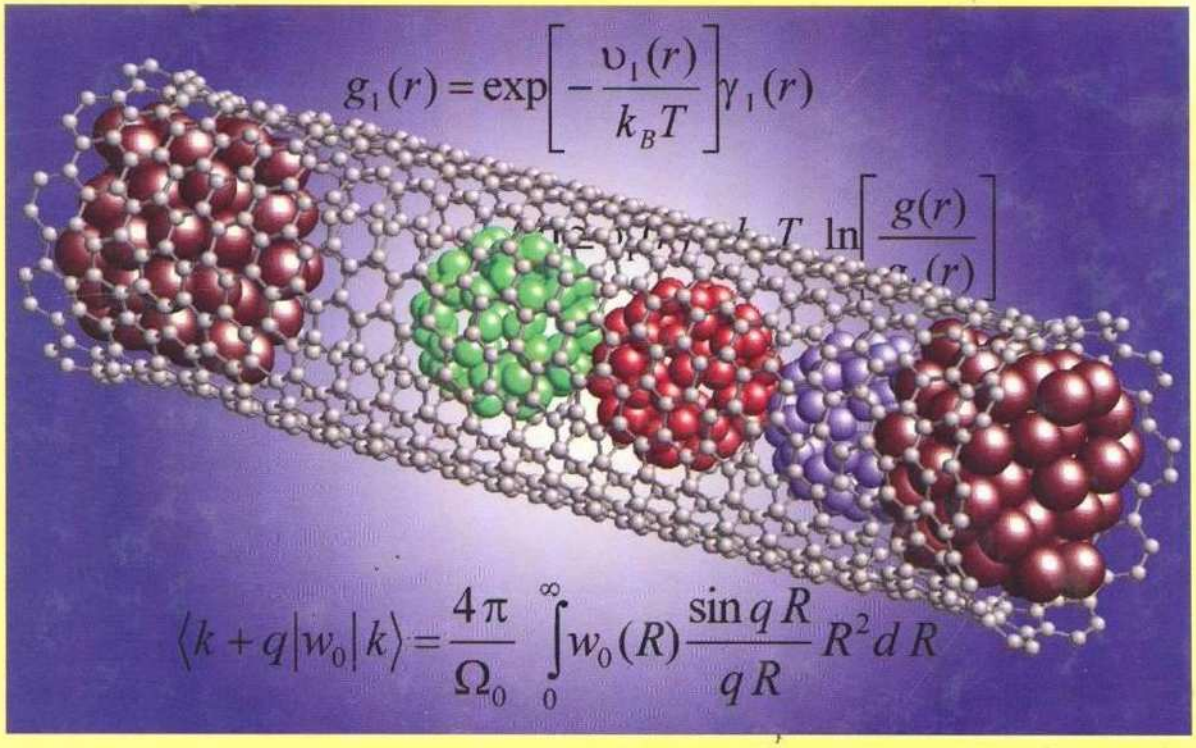
Handbook of

7

THEORETICAL and COMPUTATIONAL NANOTECHNOLOGY

Magnetic Nanostructures and Nanooptics



Edited by

M olfram Schommers

Fo

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 7

Magnetic Nanostructures and Nano-optics

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 nanolevel, 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

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

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 Naturalis* 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.

Julio A. Alonso (209)
Departamento de Física Teórica, Universidad de Valladolid, 47011 Valladolid, Spain, and Donostia International Physics Center, San Sebastián, Spain

V. I. Balykin (1)
Institute of Spectroscopy, Russian Academy of Sciences, Troitsk, Russia

GuanHua Chen (79)
Department of Chemistry, The University of Hong Kong, Hong Kong, People's Republic of China

Luiz Davidovich (329)
Instituto de Física, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil

Josef Fidler (623)
Vienna University of Technology, Institute for Solid State Physics, Vienna, Austria

C. Henkel (463)
Institut für Physik, Universität Potsdam, Potsdam, Germany

Ulrich Hohenester (395)
Institut für Physik, Theoretische Physik, Karl-Franzens-Universität Graz, Graz, Austria

Tapas Kar (127)
Department of Chemistry and Biochemistry, Utah State University, Logon, Utah, USA

V. V. Klimov (1)
Institute of Spectroscopy, Russian Academy of Sciences, Troitsk, Russia

Jens Kortus (503)
Max-Planck-Institut für Festkörperforschung, Stuttgart, Germany, and Institut de Physique et Chemie des Matériaux de Strasbourg, France

V. S. Letokhov (1)
Institute of Spectroscopy, Russian Academy of Sciences, Troitsk, Russia

Zhigang Li (127)
State Key Laboratory of Applied Optics, Changchun Institute of Optics and Fine Mechanics, Chinese Academy of Sciences, Changchun, People's Republic of China

Chi-Chiu Ma (79)
Department of Chemistry, The University of Hong Kong, Hong Kong, People's Republic of China

J. Mathon (677)
Department of Mathematics, City University, London, United Kingdom

Wounjhang Park (263)
Department of Electrical and Computer Engineering, University of Colorado, Colorado, USA

Andrei V. Postnikov (503)
Institute of Metal Physics, Russian Academy of Sciences, Yekaterinburg, Russia, and Osnabrück University, Department of Physics, Osnabrück, Germany

Angel Rubio (209)
Departamento de Física de Materiales, Facultad de Química, Universidad del País Vasco, Centro Mixto CSIC-UPV, and Donostia International Physics Center, San Sebastián, Spain

Willi Schepper (563)
University of Bielefeld, Bielefeld, Germany

Werner Scholz (623)
Vienna University of Technology, Institute for Solid State Physics, Vienna, Austria

Thomas Schrefl (623)
Vienna University of Technology, Institute for Solid State Physics, Vienna, Austria

A. Umerski (677)
Department of Applied Mathematics, Open University, Milton Keynes, United Kingdom

XiuJun Wang (79)
Department of Chemistry, The University of Hong Kong, Hong Kong, People's Republic of China

Rui-Hua Xie (127)
Department of Chemistry, Queen's University, Kingston, Ontario, Canada

Yang Zhao (79)
Department of Chemistry, The University of Hong Kong, Hong Kong, People's Republic of China

Handbook of Theoretical and Computational Nanotechnology

Edited by

Michael Rieth and Wolfram Schommers

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 NANO-OPTICS

Volume 8. FUNCTIONAL NANOMATERIALS, NANOPARTICLES, AND POLYMER DESIGN

Volume 9. NANOCOMPOSITES, NANO-ASSEMBLIES, AND NANOSURFACES

Volume 10. NANODEVICE MODELING AND NANOELECTRONICS

CHAPTER 1

Atom Nano-optics

V. I. Balykin, V. V. Klimov, V. S. Letokhov
Institute of Spectroscopy of Russian Academy of Sciences Troitsk, Russia

CONTENTS

ISBN: 1-58883-049-7

Handbook of Theoretical and Computational Nanotechnology
Edited by Michael Rieth and Wolfram Schommers
Volume 7: Pages (1–78)

1. INTRODUCTION

Atom nanooptics is a part of a more general domain of *nanooptics* that includes both photon nanooptics and atom nanooptics. Figure 1 schematically illustrates these individual areas of nanooptics, which are intimately interrelated by the configurations of the light fields produced and used therein: light field in the focus of a light beam (light far field), standing light waves, laser near field, and evanescent laser field.

The term of *atom optics* is due to the natural analogy with *light optics* or the optics of photons. Light optics is based on two fundamental principles: (a) the wave properties of light and (b) the electromagnetic interaction between light field and matter or, in other words, between light and bound charged particles (electrons or ions) in a medium. Owing to this interaction, the light field can be reflected by the medium or diffracted by it, or else light can propagate through the medium with some velocity other than the velocity of light in a vacuum, and so on [1]. Atom optics is considered in two books [2, 3].

According to the de Broglie idea, wavelike properties are associated with any particles of matter, and the de Broglie wavelength is defined by the fundamental relation (Fig. 2)

$$\lambda_{\mathrm{Br}} = \frac{h}{p} = \frac{h}{M\mathrm{v}} \tag{1}$$

where h is Planck's constant and p, M, and v are the momentum, mass, and velocity of particle, respectively. The wave properties of massive particles were verified in experiments on the diffraction of electrons and used in the first light optics analog for particles—*electron optics* [4]. Electron optics is based on (a) the wave properties of electrons and (b) the electromagnetic interaction between moving electronic charge and electrical and magnetic fields of appropriate configuration [4]. The most familiar application of electron optics is electron microscopy [5].

Another light optics analog is *neutron optics* based again on (a) the wave properties of ultra cold neutron and (b) the interaction between neutrons and atomic nuclei, which can be described by means of what is known as the optical potential [6]. As distinct from electron

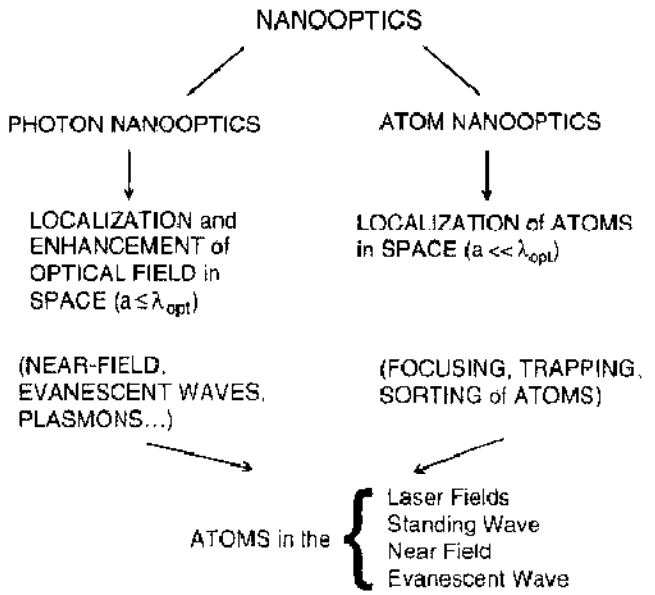


Figure 1. Nanooptics as combination of photon nanooptics and atom nanooptics. Common features of both trends: atoms in laser field of various configurations (standing wave, near-field, evanescent wave).

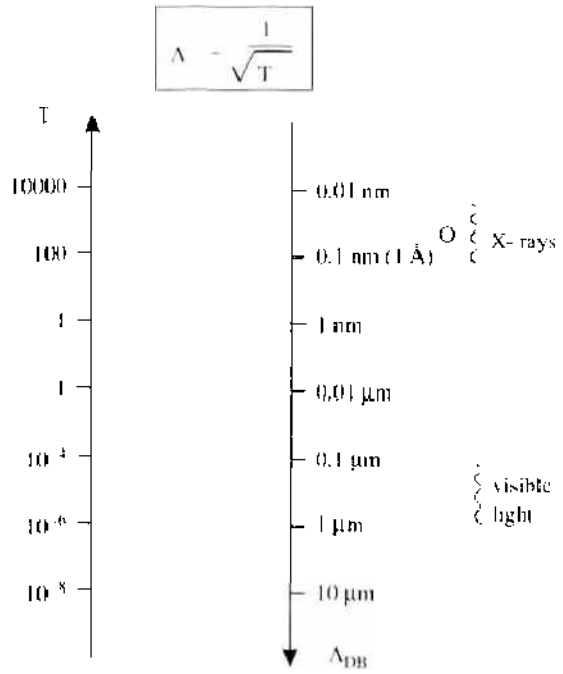


Figure 2. Correspondence between Kelvin temperature and the thermal de Broglie wavelength of a typical atom.

optics we deal here with more massive particles (ultracold neutrons) whose wave properties manifest themselves at low temperatures [7]. The effect of gravitation and low intensity of ultracold neutron sources make experiment in neutron optics more complex than in electron optics. Nevertheless neutron interferometers [8] and microscopes [9–11] have already been successfully realized.

The next natural object are neutral atoms or molecules. The wave properties of atoms and molecules and various types of their interaction with matter and electromagnetic fields (from static to optical) make it possible to implement *atom* and *molecular optics*. It is precisely the great variety of methods for exerting effect on an atom (or molecule) possessing a static electrical or magnetic moment, a quadruple moment, and optic resonance transitions (or a high-frequency dipole moment) that form the basis for several possible ways to realize atomic (molecular) optics. Let us consider them briefly.

2. METHODS OF REALIZATION OF ATOM OPTICS

The known methods to implement atom optics (atomic-optical effects) can be classed in the following three categories:

(1) Methods based on the interaction between atoms and matter.
(2) Methods based on the interaction between atoms having a magnetic or electrical dipole moment and a static electrical magnetic field of a suitable configuration.
(3) Methods based on the resonance (or quasi resonance) interaction between atom and a laser field.

The first experiment on atom optics realized by method (1) and (2) were successfully conducted almost a century ago. The advent of tunable laser allowed the possibility to demonstrate atom optics based on the atom-light interaction. It is exactly this type of atom optics that the present review is devoted to. However, for the sake of generality of the physical picture, it seems advisable to recall briefly the milestones in all approaches to atom (molecular) optics.

2.1. Interaction Between Atoms and Matter

In his classical monograph [12], Ramsey (Chapter 2, Section 5) considered the mirror reflection and diffraction of molecular beams on the surface of a solid. According to Ramsey, for

mirror reflection to occur, it is necessary that the following two conditions be satisfied:

1. The projection of the height of surface irregularities on the direction of molecular beam must be shorter than the de Broglie wavelength. Recall an example from light optics: smoked glass is a poor reflector in the case of perpendicular incidence and a good reflector in the case of grazing incidence. If δ is the average height of surface irregularities and ϕ is the grazing angle of incident beam, the above requirement may be expressed as (Fig. 3[a]).

$$\delta \sin \phi < \lambda_{Br} \tag{2}$$

2. The average residence time of the particle on the surface must be short. In this case, the state of reflected particles will be the same as that of incident particles. The roughness of most thoroughly mechanically polished surfaces is of the order 10^{-5} cm, whereas the de Broglie wavelength of hydrogen at 300 K amounts to 10^{-8} cm. Therefore, according to Eqs. (1) and (2), the condition for the reflection has the form $\phi < 10^{-3}$rad.

It was more than 50 years ago that they managed to observe a 5% reflection of hydrogen beam from polished bronze mirror at the grazing angle of $\phi = 10^{-3}$grad [13]. Cleaved crystal surfaces are much smoother. The thermal vibrations of the crystal lattice limits the roughness of the surface to about 10^{-8} cm. In that case a beam of He atoms should undergo reflection at grazing angles less than 20–30 grad. This was confirmed in the experiments [14] with He atoms and LiF crystal. (Fig. 3[b]). The temperature dependence of the grazing angle marking the onset of simple reflection of atoms bears witness to the fact that thermal vibrations have an effect on the surface roughness of crystal.

Experiments on the simple reflection of atoms at the surface of condensed medium continue to draw investigator's attention. Recall the experiments on the reflection of ^{3}He atoms grazing the surface of liquid ^{4}He [15] and thermal Cs atoms grazing a polished glass surface [16].

The first experiment aimed at observing the diffraction of atoms by a cleaved crystal surface acting as a two-dimensional plane grating were conducted by Stern [17] and the results of detailed research into this phenomena were presented in Ref. [18]. The diffraction of atoms by a fabricated periodic structure (a slotted membrane) with a much more grating period was observed in the work reported in Ref. [19].

The effect of quantum-reflection of ^{4}He and ^{3}He beams at a liquid-helium-vacuum interface was successfully used to focus hydrogen atoms with a concave grating [20], and the authors of Ref. [21] were successful in conducting an experiment on focusing a beam He atoms by means of a zone plate.

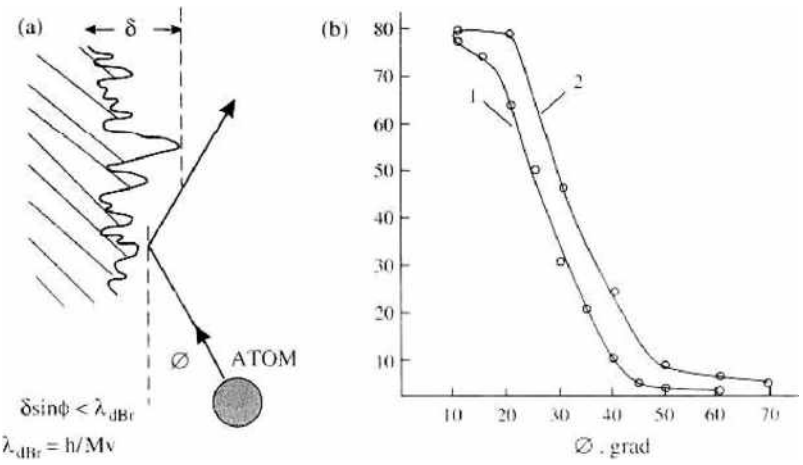


Figure 3. Reflection of atom in grazing incidence upon the surface of a solid: (a) requirement for the surface roughness δ, the glancing angle of the incident beam ϕ, and the de Broglie wavelength; (b) reflectivity of a beam of He atoms from the surface of LiF crystal at two different temperatures (1, 195 K; 2, 100 K) [14].

Atomic interferometry based on the microfabricated structures was realized in two elegant experiments: The atomic Young's two slit interferometer [22] and the atomic Michelson interferometer [23].

2.2. Interaction Between Atoms and Static Electric or Magnetic Fields

Some elements of the optics of atoms and molecules, based on the interaction between spatially nonuniform static magnetic or electrical fields and the magnetic or electrical dipole moment of the particles, have long been known and used fairly successfully in experimental physics. An excellent review on the early experiments in this field was presented by Ramsey (1956) [12].

In the presence of a magnetic or electric field, the quantum state of atom or molecule are shifted, the shift depending on the initial quantum state of the particle and the field strength (the Zeeman and Stark effects). In the adiabatic approximation (the field varies in the time and space not very rapidly; the particles move slower enough), the internal state of particles follows the field-strength variation, or, in other words, the particles remains at one and the same quantum sublevel whose energy W depends on the field strength.

In the adiabatic approximation, the motion of the center of mass of a neutral particle with a mass M obeys the Schrödinger equation for the wave function $\psi(r, t)$:

$$ih\left[\frac{\partial\psi_i(\mathbf{r}, t)}{\partial t}\right] = \left\{\left(\frac{-\hbar^2}{2M}\right)\nabla^2 + W_i(\bar{\mathbf{r}})\right\}\psi_i(\mathbf{r}, t) \tag{3}$$

where $W_i(\mathbf{r})$ is the internal energy of the particle in the quantum state i at the point $\mathbf{r}$ that depends on the electrical field strength $E(\mathbf{r})$ or the magnetic field strength $H(\mathbf{r})$.

Magnetic Interaction In the simple case of a constant magnetic moment μ, the effective potential energy W of an atom or a molecule in an external magnetic field of strength H is given by

$$W = -\mu\mathrm{H} = -\mu_{\mathrm{eff}}H \tag{4}$$

where μ_{eff} is the projection of μ, on the direction of H. It follows from the relationship between force and potential energy that the force acting on the atom or molecule is

$$\mathrm{F} = -\nabla W = -\left(\frac{\partial W}{\partial H}\right)\nabla H = \mu_{\mathrm{eff}}\nabla H \tag{5}$$

A particle in nonuniform magnetic field is acted on by the force directed along the field strength gradient.

The authors of Refs. [24–27] proposed to use nonuniform magnetic field to focus molecular beams issuing at different angle from the source.

Figure 4 shows the configuration of the focusing magnetic field used by Frieburg and Paul [24, 25]. The method was extended by Ref. [28] to atoms whose magnetic moments depend on the strength of the external magnetic field.

The focusing properties of a magnetic lens depends on the magnetic sublevel of atoms. That was successfully used by Ramsey and co-workers to create the hydrogen maser [29, 30]. The hydrogen atoms in the state $F = 1$, $M = 0$ were focused in a small hole in the wall of the storage cell and accumulated there, while the atoms in me lower atoms state $F = 0$ are defocused.

Electrical Interaction. Since the energy of atom or molecule in an electrical field depends on the strength of the latter, it then can be presumed, by analogy with Eqs. (4) and (5), than the atom or molecule possesses an effective dipole moment given by

$$\mu_{\mathrm{eff}} = -\left(\frac{\partial W}{\partial E}\right) \tag{6}$$

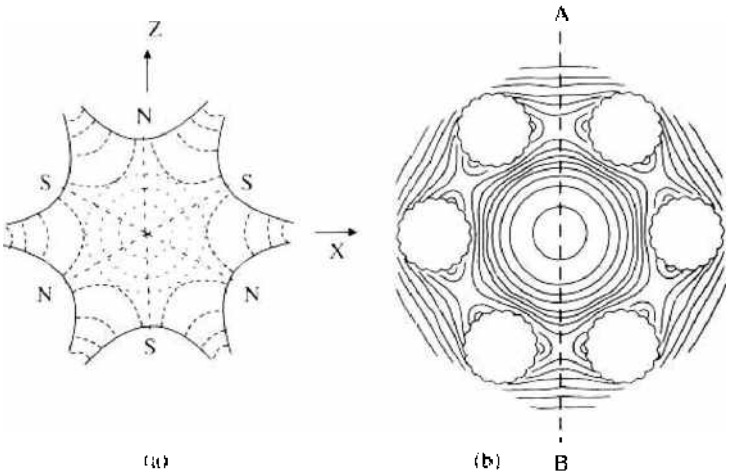


Figure 4. (a) Ideal focusing magnetic sextupole field. Dashed, magnetic field lines; dotted, lines of equal magnetic potential. B = const. (b) Linear sextuple made of six straight current leads with alternating current direction [24].

The force acting on an atom or a molecule in a nonuniform electrical field is defined by the following expression similar to (5):

$$\mathbf{F} = \mu_{\text{eff}} \nabla E = \mu_{\text{eff}} \left(\frac{\partial E}{\partial z} \right) \tag{7}$$

where the direction of the field strength gradient is taken to be the z-axis.

Paul and co-workers [28] created focusing electrical fields for a beam of polar molecules. The electrical focusing of a beam of molecules in a certain (excited) quantum state was used by Townes in developing the NH_3 maser [31, 32]. The hexapolar electrical field configuration (as in Fig. 4) possesses not only focusing properties but also selectivity with respect to the quantum state of the molecule, for the quantity μ_{eff} depends on its quantum numbers J, K, and M.

This latter property was successfully used in experiments on molecular dynamics with a beam of molecules in a specified quantum state, including the experiments on the orientation of molecules [33].

When speaking of the optics of atomic or molecular beams, we almost always mean their focusing, for it is exactly this effect that has found practical application. But one can also speak electrical or magnetic mirrors and gratings for slow-moving neutral atoms and molecules [34].

2.3. Interaction Between Atoms and Light Field

Atoms or molecules having no static, magnetic, or electrical dipole moment cannot change their mechanical trajectory in a static magnetic or electrical field. However, new possibilities are being opened up for particles based on the induction in them of a high-frequency (optical) electrical dipole moment in a quasi resonant or resonant laser light field. Before the advent of the laser, it was only possible to induce microwave transitions in atoms and molecules, which allowed one to alter efficiently their quantum state and thereby the character of motion in external spatially nonuniform electric or magnetic field [12].

An atom in quasi resonant laser field acquires a high-frequency polarizability, and if the intensity of the laser field is spatially nonuniform, the atom is acted on by the *gradient* (*dipole*) force [35]. For example, the gradient force in a standing laser light wave may cause the channeling of atoms moving along the wave front [36]. The gradient force was successfully used by Ashkin and co-workers to focus an atomic beam [37].

In the optical region of the spectrum, the recoil effect resulting from atom light interaction is significant. This effect was predicted by Einstein [38] as far as back as 1909 and was experimentally corroborated by the slight deflection of a beam of sodium atoms scattering the resonant radiation of D line of Na [18].

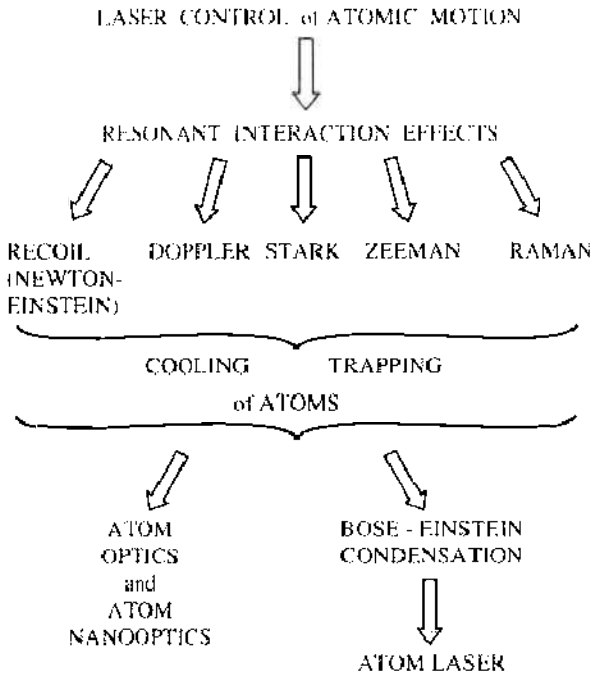


Figure 5. Trends of laser control of atomic motion based on various resonant effects of laser resonant interaction effects with atoms leading to cooling and trapping effects of atoms and atom optics, atom nanooptics, Bose-Einstein condensations, and atom laser.

An intense laser radiation tuned to the resonance with some allowed dipole transition in an atom can make it re-emit millions of photons, and as a consequence, radiation can exert a substantial effect on the atomic velocity and mechanical trajectory. Hänsch and Schawlow [39] proposed to use the resonance force due to spontaneous re-emission of photons for cooling neutral atoms and Wineland and Dehmelt [40] to cool ions in an electromagnetic trap.

Since that time a new line of inquiry has started developing in atomic physics, based on the effects of resonance interaction between laser light and atoms and the use of the well-known atom physics effects: the recoil effect, Doppler effect, Stark effect, Zeeman effect, and Raman effect. This has led to the development of very effective methods for the laser cooling and trapping of atoms, which in turn led to the advent of atom optics, including atom nanooptics and the physics of ultracold atoms, including quantum degenerate gases (Bose-Einstein condensation, Fermi degenerate gases, atom lasers, etc.). This evolutionary sequence of ideas that led to the development of atom nanooptics is illustrated in Fig. 5.

The gradient and spontaneous forces are at the root of a great many experiments on controlling the motion of atom and atomic optics by means of light, which was already considered in the reviews [41–50], special issue of scientific journals [51–53], monographs [54, 55], and textbook [56].

3. RADIATION FORCES FOR MANIPULATION OF ATOMS

Radiation force means the total force arising upon interaction between laser light and an atom. Depending on the spatial and temporal structure of a light field, its strength and wavelength, the radiation force may be a very complex function of the atom's position and velocity. But, since all the known studies on the application of the radiation forces have been carried out mainly using three types of light fields, namely, a plane-traveling light field, a Gaussian laser beam, a standing light wave, or their combination, we will restrict ourselves to qualitative consideration of these types of fields only.

In this review, we will consider the motion of atoms under different types of radiation forces. Since the conception of force is a classical one, it is necessary to mention under what condition the motion of atom can be considered classical [50, 54].

There are two such conditions. One of them directly follows from the fact that for classical atomic motion the quantum fluctuation in the atomic momentum must be negligibly small,

compared to the change in the average atomic momentum due to the force. The minimum atomic momentum variation in a laser field is equal to the photon momentum. The change in the average atomic momentum under action of the radiation force is to be considered significant, if it breaks of the resonant atomic interaction with the light field. The frequency range within which an atom absorbs the resonant radiation is determined by natural width of spectral line 2γ. The resonance between atom and light is interrupted when the average atomic momentum varies by the value $\delta p \approx M\gamma/k$. If one requires that the value of this atomic momentum δp is larger than the photon momentum, the first condition for a classical atomic motion will be

$$hk \ll \frac{M\gamma}{k} \tag{8}$$

or

$$R \ll \hbar\gamma \tag{9}$$

where $R = \hbar^2 k^2/2M$ is the recoil energy.

The second condition for classical motion follows from the fact that the process of absorption—spontaneous emission takes place in time interval of the order of $\tau_{sp} \approx \gamma^{-1}$. Because the classical description of atomic motion cannot allow for such small-scale momentum variations, it is necessary to restrict classical time scale to the condition:

$$\delta t \gg \gamma^{-1} \tag{10}$$

We could look at the problem of classical description of motion of atom from the point of view of an evolution of the atomic wave packet. The quantum description of the evolution of the atomic wave packet will be close to the classical one if the atom has a well-defined position and momentum. The momentum width Δp of wave packet related to the position width Δr by the Heisenberg inequality

$$\Delta p * \Delta r > \hbar \tag{11}$$

The force exerted by laser field on the atom varies over distances on the order of laser wavelength λ or larger. The force could be considered classical if the position spread of wave packet is a smaller than the laser wavelength:

$$\Delta r \ll \lambda \tag{12}$$

The force also depends on the velocity of atom because of the Doppler shift. The appreciable change of the atomic response to the laser excitation through the Doppler shift will be at the velocity change:

$$\delta v \cong \frac{\gamma}{k} \tag{13}$$

It is clear that velocity (momentum) spread Δv of wave packet (for classical description of atomic motion) must be smaller than δv:

$$\Delta v \ll \delta v = \frac{\gamma}{k} \tag{14}$$

or

$$\Delta p \ll \frac{\gamma M}{k} \tag{15}$$

Equations (12) and (15) impose upper bound on the momentum and space spread of wave packet, which can be in conflict with the Heidelberg inequality (11). From (11), (12) and (15) follows one condition for classical description of atomic motion:

$$hk \ll \frac{M\gamma}{k} \tag{16}$$

which also coincides with Eqs. (8) and (9).

To date, the theory of atomic motion in laser fields has been developed quite well [50–52, 54, 55].

3.1. A Plane Traveling Wave: Light Pressure (Spontaneous) Force

Consider a plane wave directed along the z-axis and having its frequency tuned to resonance with the absorption frequency of an atom placed in it (see Fig. 6):

$$\mathbf{E} = \mathbf{e}E_0 \cos(kz - \omega t) \tag{17}$$

The atom absorbs laser photons directed along the z-axis and re-emits spontaneous photons symmetrically in all directions. As a result of these processes, a radiation force is determined by the product of the momentum of a photon and the rate of photon scattering:

$$F_{l,p} = \hbar k Q \tag{18}$$

where $Q = 2\gamma n_2$ is the rate of photon resonant scattering by atom, and n_2 is a steady state population of exited state.

The atom is acted on in the direction of the wave by radiation force whose maximum magnitude is given by the product of photon momentum $\hbar k$ by the photon scattering rate Q (i.e., $F_{max} = \hbar k\gamma$, where $k = 2\pi/\lambda$). If the atom is not in exact resonance with laser radiation and has a velocity v projected onto the z-axis in the direction of a light wave, a steady state population of an excited state depends on the projected velocity and on the detuning of the field frequency ω from the absorption frequency ω_0 of the atom:

$$n_2 = \left[\frac{G}{2}\left(1 + G + \frac{(\Omega - kv)^2}{\gamma^2}\right)^{-1}\right] \tag{19}$$

where $G = I/I_s$, I is the intensity of the wave, I_s is the saturation intensity of the atomic transition, and $\Omega = \omega - \omega_0$ is a detuning of field frequency from an absorption frequency.

From Eqs. (18) and (19), follow the expression for the radiation force in a traveling wave, which is called a *light pressure force* [57]:

$$F_{l,p} = \hbar k\gamma\left[\frac{G}{(1 + G + (\Omega - kv)^2/\gamma^2)}\right] \tag{20}$$

The force has a Lorentzian dependence on the velocity of atom (Fig. 6) and reaches its maximum value at exact resonance. The ultimate, value of the force is limited by the saturation effect to F_{max}. The acceleration of an atom under the action of this force reaches a magnitude, of 10^8 cm/s^2, which is 10^5 times greater than the acceleration of the earth gravity.

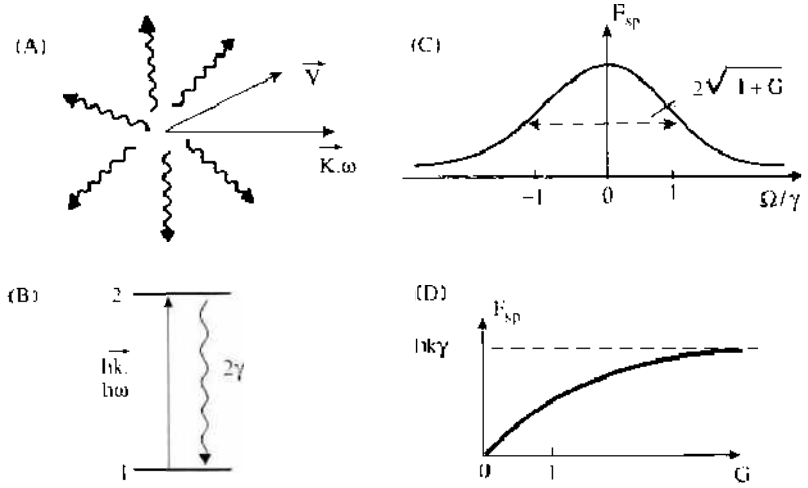


Figure 6. Origin and properties of light pressure (spontaneous) force F_{sp}: (A) absorption of directed ($\mathbf{k}$, ω), photon and isotropic remission of spontaneous photon by atom with velocity v; (B) an ideal two-level diagram of resonant interaction of light wave with frequency ω and the quantum transition frequency ω_0 with the radiative damping rate 2γ of the excited state; (C) resonant dependence of F_{sp} as a function of frequency detuning $\Omega = \omega - \omega_0 = kv$; (D) saturable dependence of F_{sp} as a function of saturation parameter G.

Under optimum condition, dissipative light pressure force Eq. (20) allows the atoms to be cooled to the temperature of tthe so-called Doppler limit [58, 59]:

$$\mathbf{T}_{\text{min}} = \frac{\hbar\gamma}{2k} \tag{21}$$

3.2. Traveling Gaussian Light Beam: Gradient (Dipole) Force

In the case of a Gaussian light beam of spatially nonuniform distribution of amplitude and phase, the electrical field is

$$\mathbf{E} = \mathbf{E}_0(r)\cos(kr - \omega t), \quad \mathbf{E}_0(r) = \mathbf{E}_0 \exp\left(\frac{-r^2}{2\rho^2}\right) \tag{22}$$

In such a field an atom is acted upon the radiation force F_{RAD}, which can be expressed as the sum of the light pressure force F_{LP} and what is known as the gradient force F_{GR} (Fig. 7):

$$\mathbf{F}_{\text{RAD}} = \mathbf{F}_{\text{LP}} + \mathbf{F}_{\text{GR}} \tag{23}$$

For a Gaussian laser beam, the light pressure force is a direct generalization of the force (18).

Depending on the detuning of laser radiation frequency with respect to atomic transition frequency, the laser field either expels atoms out of the beam ($\Omega > 0$) or draws it toward the beam center ($\Omega < 0$). In terms of atomic and laser parameter the gradient force is expressed as Refs. [54, 60]:

$$F_{\text{GR}} = \hbar\left(\frac{\rho}{\rho_{10}^2}\right)(\Omega - k\mathrm{v}_z)\left[\frac{G(r)}{(1 + G(r) + (\Omega - k\mathrm{v}_z)^2/\gamma^2}\right] \tag{24}$$

where the laser beam radius ρ_{10}, ρ the distance between the atom and the laser beam axis.

The dependence of the magnitudes and signs of the spontaneous and gradient forces are of entirely different character (Figs. 6 and 8). Unlike the light pressure force, the value of gradient force is not limited as the saturation parameter is increased. The maximum value of the force increases proportionally to $G^{1/2}$. The velocity dependence of gradient force is determined by the curve of a dispersive form (Fig. 8).

The physical reason for the existence of the gradient force is an effect of a spatially nonuniform optical field on the atomic dipole moment that is induced by the laser field. From the standpoint of quantum mechanics, the gradient force results from the absorption and stimulated emission of photons by the atoms in a *nonuniform* laser field (Fig. 8[a]).

Both forces (light pressure force and gradient) are essential to the control of atomic motion. For example, the gradient force allows one to realize atom optics. The spontaneous force makes it possible to cool atom and ions.

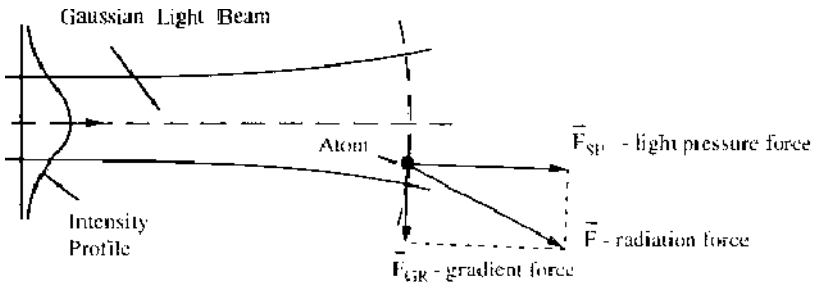


Figure 7. Radiation pressure force acting on an atom in traveling Gaussian laser beam. The Gaussian laser beam exerts a gradient force on the dipole moment that intense light beam induces in atom. The resultant force of the laser field is not simply in the propagation direction of the laser beam.

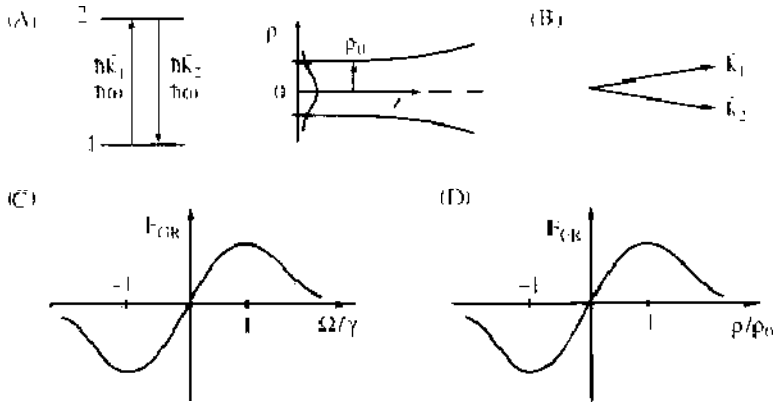


Figure 8. Origin and properties of gradient (dipole) force F_{gr}: (A) absorption and stimulated emission of directed photons ($\mathbf{k}_1$, ω_1) and ($\mathbf{k}_2$, ω_2); (B) the Gaussian beam with transversal profiles of intensity $I(\rho)$; (C) resonant dependence of F_{gr} as a function of frequency detuning Ω; (D) dependence of F_{gr} as a function of atomic displacement of atom ρ from optical axis.

3.3. Plane Standing Wave: Two-Level Atom

This wave is formed from two plane waves traveling toward each other:

$$\mathbf{E} = \mathbf{e}E_0 \cos(kz - \omega t) + \mathbf{e}E_0 \cos(-kz - \omega t) = \mathbf{e}2E_0 \cos\omega t \cos kz \tag{25}$$

At low radiation intensity ($G \ll 1$) and for a two-level atom, the radiation force in the standing wave is determined by the sum of the forces Eq. (18) from each traveling waves.

However, at high saturation G, this no longer holds. The interaction between a two-level atom and the two counterpropagating waves at $G > 1$ gives rise to the new effects beyond the scope of the spontaneous light pressure and gradient forces.

In order that we can consider classically the motion of the atom in such a wave, it is necessary that the following condition be fulfilled:

$$p_z \gg \hbar k \tag{26}$$

where $p_z = \mathrm{M}v_z$, the component of atomic momentum along the standing wave. It can be clearly seen that only in this case, according to uncertainty principle and with the reasonable relation $\Delta p_z < p_z$, do we uncertainty of coordinate $\Delta z < \lambda$.

We assume also the following condition on the atomic and laser parameters:

$$kv_z < \gamma \tag{27}$$

$$kv_z < \frac{\Delta^2}{(\gamma G_0^{1/2})} \tag{28}$$

where G_0 is the transition saturation parameter of one traveling wave. The inequality Eqs. (27) and (28) reflects the absence of the Landau-Zener transition [61]. Also, we assume that the relation $t_{\mathrm{int}} \gg \gamma^{-1}$.

3.3.1. Light Pressure Force

In a single standing light wave, two resonances occur under weak saturation, one for each traveling waves: $\pm kv_z = \Omega = \omega - \omega_0$ (Fig. 9). At $\omega < \omega_0$, the counterpropagating waves decelerate the atom, whereas, its co-propagating counterpart accelerates it. The combine effect of the two traveling waves acting independently is described by a curve of dispersive character (Fig. 9). At exact resonant ($kv_z = 0$), the friction force is zero, and the slope of the curve at $v_z = 0$ gives the friction coefficient η for atom, which governs its cooling rate.

For low velocity of atoms,

$$kv \ll \gamma$$

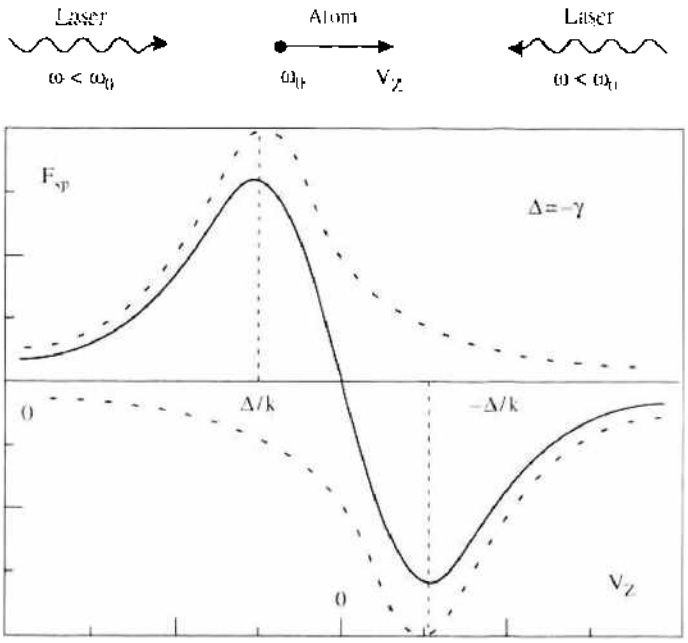


Figure 9. Dependence of light pressure force as a function of velocity projection v_z for traveling wave (dotted line) and for standing wave (solid line); frequency detuning $\Omega = \omega - \omega_0 = \gamma$.

the sum of the two gives a damping linear in the atomic velocity:

$$F_{LP} = \eta v \tag{29}$$

$$\eta \cong -4\hbar k^2 \left[\left(\frac{\Omega}{\gamma} \right) \frac{G}{(1 + \Omega^2/\gamma^2)^2} \right] \tag{30}$$

where η is a damping (or friction) coefficient. In a weak standing wave at $\omega < \omega_0$, the traveling wave propagating counter to the moving atom cools it. But when saturation is strong, the stimulated re-emission of photons from one traveling wave into the other becomes predominant. At $\omega < \omega_0$, this stimulated re-emission change the sign of friction coefficient so that in the neighborhood of $kv \cong \gamma$; the cooling of the atom turns to its heating [62]. However, where the frequency, detuning of the laser field is positive ($\omega > \omega_0$), such re-emission of photons gives rise to a new cooling force that depends on stimulated emission and a light-induced shift of energy levels.

3.3.2. Gradient Force

A spatial modulation of the laser field intensity in the standing wave with a period $\lambda/2$ begins to show up with increasing radiation intensity, which leads to the appearance of the gradient force of potential nature:

$$\mathbf{F}_{GR} = -\left(\frac{dU_g}{dz} \right) \mathbf{e}_z \tag{31}$$

The associated potential has the form [63]

$$U_{GR} = \left(\frac{\hbar\Omega}{2} \right) \ln(1 + s) \tag{32}$$

where

$$s = \frac{G}{[1 + (\Omega/\gamma)^2]} \tag{33}$$

is an off-resonance atomic transition saturation parameter, $G = 4G_0 \cos^2(kx)$ is the atomic transition saturation parameter the standing light wave and $G_0 = d^2E^2/2\hbar^2\lambda^2\gamma^2$.

In the limiting case of weak saturation and great frequency detuning,

$$G \ll |1 + (\Omega/\gamma)^2| \tag{34}$$

the expression for the potential U_{GR} reduces to the simple form of the light wave, and a negative detuning ($\Omega < 0$), they coincide with the loops of the wave:

$$U_{GR} = \left(\frac{2\hbar\Omega_R^2}{\Omega}\right)\cos^2(kx) \tag{35}$$

where $\omega_R = dE_0/\hbar$ is the Rabi frequency. It follows from expressions (32) and (35) that the period of the potential is equal to half the optical wavelength. At positive frequency detuning ($\Omega > 0$), the potential wells coincide with the nodes of the light wave, and at a negative detuning ($\Omega < 0$), they coincide with loops of the standing wave.

3.3.3. Stimulated (Retarded) Force

As noted in Section 3.3.1, the saturation parameter is increased, the force-versus-velocity curve starts developing the high-order resonances due to the nonlinear interaction between atoms and the both counterpropagating waves. Stimulated transition in a standing light wave, which destroys the cooling effect in the case of strong saturation at a negative frequency detuning, can nevertheless be used to cool atoms, but at oppositive detuning this time [64–66]. That this is possible was demonstrated with success in [67].

To gain an insight into this interesting effect, account should be taken of two quantum-stimulated transitions in the standing light-wave field, which give rise to the spatially periodic atom-field interaction potential $U(z)$ (or the gradient force $F(z) = \text{grad}\, U(z)$) in quantum mechanical terms, a light induced shift of atomic energy level that also oscillates in space. Combined with the spontaneous transition of the atom moving in the standing light wave, this gives rise to the friction force [64–68].

The quantum mechanical treatment of the interaction of a two-level atom with a standing light wave shows also that the total radiative force depends on the atomic velocity. In a zeroth order of approximation of this force (for small parameter $|kv_z/\gamma| \ll 1$) the gradient force (32) appears, and in first order, the friction force [66, 68]:

$$\mathbf{F}_f = 2\hbar\left(\frac{\Omega}{\gamma}\right)G\left\{\frac{[1 + (\Omega/\gamma)^2 - G(1 + G/2)]}{[(1 + (\Omega/\gamma)^2 + G)^3]}\right\}\mathbf{k}(\mathbf{vk})\tan^2(\mathbf{kz}) \tag{36}$$

which is called stimulated, or retarded force.

Let us give the explanation of appearance of this force in two physically equivalent languages. The well-known light-induced shift of the energy levels of the two-level atom is given by [70–73]

$$d\omega \cong \left(\frac{1}{4}\right)E^2\left(\frac{d^2}{\hbar^2\Omega}\right) \tag{37}$$

The levels are shifted one and the same distance but in opposite directions. With $\omega < \omega_0$, the perturbation caused by light field increases the distance between the level, and vice versa with $\omega > \omega_0$, it decreases the spacing between them (Fig. 10).

The stimulated cooling of the atoms in a standing light wave can occur on account of the following predominant sequence of stimulated and spontaneous processes. Since the rate of stimulated processes W_{st} (stimulated two-quantum processes responsible for the light-induced shift of energy levels and stimulated absorption responsible for excitation of atom) is proportional to E^2, in the standing light wave loops, there predominantly occurs the excitation of the atom upon absorption of the red shifted light (Fig. 10). If the condition

$$kv_z \approx \gamma \tag{38}$$

is satisfied, the excited atom may enter the region of standing wave where the field intensity is its minimum and the light-induced level shift are small and then spontaneously emit a shorter-wavelength photon. The spontaneous emission rate is independent of the periodic

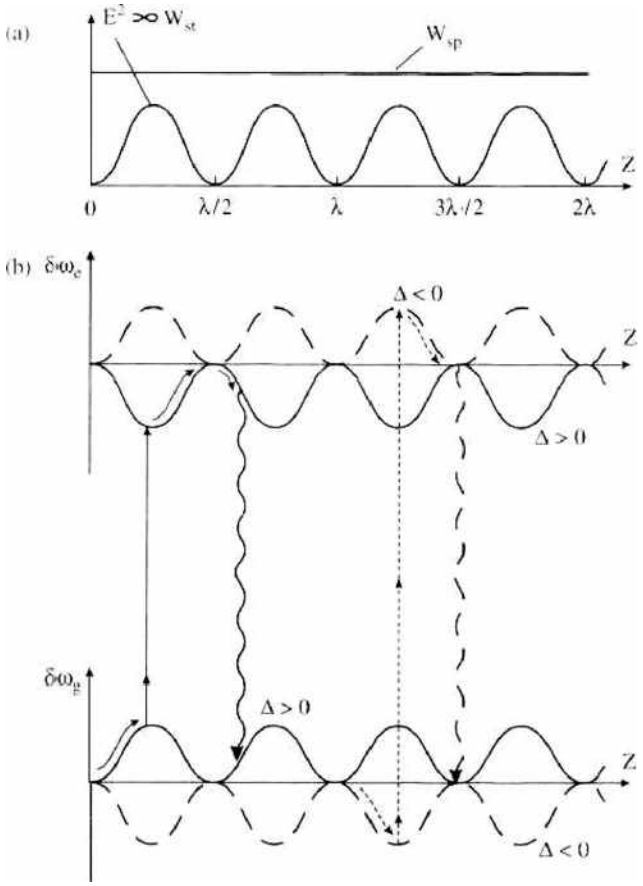


Figure 10. (a) Distribution of standing wave intensity E^2 and the spontaneous and stimulated emission probabilities W_{sp} and W_{st}; (b) and (c) energy shift $\delta\omega_g$ of the ground and excited states for positive ($\Omega > 0$) and negative ($\Omega < 0$) light-field frequency detuning. Vertical straight lines corresponds to induced absorption, vertical wavy lines correspond to spontaneous emission.

field-intensity variations. The energy difference between the absorbed and the spontaneously emitted photon is derived from kinetic energy of the atom. Such an absorption of photon in the standing tight wave loops and their spontaneous re-emission at the field minima may take place repeatedly, thus giving rise to a friction force Eq. (36) proportional to the atomic velocity.

A clear and consistent theoretical description of the cooling process in a standing light wave was made in [66] on the basic of the dressed-atom approach [73], providing a qualitative understanding of the main characteristic of the stimulated force (its mean value, fluctuations, velocity dependence, etc.) in high-intensive limit.

An important feature of stimulated force in a standing light wave (which is a consequence of the light-induced energy-level shift) that it cannot be saturated by increasing the light-field intensity [64–66, 74]. This is naturally explained by the fact that the atomic cooling rate is proportional to the energy difference between the absorbed and re-emitted photons governed by the magnitude of light-induced shift Eq. (37) proportional to the field intensity.

3.4. An Arbitrary Monochromatic Wave: Multilevel Atom

When the monochromatic wave interacts with the multilevel atom, the number of new effects grows still greater. A general description of radiation forces with an arbitrary number of degenerate or nearly degenerate sublevel and an arbitrary monochromatic radiation field were given in Ref. [75]. It allows also for an arbitrary magnetic field that adds Zeeman precession to atomic evolution. The total radiative force on the atom may be generally separated into four terms. Apart from the radiation pressure and the gradient force (which we have considered in Sections 3.3.1 and 3.3.2), there are two terms that are from the gradient of the polarization direction. One of these terms corresponds to fluorescent scattering of

absorbed photons. The other term arises from the redistribution of photons between plane waves that compose the radiation field.

The experiments [76–81] have demonstrated that cooling below the Doppler limit can occur for the system having multistate structure in the field of driven transition. An essential ingredient for this damping force is a variation of the relative orientation of the atomic dipole with the respect to the light polarization during the transversal of a wavelength: Such situation arises when atom moves through a field with polarization gradient [78, 82] or when a magnetic field causes Zeeman precession of the atomic dipole [79].

Let us consider briefly the new possibilities that are opening up.

3.4.1. Polarization Gradient Force

The effect of periodic modulation of the light-induced shift of atomic energy level (see Fig. 10) is especially manifest in the case of the three-level atom, where it enables a new mechanism of deep (sub-Doppler) atomic cooling and trapping to come into play. An excellent description of the mechanism allowing the Doppler cooling limit to overcome has been presented in the brief review [47]. These mechanisms also include, in addition to the effect of light-induced energy shifts, optical pumping and laser polarization gradients [52, 78].

The two counter-propagating laser beams with orthogonal linear polarization form a standing light wave whose local polarization changes every eight's of a wavelength from linear to circular type (Fig. 11[a]). Depending on the sign of the local circular polarization, there takes place, thanks to the optical pumping, the accumulation of atoms either at the sublevel $g_{-1/2}$ (for σ^-) or at the sublevel $g_{+1/2}$ (for σ^+). Thus, the spatially periodic modulation of the circular polarization sign causes a spatially periodic variation of the population of the Zeeman sublevels of the ground state. The shift of Zeeman sublevels also depends on the circular polarization sign: the σ^+ wave shifts only the $g_{+1/2}$ sublevels, whereas its σ^- counterpart wave shifts only the $g_{-1/2}$ ones. So, the light shift energies and population of two ground-state sublevels of the atom vary with the local polarization of the standing light wave, hence with the atom's position, as shown in Fig. 11(c), where the sublevel population is proportional to the size of the full circles.

Let the atom move at a small angle to the light-wave front, so that its velocity projection is

$$0 < v_z \ll \frac{\gamma}{k} \tag{39}$$

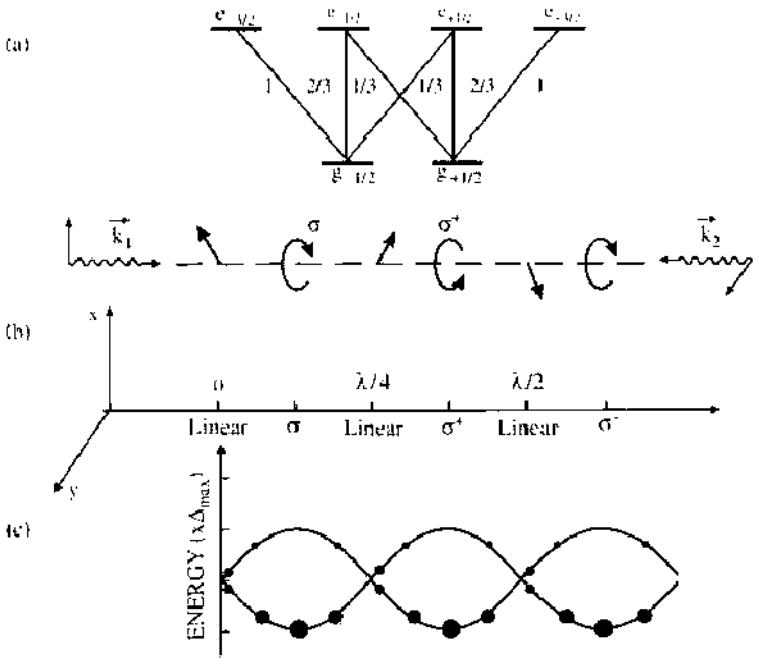


Figure 11. Explanation of atomic cooling based on the effects of polarization gradient, light-induced level shifts, and optical pumping effect for multilevel atom [47].

that is much smaller than required by condition Eq. (38). Consider for the case of definiteness an atom starting to move from the point $z = \lambda/8$ (Fig. 11). Such an atom will cover the distance of $\lambda/4$ in a time shorter than optical pumping time τ_p, and so it will climb from the potential minimum to the maximum, while residing at one and the same sublevel as illustrated in Fig. 11. The probability that the optical pumping will cause the atom to move from one sublevel to the other here becomes higher. But the spontaneous decay time is much shorter than the optical pumping time, $1/\gamma \ll \tau_p$, and therefore the spontaneous transition of the atom to the other sublevel takes place while it travels the distance of $\Delta z \ll \frac{\lambda}{4}$. As a result, the atom absorbs long-wave photon and emit short-wave ones. The energy difference between these photons is equal to the level shift (37), and so it is proportional to the light-field intensity. Thus, the rate of atomic cooling by this mechanism is proportional to the laser intensity. Naturally, this process of absorption at the top of potential hill and spontaneous descent in the valley takes place repeatedly, providing for the deep cooling of the atom to that of the order of recoil velocity v_{rec}. A detailed theoretical analysis of this atomic cooling mechanism has been made in [78, 82].

Its substantial difference from the Doppler mechanism is that it operates effectively at atomic velocities too low for Doppler mechanism to be efficient and makes it possible to reach temperatures (velocities) much below the Doppler limit (Eq. [21]), down to the recoil velocity and temperature [78]:

$$v_{rec} = \frac{\hbar k}{M} \quad \text{or} \quad T_{rec} = \frac{R}{k_B} \tag{40}$$

where R is the recoil energy.

3.4.2. Magneto-Optical Forces

In the presence of a static magnetic field, the radiative forces acting on a multilevel atom in a laser field may acquire new properties [83, 80]. Of particular interest for the creation of new types of atomic traps is a confining potential—like character of the radiation force. To produce a magneto-optical force pseudopotential character [83], a scheme based on the spontaneous force has been used. In this case, the spatial dependence of this force is determined by the Zeeman shift in the magnetic sublevel of the atom induced by spatially noni-iniform static magnetic field. One of the shortcomings of this scheme is that the maximum of this magneto-optical force is restricted by the upper limit of spontaneous force $\hbar k\gamma$, where 2γ is the natural width of exited state.

It has been shown [84] that there is magneto-optical force based on the induced photon re-emission between two counterpropagating light waves having different direction of polarization vectors. It is essential that the value of this magneto-optical force may exceed $\hbar k\gamma$.

4. LIGHT FIELD CONFIGURATIONS FOR NANOOPTICS

To implement atom optical elements using the gradient force requires spatially inhomogeneous fields with inhomogeneities on the scale of less than the wavelength of light (i.e., on a submicron scale). The search for the ways to optimize the suitable field configurations is continuing.

First, already at the onset of investigations into the manipulation of atoms in a light field, they suggested channeling atoms over the wavefront surfaces of a standing light wave [36], as illustrated in Fig. 12. The intensity of inhomogeneities of the standing light wave field with a period of $\lambda/2$ allow one to effect the guiding (one-dimensional [1D] trapping) of atoms. Such a motion of atoms contains elements of their reflection and focusing by light. It was precisely the latter fact that was used in the first experiments on atomic photolithography using standing light waves [85] (for details, see Section 7).

Next there was proposed an "atomic objective lens" with a spatial resolution of a few angstrom units. Such a lens is essentially a focused laser field with its frequency well away from the atomic-transition frequency. The atomic beam propagates along the lens axis. The tight focusing of the laser radiation produces an efficient "thin" lens, and the choice of

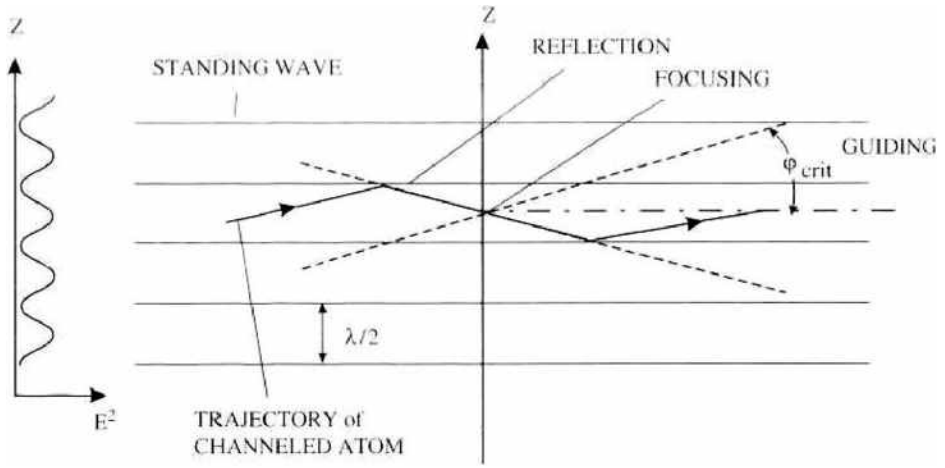


Figure 12. Idea of channeling (guiding, one-Dimensional trapping) of neutral atoms in laser standing wave with elements of atomic optics: reflection and focusing [36].

the TEM_{01}^{*} mode and the large frequency detuning make it possible to achieve a good approximation to an "ideal" objective lens (for details, see Section 6).

Another light-field configuration in the form of the so-called evanescent wave was proposed [86] to serve as a light mirror for atoms (Fig. 13(a)). This thin (surface) wave is formed upon the total internal reflection of a laser beam at a dielectric-vacuum interface. The thickness of this surface wave is of the order of the wavelength of light. The laser radiation intensity on the surface of the dielectric equals the intensity of the initial laser wave inside the dielectric, but in the vacuum, it abruptly drips to zero within a few wavelengths from the surface. In such a surface wave, one has an enormous light-intensity gradient—the greatest gradient seen anywhere in optics. An atom placed in such a laser surface wave of frequency higher than the atomic transition frequency experiences a gradient force tending to expel it from the wave into the vacuum. If the atom approaches the surface wave from the vacuum side, it slows abruptly down to a zero normal velocity in the surface wave and then accelerates in the wave in the opposite direction away from the surface. Using a two-frequency evanescent wave, one can effect the 1D-trapping of atoms [87], as illustrated in Fig. 13(b).

Evanescent waves can exist not only at a flat interface of two media but also in wave-guide structures. This effect can help effect the transportation (guiding) of atoms in a hollow fiber [88] (Fig. 13[c] and 13[d]). All these atom optical effects will be considered in more detail in Section 6.

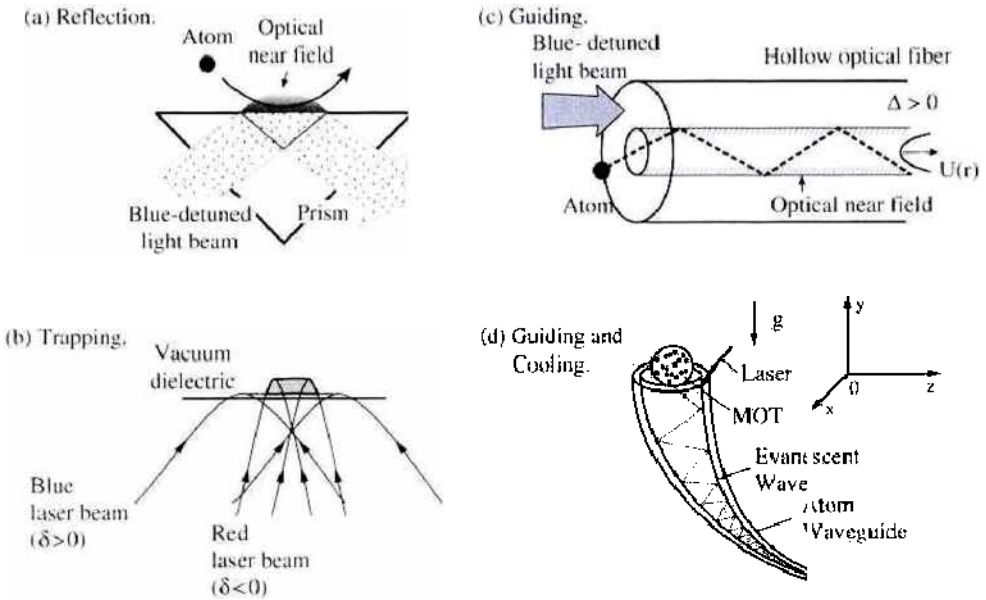


Figure 13. Atom optics experiments with evanescent laser wave.

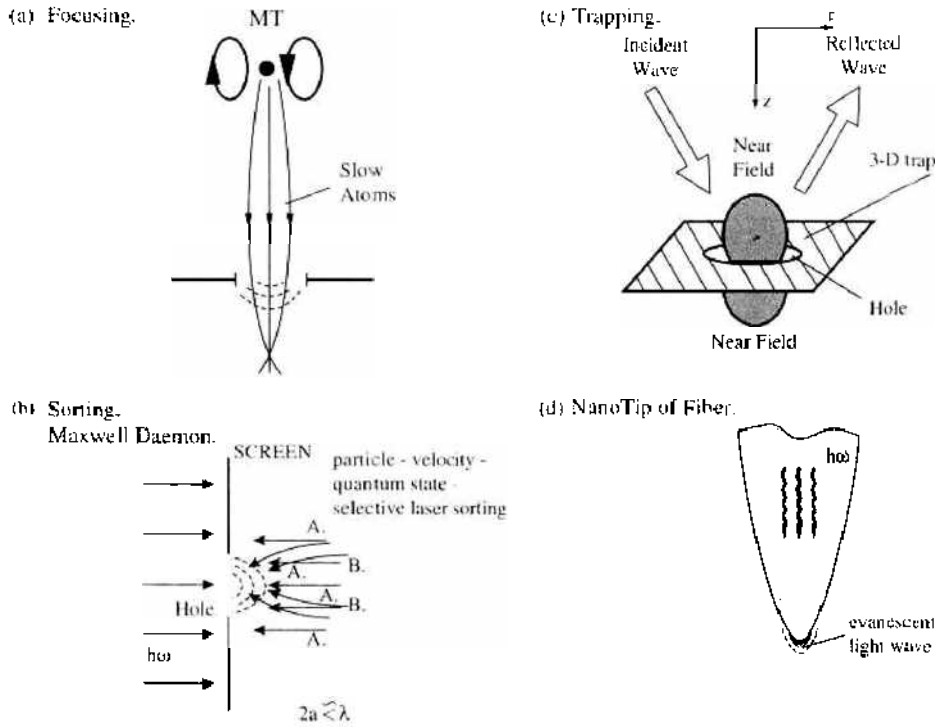


Figure 14. Atomic optics experiments in the near laser field.

One more light-field configuration with submicron inhomogeneities is associated with the "near field" effect arising when a light field "leaks" through an aperture $\ll\lambda$ in diameter. Such a field has a perceptible intensity at distances of the order of the diameter of the aperture from it and then dies out very rapidly. This effect formed the basis of what is known as "near-field optics" that, in contrast to "far-field" focusing optics is capable of a spatial resolution as high as 30–100 nm [89, 90]. The spatially inhomogeneous light near field is also of great interest to atom nanooptics [91], which is illustrated by Fig. 14. Light near fields can be used for the focusing of atoms (Fig. 14[a]), trapping of atoms (Fig. 14[c]) and for selective sorting of various atomic species (Maxwell demon) [92], as shown in Fig. 14[b]. Moreover, a near field issuing from an aperture into a fiber nanotip (Fig. 14[d]) can be used to manipulate single atoms. The guiding of atoms is considered in Section 5, and their focusing in near field, in Section 6.

5. GUIDING OF ATOMS

The guiding of atoms can to a large measure be effected in the same way as that of photons in an optical waveguide. But serving as waveguides for atoms are static electric, magnetic, and high-frequency electromagnetic fields and their combinations of various one- and two-dimensional configurations.

5.1. Guiding of Atoms with Electric Fields

Static electric fields were successfully used to focus polar molecules with an induced electric dipole moment [12]. But is it possible to use static electric fields to guide atoms and molecules? Let us consider the physics of the spatial localization of neutral atoms. For one to hold a neutral particle in static equilibrium, one should satisfy the following two conditions. First, the holding force should be identically equal to zero at a certain point r_0 in space:

$$\mathbf{F}(\mathbf{r}_0) = 0 \tag{40}$$

Second, the force should bring the particle back to the point of equilibrium, r_0. Mathematically, the second condition means that the partial derivatives $\partial F_x/\partial x$, $\partial F_y/\partial y$, and $\partial F_z/\partial z$ should be negative:

$$\nabla \mathbf{F} < 0 \tag{41}$$

To guide neutral particles in a specified direction (along the z-axis, for example), the holding force should meet not so stringent requirements: it is only two partial derivatives, namely, $\partial F_x/\partial x$ and $\partial F_y/\partial y$, that should be negative.

The potential energy of a neutral particle in a static electric field is defined by the expression

$$V_L = -\frac{1}{2}\alpha|E|^2 \tag{42}$$

where α is the polarizability of the particle and E is the strength of the external electric field. In a space free of charge, the Maxwell equation for the electric field is

$$\nabla \mathbf{E} = 0 \tag{43}$$

and the field can be expressed in terms of the scalar potential ϕ alone as

$$E = -\nabla\phi \tag{44}$$

Since the force acting on the particle is governed by the potential energy gradient V, the components of this force are given by

$$F_j = \sum_i \left(\frac{\partial\phi}{\partial x_i}\right)\left(\frac{\partial^2\phi}{\partial x_i \partial x_j}\right) \tag{45}$$

Using expression (45) for the force components, one may write the necessary condition (41) in the form

$$\nabla \mathbf{F} = \alpha \sum_{ij} \left(\frac{\partial^2\phi}{\partial x_i \partial x_j}\right)^2 < 0 \tag{46}$$

Inasmuch as the sum of the squared partial derivatives in expression (46) is a positive quantity, condition (46) is only satisfied for particles of negative polarizability.

For polar molecules, condition (41) can easily be satisfied by selecting their suitable rotational quantum states. Practically such molecules are guided by means of a four- or a six-pole electrostatic field. The strength of the field in the quadrupole configuration is defined by the expression

$$|E| = \left(\frac{U_0}{r_0^2}\right) r \tag{47}$$

where U_0 is the voltage applied across the conductors, r_0 is the distance between the surface with the potential U_0 and the center of the multipole, and r is the distance from the multipole axis. The strength of the field in this configuration increases as the distance from the multipole axis increases, and so a particle placed inside the multipole is pulled toward its axis.

The polarizability of neutral atoms is positive, so that finite motion condition (41) is not satisfied. Nevertheless, atoms can be guided in quasi-static electric fields by way of their so-called dynamical focusing. There are several dynamical focusing schemes. The best-known scheme is based on the use of alternating sections of two-wire lines (charge elements), wherein the planes of the adjacent two-wire charge elements are tuned through 90° relative to each other. A potential is applied across the wires and the atoms move along the axis between them. By virtue of condition (41), in the plane of the wires the atoms are acted upon by a force that tends to push them away from the axis of the wires, and in the perpendicular plane, by a force that tends to pull them to the axis. Near the axis, the force constants for the atoms are defined by the expression

$$K_x = -K_y = \frac{2E_0\alpha}{d^2}, \quad K_z = 0 \tag{48}$$

where d is the distance between the wires and E_0 is the electric field strength at the z-axis. While moving through the successively alternating pairs of wires, an atom alternately

experiences the action of the pulling and the expulsive force in the XZ- and the YZ-plane. The atomic trajectory in such a system is similar to the trajectory of a light beam passing through a sequence of positive and negative lenses with equal focal lengths. Such a scheme was used for focusing and guiding a beam of ultracold neon atoms [93]. The atomic beam passed through a set of three electrostatic lenses. Each lens consisted of two parallel cylinders that functioned as a concave lens along one axis and as a convex lens along the other axis.

An alternative approach to the dynamical guiding of atoms is the use of a quadrupole of two pairs of wires with a time-variable voltage applied across them, the phase shift between the pairs of wires being equal to 90° [94].

One more approach to the dynamical guiding of atoms is based on their interaction with the electric field of a charged wire. The atomic interaction potential has the following form:

$$V_a = -2\alpha\frac{q^2}{r^2} \tag{49}$$

where q is the wire charge per unit length. The interaction potential of Eq. (49) is attractive and has the form $1/r^2$. No stable localization of atoms is possible in such a potential: the classical trajectory of an atom moving around the wire comes to an end either on the surface of the wire or at infinity [95–101]. The motion of an atom around a charged wire can be stabilized by using a charge oscillating in time [102]. A stable motion of an atom about a charged wire can be attained by producing a repulsive potential along the wire surface. Folman and co-workers [103] suggested using as such a repulsive potential the dipole potential of an evanescent wave with the field frequency detuned toward the blue side relative to the atomic transition frequency. This can be realized by replacing the charged wire with an optical waveguide coated with a thin layer of a conductive material.

5.2. Magnetic Waveguides

All static magnetic waveguides are based on the localization of atoms by inhomogeneous stationary magnetic fields. The potential of interaction between an atom and a static magnetic field **B** has the form

$$V_M = -\mu B = -g_F\mu_B m_F B \tag{50}$$

where $\boldsymbol{\mu} = g_F\mu_B\mathbf{F}$ is the magnetic moment of the atom, g_F is the Lande factor of the atomic hyperfine state F, μ_B is the Bohr magneton, and m_F is the magnetic quantum number. In the inhomogeneous magnetic field $\mathbf{B} = \mathbf{B}(\mathbf{r})$, an atom with a constant magnetic moment of μ is acted upon by the magnetic dipole force

$$\mathbf{F} = \nabla(\boldsymbol{\mu}\mathbf{B}) = \mu_t\nabla B \tag{51}$$

where μ_t is the projection of the magnetic moment onto the direction of the field. When the adiabatic approximation condition is satisfied, that is, when the Larmor precession frequency $\omega_L = \mu_B B/\hbar$ of the magnetic moment is higher than the variation rate of the magnetic field direction, the interaction potential is proportional to the magnetic field magnitude $|\mathbf{B}|$. Depending on the orientation of the magnetic moment relative to the magnetic field direction, the interaction potential can be either positive or negative. If the magnetic moment is directed with the magnetic field, the interaction potential $V_M < 0$ and the atom is pulled into the strong-field region. The correspondent quantum state of the atom has come to be known as the strong-field seeking state. The potential energy minimum corresponds to the magnetic field maximum. The strong-field seeking state is the lowermost energy state of the atom-field system.

If the direction of the magnetic moment of the atom is opposite to that of the magnetic field, the interaction energy $V_M > 0$. In that case, the quantum state of the atom is called the weak-field seeking state, and the atom is expelled from the strong-field region. The weak-field seeking state is not the lowermost energy state of the atom-field system. Since the Maxwell equations allow no static field maximum in free space, force Eq. (51) can only be used to guide atoms about the minimum of a static magnetic field [104]. The atoms in that

case reside in a weak-field seeking state that is not the lowermost energy state of the atom-field system. The behavior of an atom in a magnetic waveguide depends substantially on the magnitude of the magnetic field minimum B_{min}. Magnetic waveguides wherein $B_{min} = 0$ have a simple geometrical shape and contain relatively simple, highly symmetrical magnetic fields. In such traps, however, a material negative role is played by the nonadiabatic Majorana transitions in the zero-field region that change the mutual orientation of the atomic magnetic moment and the magnetic field and thus cause the atoms to leave the waveguide. This escape mechanism usually limits the channeling time of the atoms to a few seconds. Magnetic waveguides wherein $B_{min} \neq 0$ are free from this disadvantage, but their geometrical shape is complex and they contain fairly intricate asymmetric fields. As an example, the solid lines in Fig. 15 show the lower magnetic sublevels of the ^{23}Na atom, in which the atom can be trapped at the minimum of a static magnetic field.

At the usual value of the atomic magnetic moment, approximately equal to the Bohr magneton ($\mu \approx \mu_B$), and a moderate magnetic field strength ($B = 100$ G) at the edges of a magnetic waveguide, the latter can hold atoms with a temperature of the order of 10 mK.

5.2.1. One-Dimensional Magnetic Waveguides

5.2.1.1. Single Current-Carrying Wire A most simple waveguide configuration for guiding and trapping atoms is a single current-carrying wire. It was first suggested by Vladimirskii [95] for the localization of slow neutrons. The development of the laser techniques for cooling atoms (see the review papers by Metcalf and van der Straten [56] and Balykin et al. [105]) has made it possible to lower the kinetic energy of atoms to such values as allow their binding to current-carrying wires [95–97, 98–101, 106]. The potential of interaction between an atom and the magnetic field of a current-carrying wire has the form

$$V_M = -\boldsymbol{\mu}\mathbf{B} = -\left(\frac{\mu_0}{2\pi}\right)\left(\frac{I}{r}\right)(\mathbf{e}_\varphi \boldsymbol{\mu}) \tag{52}$$

where I is the current through the wire, r is the distance between the atom and the center of the wire, μ_0 is the permeability of vacuum, and $\mathbf{e}_\varphi$ is the azimuthal unit vector. When the atomic magnetic moment $\boldsymbol{\mu}$ is parallel to the direction of the magnetic field $\mathbf{B}$, the atom is pulled into the region of the field maximum (i.e., toward the wire surface). To hold the atom outside of the wire, it is necessary that an additional expulsive potential be present. The attractive potential depends on the coordinate r as $1/r$. It is well known that the centrifugal potential barrier V_L can compensate for all the regular potentials that diverge less rapidly

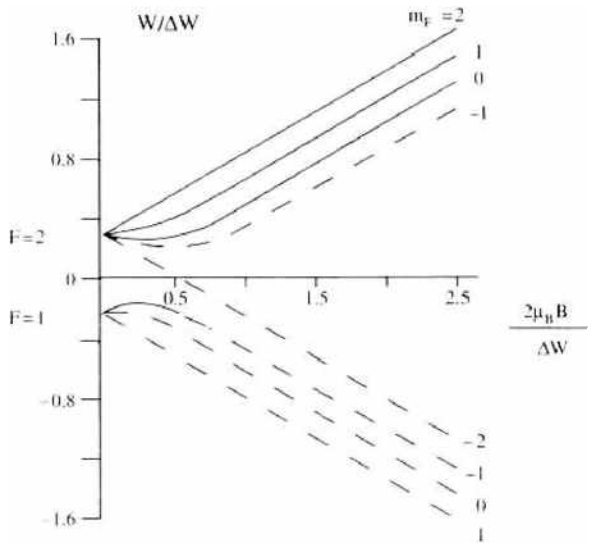


Figure 15. Hyperfine-structure magnetic sublevels for ^{23}Na as a function of the magnetic field strength. $\Delta W = 1772$ MHz is the ground-state hyperfine splitting. Solid lines correspond to the lower magnetic sublevels of the ^{23}Na atom at which it can be held at the minimum of a static magnetic field.

than $1/r^2$ when $r \to 0$ and can thus provide for a stable circular motion of the atom about the wire.

The atom + wire quantum system can be seen as a two-dimensional hydrogen atom. The wire resembles the "nucleus" and the atom plays the part of the "electron." The quantum mechanical properties of the atom + wire system were investigated in detail: its bound-state spectra were found and the lifetimes of its excited states determined [107–113].

The Keplerian guiding of atoms was for the first time realized with a beam of thermal Na atoms propagating along a meter-long current-carrying wire [97]. The guiding and trapping of atoms about a current-carrying wire were experimentally studied in detail using cold Li atoms [98–101]. The atoms were first cooled and localized in a magneto-optical trap (MOT) located at a distance of 1 mm from the wire 50 μm in diameter. A cloud of cold atoms was then liberated from the trap and reached the wire and at the same time a current of around 1 A was passed through it to produce an attractive potential for the atoms. The atoms caught performed circular motions about the wire while traveling along it in accordance with their initial velocities. Figure 16 schematically illustrates the localization and channeling of Li atoms along a current-carrying wire realized by Denschlag and co-workers.

A slight modification of the above scheme by the addition of a constant magnetic field B_c in a direction normal to the wire makes it possible to implement another scheme of guiding atoms along a current-carrying wire. The additional constant field cancels out the magnetic field of the wire along a line passing parallel to it at a distance of $r = (\mu_0/2\pi)(I/B_c)$. A potential energy minimum is poduced along this line for the atoms in the weak-field seeking state.

5.2.1.2. Two-Wire Guides The magnetic field produced by two wires carrying equal currents I of the same direction has zero strength along the central line between the wires:

$$B(x, y) \simeq 4\mu_0 I \frac{(y\hat{x} + x\hat{y})}{\pi d^2} \tag{53}$$

As follows from expression (50) for potential energy, the transverse guiding potential increases linearly with the applied current, and the given guide configuration can hold atoms along the central line. The distance between the wires, the current through them, and the axial velocity of atoms determine the minimum radius of curvature of the guide capable of holding atoms therein [100–101, 113]. With such a two-wire line, it proved possible to realize storage rings for cold atoms [114] (Fig. 17). The two-wire line is attractive from the standpoint of development of atomic waveguides on surfaces by lithographic techniques [103]. To avoid direct contact between the atoms and the surface, a constant magnetic field parallel to the plane of the wires is added to displace the potential minimum away from the surface.

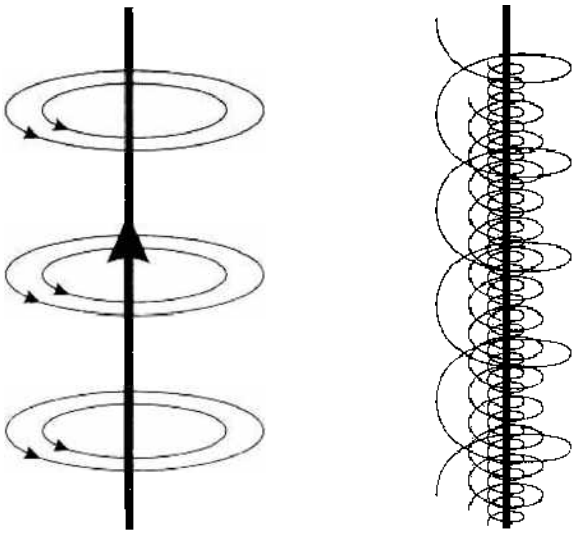


Figure 16. Guiding of atoms with a current-carrying wire. *Right*, the current-carrying wire producing a magnetic field; *left*, schematic illustration of the classical trajectories of an atom about a current-carrying wire.

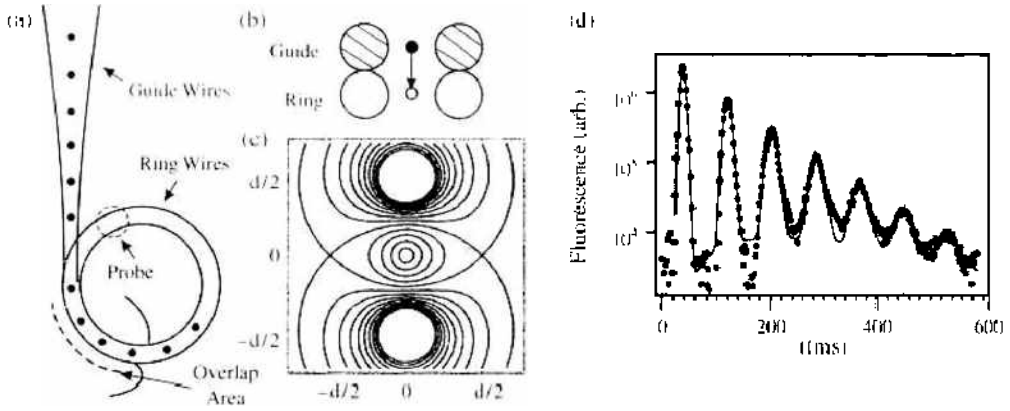


Figure 17. Guiding and trapping of neutral atoms with a two-wire configuration: (a) schematic diagrams of a waveguide and a storage ring for atoms; (b) cross section of the overlapping region between the waveguide and the storage ring; (c) experimental evidence of the successive revolutions of atoms in a storage ring. Courtesy of M. Chapman.

Even with moderate currents and constant magnetic field one can localize and transport atoms at a distance of a few micrometers from the surface [103, 115].

5.2.1.3. Four-Wire Guides The magnetic field produced by four wires equally spaced about a cylinder of radius R and alternately carrying the currents $+I$ and $-I$ can be approximated well enough by the following simple quadrupole formula:

$$B(r) = B_0 \frac{r}{R} \tag{54}$$

where $B_0 = 2\mu_0 I/\pi R$. Such a configuration (with six wires) was first used by Friedburg and Paul [24] for focusing thermal atomic beams. It was experimentally demonstrated [116] that this configuration could successfully be used to transport cold atoms. A cloud of cooled Rb atoms was coupled via a magnetic funnel into a miniature waveguide formed by four wires embedded in a silica fiber. The atomic cloud of 100 μm diameter could travel for a few centimeters inside of the waveguide.

This type of guide is being considered as a prototype of a single-mode waveguide for de Broglie waves. If the atoms of interest move in the magnetic waveguide field slowly enough (the adiabatic approximation), the effective atom-field interaction potential has the form

$$V(r) = V_0 \left(\frac{r}{R}\right)^{n-1} \tag{55}$$

where V_0 is a constant characterizing the atom-field interaction force and R is the waveguide radius. In a cylindrical waveguide, the center-of-mass motion of the atoms is described by the wave function $\Psi(r, \varphi, z)$ of the system. By virtue of the symmetry of potential Eq. (55), this motion can be resolved into its axial, azimuthal, and radial components [117] as follows:

$$\Psi(r, \varphi, z) = Y(r) e^{ikz} \tag{56}$$

The momentum of the atom along the waveguide, $\hbar k$, can assume any value, while its angular momentum $\hbar e_\varphi$ about the z-axis of the waveguide is quantized. The radial wave function satisfies the radial equation derived from the Schrödinger equation:

$$\left[-\frac{\hbar^2}{2m}\left(\frac{1}{r}\frac{d}{dr} + \frac{d^2}{d^2 r}\right) + \frac{\hbar^2 \lambda^2}{2mr^2} + V(r) - E\right] Y(r) = 0 \tag{57}$$

where m is the mass of the atom and E is the energy of its transverse motion. Upon substitution of $\tilde{r} = r/\eta$, $\tilde{E} = E/\varepsilon$, the above equation assumes the following dimensionless form:

$$\tilde{r}^2 Y'' + \tilde{r} Y' - (e^2 + \tilde{r}^{n+1} - \tilde{r}^2 \tilde{E}) Y = 0 \tag{58}$$

where η is the characteristic length comparable with the width of the ground-state wave function

$$\eta = \left(\frac{1}{V_0}\frac{\hbar^2}{2mR^2}\right)^{1/n+1} R \tag{59}$$

and ε is the characteristic energy commensurable with the ground-state energy

$$\varepsilon = \left(\frac{1}{V_0}\frac{\hbar^2}{2mR^2}\right)^{(n-1)/(n+1)} V_0 \tag{60}$$

To achieve a single-mode propagation regime for the atoms inside the waveguide, their thermal energy $k_B T$ should be of the same order of magnitude as the characteristic energy ε for the quantum modes, defined by expression (60).

A single-mode atom waveguide can be achieved when the thermal energy of atoms $k_B T$ is of the order of the characteristic energy for the quantum modes ε Eq. (60). Considering relations (55) and (60), we have the following requirement for the current to be carried by the single-mode waveguide:

$$I = \frac{\pi\sqrt{m}}{\sqrt{2}\mu\mu_0\hbar} R^2 (k_B T)^{3/2} \tag{61}$$

One can readily see from relation (61) that given realistic values of the current through the waveguide, its single-mode regime can only be attained if its radius is extremely small. To illustrate, even for a waveguide with a radius as small as a mere 10 μm the required current is of the order of 0.1 A [117].

5.2.2. Two-Dimensional Waveguides

5.2.2.1. Zeeman Effect Surface Trap (ZEST) One of the plane waveguide schemes using a magnetic field is based on a modification of a magnetic mirror [118]. A magnetic mirror can be formed by an exponentially decaying magnetic field produced by a grid of permanent magnets with a periodically varying magnetization direction [34]. A magnetostatic mirror of this type was already used earlier to reflect cold neutrons [95].

Figure 18 presents a schematic diagram of a magnetostatic mirror for cold atoms. The magnetic field of the mirror is produced by a grid of magnets having their length along the x-axis much greater than the grid period a and a variable magnetization of $M(y, z) = \pm M_0$. Away from the surface of the magnets, $z \gg a/2\pi$, the magnetic field may be represented in the form

$$B(y, z) = B_0 e^{-kz}\left(1 + \frac{1}{3}e^{-2kz}\cos 2ky + \cdots\right) \tag{62}$$

where $B_0 = 2\mu_0 M_0/\pi$, $k = 2\pi/a$. When the atom resides at the hyperfine structure magnetic sublevel m_F with the total momentum F, the atom-field interaction energy is

$$U(y, z) = m_F g_F \mu_B B(y, z) \tag{63}$$

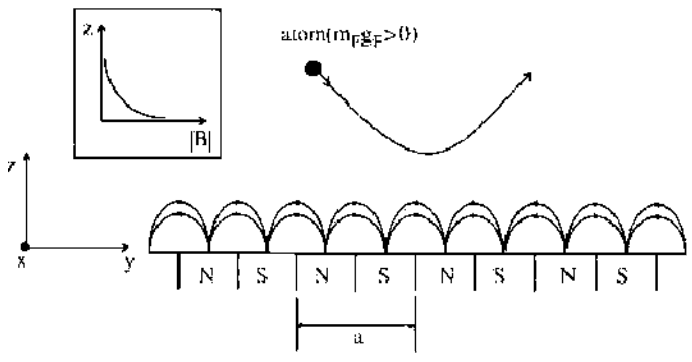


Figure 18. The Schematic diagram of magnetostatic mirror for cold atoms.

where g_f is the g-factor and μ_B is the Bohr magneton. For an atom in a "positive" magnetic state, where $m_f g_f > 0$, the energy of the atom decreases as it moves farther away from the surface of the magnets. This means that when the kinetic energy of atoms is lower than the barrier height, potential Eq. (63) reflects the particles incident thereon.

For slow atoms whose magnetic moment adiabatically follows the varying magnetic field, the quality of the magnetic mirror based on potential Eq. (63) is mainly governed by the ratio between the first two terms in relation (62). The second term in this relation that depends on the y-coordinate causes distortion of the atomic wave front, hence the diffuse scattering of the reflected atomic wave. At the same time, this term decay much faster than the first term in relation (62). In this connection, its contribution is essential for atoms penetrating deep into the magnetic field and is less important for slow, cold atoms.

The practical realization of the magnetostatic mirror can be exemplified by a device assembled of rare-earth neodymium-iron-boron permanent dipole magnets 1.04 mm wide, 20 mm long, and 12 mm thick. A set of 18 such magnets produced a field of form Eq. (62) with a strength of $B_0 = 4200$ Gs and a characteristic decay length of $a/2\pi = 0.33$ mm [119]. This magnetic mirror was used to reflect cold cesium atoms residing at the magnetic sublevel $F = 4$, $m = 4$ of the ground state $6^2S_{1/2}$. With atoms at a temperature of 20 μK, the experimentally measured reflectivity of the mirror proved close to 100%.

Another closely similar approach to the development of a magnetic mirror for cold atoms is to use a periodically magnetized ferromagnetic surface that reflects atoms owing to its magnetic potential. The experimental testing of this approach demonstrated that the ferromagnetic mirror could also reflect practically 100% of the incident atoms [120].

The successful implementation of the above two types of magnetic mirrors made it possible to design gravitational-magnetic traps, resonant cavities, and waveguides for cold atoms [121, 122].

Figure 19 presents a schematic diagram of a plane waveguide relying for its operation on a magnetic mirror [118]. The inset shows the Zeeman shift of the ground-state hyperfine levels as a function of the magnetic field strength, hence the distance from the surface of the magnetic mirror. As can be seen, the atom in the state f has its minimal potential energy at a certain distance from the surface; that is, the atoms in this state can be localized over the surface. Hinds and co-workers [118] proposed a scheme for loading atoms into such a plane waveguide: atoms in the state i move toward the mirror in its repulsive potential that slows them down. At the maximum point of the mirror potential, which corresponds to the minimum point of the potential energy of the atom at the sublevel f, laser radiation moves the atom via the intermediate level j to the level f at which it finds itself localized. With realistic magnetic mirror parameters (surface magnetization $B_1 = 2$ kGs, period $a = 1$ μm),

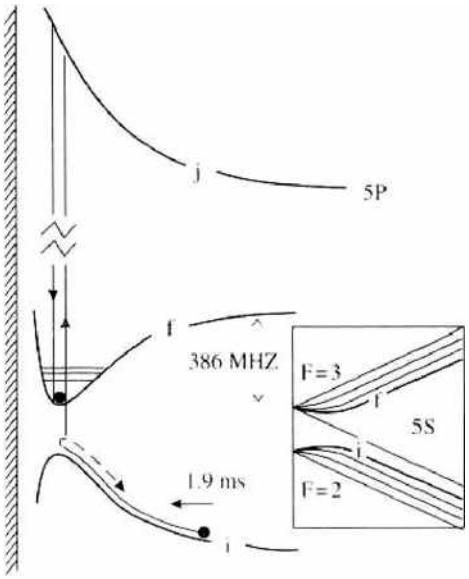


Figure 19. Scheme of loading of atoms a two-dimensional magnetic waveguide.

atoms can be localized over a plane at a distance of 162 nm from the mirror surface. In that case, the width of the center-of-mass wave function of the atom amounts to some 8 nm.

5.2.2.2. Combined Electric-Magnetic Waveguides The polarizability of neutral atoms is always positive, and so they are pulled into the strong electric field region. This circumstance can be used to develop a plane electric-magnetic waveguide.

A simple realization of this idea is to apply onto a surface an array of parallel wires alternately charged to a potential of $\pm\phi$. In that case, the energy of an atom in the electric field over such a surface is

$$U_E = -\frac{1}{2}\alpha K_E^2 \varphi^2 e^{-2K_E Y} \tag{64}$$

where $K_E = 2\pi/a$, a being the period of the array, and Y is the distance between the atom and the surface. For particles of negative polarizability (for example, polar molecules in a suitable rotational quantum state), such a periodic structure on surface is a plane mirror (an analog of a magnetic mirror). But atoms (particles of positive polarizability) are attracted to such a surface. A combined electric-magnetic waveguide consists of a magnetic mirror with an array of charged wires on its surface. Cold atoms are magnetically repelled from the surface of the mirror and are then strongly attracted to it by its electric field. Such a two-dimensional localization of atoms can be realized if the range $1/K_M$ of their magnetic repulsion is shorter than the range $1/K_E$ of their electric attraction. Electrostatic forces can be used to design both two- and one-dimensional waveguides.

Imagine a magnetic mirror with a thin wire carrying a charge of λ per unit length on its surface. The field over the wire is $\lambda/(2\pi\varepsilon_0\rho)$, where ρ is the distance from the wire. And the total potential over the surface will be [123]

$$U = -\mu B_1 e^{-K_M Y} - \frac{\alpha\lambda^2}{8\pi^2\varepsilon_0^2\rho^2} \tag{65}$$

Figure 20 shows the behavior of potential Eq. (65) directly above the wire for the Rb atom as a function of the distance from the surface. The magnetic mirror has a surface field strength of 1.1. kGs and a period of 10 μm, the charge of the wire being equal to 1 nC $\cdot$ m^{-1}. The profiles in Fig. 20b show the guiding tube formed along the wire at a distance of 9 μm from the surface.

Miniaturizing the current- and charge-carrying structures used to confine cold atoms offers prospects for a finer control over atomic ensembles [124, 125]. Following the successful trapping and guiding of thermal atoms by means of miniature self-supported wires [98–101, 116, 126, 127] and substrate-supported microfabricated wire arrays, the recent experiments [128–130] demonstrated wire traps and waveguides with Bose-Einstein condensation.

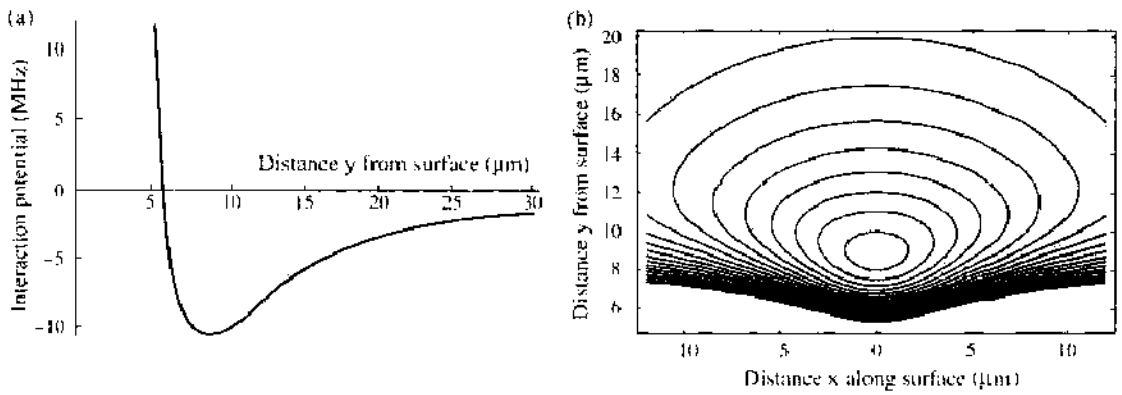


Figure 20. Potential of combined electric-magnetic waveguide.

5.3. Atom Guiding with Laser Light

The advent of lasers not only stimulated the development of an optical fiber but also gave a creative impulse for a guiding of free atoms. Probably the first proposed scheme to guide atoms was the scheme of guiding of atoms in a standing wave [36]. Later several groups realized this scheme of atom guiding, which is presently known as the channeling of atoms in a standing wave [131–133]. The Gaussian laser beam was the first laser field configuration that attracted a great deal of attention for focusing and guiding of atoms. In 1978 Bjorkholm et al. [37] demonstrated the focusing of an atomic beam that was propagating coaxially with the Gaussian laser beam. Actually, this pioneering in atom optics experiment could be considered to a great extent as the first guiding experiments with laser light.

In 1992, Ol'shanii et al. [88] proposed combining two experimental techniques: the guiding of the radiation itself in a hollow fiber and the guiding of the atomic beam in an optical fiber. Savage et al. [134], Marksteiner et al. [135], Burke et al. [136] developed the extended theory of the motion of atoms in optical fibers. The first successful experiment of the guiding of atoms in a hollow fiber was demonstrated by a Colorado group [137]. The more promising concept of atom guiding is based on the use of an evanescent wave atom mirror [86]. Since the time of the first demonstration of the atom reflection [138], the atom mirror was extensively studied (see the review papers by Dowling and Gea-Banacloche [139] and Balykin [140]). The same Colorado group [141] and the Japanese-Korean group [142–145] have successfully demonstrated the guidance of atoms by using optical near fields. Guidance of atoms by optical near fields is complicated by two processes: the diffusive scattering of laser light on the dielectric surface [151] and the attractive van der Waals force between the dielectric wall and the guided atom [146]. Guidance of atoms with a propagating "dark spot laser beam" is free from these limitations [147, 148] at the same time a spatial "rigidity" of the laser beam limits applicability of such kinds of atom waveguide.

5.3.1. Atom Waveguide with Propagating Laser Fields

In this section we consider the different configurations of propagating laser fields for the guiding of atoms.

5.3.1.1. Gaussian Laser Beam The Gaussian laser beam can create the transverse confinement of atoms and does not restrict the motion of atoms along its axis. The Gaussian laser beam has the following z-dependent beam radius (z along the laser beam)

$$w(z) = w_0\sqrt{1 + \left(\frac{z}{z_R}\right)^2} \tag{66}$$

where w_0 is the minimum spot size of the laser beam, and $z_R = (\pi w_0^2/\lambda)$ is the Rayleigh range. If we consider the Gaussian laser beam as an atom waveguide, then the Rayleigh range determines the effective length of such a waveguide. For a single-mode atom guiding the internal radius of a waveguide has to be in diameter about several laser light wavelengths. At the minimum spot size of a laser beam $w_0 = 2\lambda$, the Rayleigh range is a very short and equals only $z_R \approx 4\pi\lambda$. It means that the Gaussian laser beam can be considered only as a multimode atom waveguide.

Another limitation of the Gaussian laser beam as an atom waveguide is due to the fact that the maximum field intensity is on the beam axis. In the guiding of atoms by the Gaussian laser beam, they are localized predominantly in the region of high intensity where the impulse diffusion is the most severe.

In 1978 Bjorkholm et al. [37] demonstrated the focusing of an atomic beam that was propagating coaxially with the Gaussian laser beam. In their experiment, the light from a single-frequency laser was superimposed upon a beam of sodium atoms. For the used in the experiment negative detuning, when the laser frequency is smaller than atomic absorption frequency, the force on the atoms is directed toward higher intensity, and the atoms are attracted into the center of the laser beam. The atoms did not experience periodic focusing and defocusing (which, actually, means a guiding) only because of an optical pumping of

sodium atoms: The duration of resonant interaction of atoms with a single-frequency mode laser was less than the time of flight atoms through the superimposed laser beam.

The severe limitation imposed by the impulse diffusion was overcome with a development of a powerful Nd : YAG. Such laser has made it possible to guide atoms with a large laser detuning when the impulse diffusion is strongly suppressed. In the experiment Prusot et al. [149] it was reported on realization of a Nd : YAG laser atomic guide for cold rubidium atoms. The laser frequency is so far from atomic rubidium resonance ($1.6 * 10^7\gamma$) that emission rate is very small, the guide has no losses, and the atoms can be guided a long a large distance (~30 cm). A similar guide was used in a cold atom fountain [150].

Another limitation of the atom guiding based on the Gaussian laser beam (and all other schemes with the propagating laser beams) comes from a physical impossibility of "bending" the Gaussian beam, which closes many potential applications of such an atom waveguide. All these limitations of the Gaussian laser beam as an atom waveguide push the numerous research groups to explore other laser-field configurations.

5.3.1.2. Atom Waveguide with Propagating Light Wave Inside of Hollow Fiber Laser light itself can be easily and efficiently guided by the different types of optical fibers. A confinement of laser light inside a hollow fiber immediately solves two problems of atom guiding by the Gaussian laser beam: (1) the laser light can propagate without transverse spreading over a long distance; (2) inside an optical fiber the laser field can be "bent" almost without limitation. The first scheme for guiding of atoms with laser light inside a hollow fiber is proposed by Ol'shanii et al. [88] (Fig. 21). In their scheme, the light propagating along a hollow optical fiber is the lowest order propagating mode EH_{11}. The transverse electrical field profiles of the mode EH_{11} is a zero-order Bessel function:

$$\mathbf{E}(\rho, z, t) = \mathbf{e}E_0 J_0(\chi\rho)e^{i(\beta z - \omega t)} \tag{67}$$

and the intensity transverse profile is

$$I(\rho) = I_0 J_0^2(\chi\rho) \tag{68}$$

with a maximum intensity along the fiber axis. The mode EH_{11} can be excited in the fiber by a laser beam that propagates along the axis of the fiber [137]. With a red-detuned laser frequency, atoms are attracted to the high-intensity region along the fiber axis, as in the case of the Gaussian laser beam. This means that the scheme still has the same impulse diffusion problems as in the case of the Gaussian laser beam: atoms are localized predominantly in the region of high-laser intensity, where the impulse diffusion leads to their heating and, finally, to sticking of atoms to the inner fiber wall. The impulse diffusion limits atom guiding time and effective guiding length of this type of atom waveguide [88]. The atom guiding time is just equal to the time it takes for atom to enlarge its transverse kinetic energy to the value of the maximum potential depth $V_{\max}$ of the guiding laser field

$$\tau = \frac{mV_{\max}}{D} \tag{69}$$

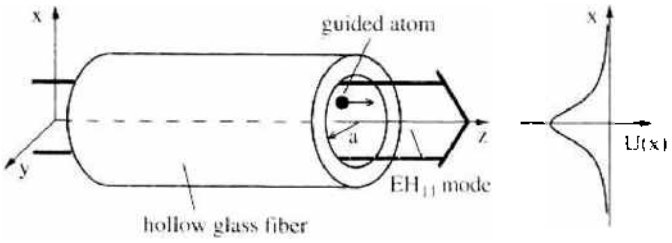


Figure 21. Scheme of an atom waveguide based on the optical mode EH_{11} propagating in the dielectric waveguide (*left*) and the optical potential for atoms (*right*) [88].

where D is the diffusion coefficient. For the large laser detuning ($\delta \gg \gamma$) the diffusion coefficient can expressed as (Ol'shanii et al. 1993)

$$D = \frac{\hbar^2 k^2 \gamma}{20} s \tag{70}$$

where $s = 2\Omega_R^2(\rho)/(\gamma^2 + 4\Delta^2)$ is the saturation parameter of the atomic transition. For guiding of thermal atoms with an average velocity $\langle v_z \rangle \approx 5 \cdot 10^4$ cm/s the corresponding atom guiding time, according to Eq. (69), is $\tau = 10^{-2}$–10^{-1} s. The corresponding guiding length is $L = \langle v_z \rangle \tau = 5 \cdot (10^2$–$10^3)$ cm. For the guidance of ultracold atoms the heating mechanism limits the guiding length to a fraction of a centimeter and in turn it limits the applicability of such a type of waveguide.

Another serious drawback of the hollow fiber atom waveguide is that these modes decay exponentially themselves as they propagate along the fiber. The propagation constant of the modes $\beta = \beta' + i\beta''$ has an imaginary part β'' which in an approximation of multimode fiber $ka \gg 1$ has the form [137]

$$\beta'' \propto a^3 \tag{71}$$

Because the attenuation length decreases as the cube of the hole radius a, even for a relatively large (40 μm) fiber diameter (used in the Renn et al. (1995) experiment) the corresponding propagation attenuation length is only $1/\beta'' \approx 6.2$ cm. The attenuation of a propagating fiber mode limits the useful length of a fiber for atom guiding. For the curved fiber the imaginary part of the propagation constant has an additional term that depends on the radius of curvature of the fiber and in turn it makes the guiding potential of an asymmetric elliptical shape.

In the JILA experiment [137, 141] to generate an atom guiding laser mode EH_{11}, the laser beam was launched into the hollow region of a glass capillary. The laser light is coupled into various modes and propagates along the fiber by grazing-incidence reflection from the glass wall. In the experiment, the capillary fiber has an outer diameter of 144 μm and hollow core diameter of 40 μm. The propagation attenuation length, for a chosen diameter fiber and at the wavelength of rubidium transition 780 nm, is only 6.2 cm. The attenuation length limits the guiding distance and the fiber length used was from 3 to 15 cm. The coupling efficiency into the lowest order EH_{11} mode was around 50% and it was achieved when the laser beam waist at the entrance of the fiber is approximately the size of EH_{11} mode, and the axes of the laser beam and the fiber coincide. The EH_{11} mode diameter is substantially smaller than the diameter of the core. It means that the guided atoms are localized in a transverse direction to a size considerably smaller than the internal diameter of the fiber and, as consequence, the effect of van der Waals interaction and quantum tunneling can be ignored in this type of atom waveguide.

In the experiment, the fiber connects two vacuum chamber: the first one, the source chamber, contains rubidium vapor with a partial pressure of $\sim 10^{-6}$ torr. Atoms with small transverse velocities and appropriate trajectories pass into the fiber and are guided through the fiber into a second detection chamber. At the exit of the fiber, the atoms are ionized and the ions detected with a channeltron electron multiplier. The main experimental data are described satisfactorily by the model based on the dipole potential: atoms were guided when the laser frequency was red detuned to the atom transition and the effective potential is in attractive one. For a blue detuning of the laser frequency there was no guiding: the efective potential is a repulsive one.

Figure 22 shows the intensity dependence for an atom flux for a red detuning of the laser of $\Delta = -8$ GHz from atom resonance. At low intensity, the flux increases linearly with laser intensity, as expected for guiding atoms in the dipole potential. However, at high-laser intensity the guided flux is decreased as a result of the action of the dissipative force and the impulse diffusion.

Renn et al. [137] showed that atoms may be guided also in curved fiber. In a curved fiber atom guiding is complicated by several additional effects: (1) bending of a fiber alters

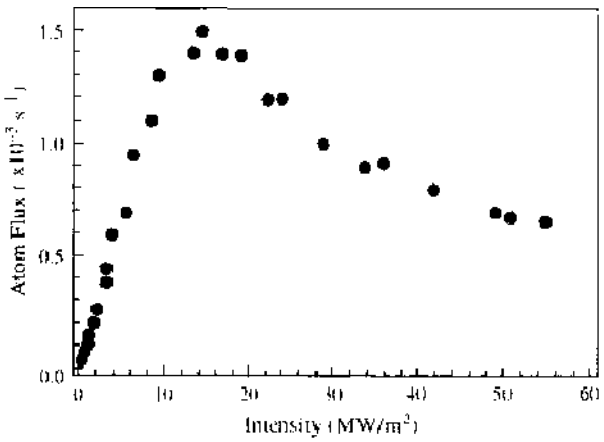


Figure 22. Atom guiding with propagating light wave in a hollow fiber. Figure shows the intensity dependence for an atom flux for a red-laser detuning of $\Delta = -8$ GHz from resonance [141].

the optical field distribution from the lowest-order grazing incidence EH_{11} mode into high-order modes; (2) there is an additional centrifugal force acting on atoms and pushing atoms into the wall; (3) for atom guiding in the bent fiber it is necessary to introduce larger laser intensity, and as a result, the optical pumping to other hyperfine sublevels starts to play a significant role. In the Renn et al. [137] experiment the atom guiding was detected at the minimum bend radius $R = 5$ cm.

5.3.1.3. Dark Spot Laser Beams A simple circumvention of the serious diffusion problem in the atom guiding can be the use of a dark spot laser beam (DSLB): a laser beam with a minimum intensity in the beam center. The best-known example for this type of laser beam is a TEM_{01}^{*} beam (known also as a donut mode). A copropagating TEM_{01}^{*} beam was first considered as an atom focusing lens [85, 152, 153]. To realize a focusing and a guiding with a donut mode, the positive laser detuning has to be used, so the dipole force is directed toward the hollow center of the laser beam. In this type of atom lens and waveguides, atoms move predominantly through a relatively low-intensity region, where the rate of spontaneous emission is a minimum [93, 140].

Nowadays a rich variety of methods exist to create the DSLB, such as the transverse mode selection method [154], geometrical optical method [155], optical holographic methods [156], computer-generated hologram method [147, 157], the method of Gaussian mode conversion [158], and the method based on the use of hollow optical fibers.

We shall only mention some of them that found an application in atom optics and, particular, in atom guiding. In atom-guiding application of the DSLB, there are three most important parameters of the beam: (1) the smallest achievable dark-spot size; (2) the degree of diffraction efficiency of the first bright ring, and (3) the divergence of the beam. The current stage of DSLB research can be found in the review paper [159].

Mode conversion method. The simplest method of generating a DSLB is to place inside a laser resonator a small absorbing dot close to the optical axis of the resonator. Then the transverse intensity distribution of the laser output recalls the distribution close to the TEM_{01}^{*} mode. Unfortunately, the quality of the laser beam is rather pure.

If, instead of the absorbing dot, a thin wire is placed inside the resonator, then the laser can be forced to generate high-quality Hermite-Gaussian modes. These modes have nodal lines at the position of the wires. The next step is a conversion of the Hermite-Gaussian mode to a DSLB. Kuppers et al. [147] used the mode conversion method to focus and guide metastable Ne atoms. First, they generated a TEM_{01} laser mode by inserting a 20-μm-diameter wire into the ring dye laser. Then, by using a mode converter consisting of two cylindrical lens, the TEM_{01} mode was transformed into a donut mode without a significant loss of power.

Computer-generated hologram method. A hologram is, in essence, an interference pattern arising from a coherent reference wave and a second wave (usually scattering from the object). Commonly, the interference pattern is recorded by a photographic emulsion (and

very often it is called a hologram). During a *reconstruction* step, the reference field illuminates the hologram and the hologram regenerates the original light field. In the rare cases when a mathematical form of the second wave is known (in our particular case it is the TEM_{01}^{*} mode), it is possible to *calculate* the interference pattern (the hologram) and to omit the *recording* process.

The interference pattern of a hologram contains the sinusoidal variations in the field intensity. In a standard technique of a photographic recording of a hologram, the sinusoidal variation in the field will lead to sinusoidal variation in optical density on the photo plate. By using a computer graphic technique, it is possible to print out the *calculated hologram pattern*. However, in practice, it is much easier to print a *binary* hologram with a "square wave" transmission function, which incorporates the sinusoidal variation in the field intensity. The next step is a photo reduction of the binary hologram onto a photographic film. The result is the amplitude hologram, which always has rather low efficiency. To obtain a hologram with a high efficiency, the amplitude hologram has to be transformed into the phase hologram.

Heckenberg et al. [160] demonstrated a computer-generated hologram method to create the TEM_{01}^{*} mode. The Gaussian mode TEM_{00} is used as a reference field. The reconstructed light beam is not strictly TEM_{01}^{*} mode donut mode but in the far field the spatial-intensity-dependence equivalent to the donut mode. By using such a DSLB, He et al. (1995) demonstrated the trapping of refractive and absorptive microscopic-size particles in the dark central spot of the focused light beam. Kuppens et al. [147] have made the extensive studies of two methods of generating a dark spot-size light beam: (1) the conversion method and (2) the computer-generated-hologram method. The quality of the donut mode obtained in the computer-generated hologram method was higher than the quality of the mode generated by the conversion method.

Microcollimation technique. Yin et al. [148] proposed and demonstrated a relatively simple and very efficient method to generate a dark hollow laser beam with a radial intensity distribution similar to one of a donut mode laser beam. Their method is based on the use of the modes of a hollow optical fiber *outside* of the fiber. It is well known that the lower-order modes of a hollow fiber are so-called linear polarized LP_{01}, LP_{11}, LP_{21} modes. All these modes are already the dark spot modes but *inside* the fiber. The experiment of Yin et al. [148] demonstrated that it is possible to extract and preserve these modes outside of the fiber. The beam divergence at the near field was equal to $6.5 * 10^{-5}$ rad, whereas at the far field it was $4.0 * 10^{-4}$ rad. The main limitations on the output parameters of the dark spot beam come from the divergence on the core of the fiber.

Guiding of atoms with the donut mode configurations overcomes the diffusion problem, however, it suffers still from inevitable divergence of a free-propagating laser beam and its "rigidity" in a vacuum.

5.3.1.4. Atom Guiding with a Dark Sport Laser Beam The Bonn group [147] successfully demonstrated the guiding and focusing of metastable neon atoms with DSLB. The DSLB was created by two methods. In the first method, the lowest-order Hermite-Gaussian TEM_{01} mode was derived from a ring-dye laser by inserting a 20-μm-diameter wire into the laser cavity. A mode converter, consisting of two cylindrical lens, transforms the TEM_{01} mode into a DSLB that contains about 1-W laser power. The DSLB intensity profile has a slightly asymmetrical ring-shape form. An ultracold beam of metastable neon atoms was injected coaxially into the DSLB. The radius of the donut mode laser beam at the injection plane was 400 μm; at the distance of 20 cm after the injection plane, the donut mode waist was 100 μm. The slow atomic beam was prepared by Zeeman slowing of a thermal beam with a further compression and deflection by a two-dimensional magneto-optical molasses [161, 162]. The atomic beam prepared in this way had a sub-Doppler transverse temperature and its longitudinal velocity was 25 m/s with a 3 m/s velocity spread.

The spatial distribution of the guided atoms was detected at the waist of the donut mode. Without the guide the width of the atomic spatial distribution of the atomic beam was a 750 μm. The guided laser beam decreased the spatial size of the neon beam to the value of 17 μm. The peak intensity of guided atoms was increased by two orders of magnitude. The total flux of guided atoms contains 10% of the initial value.

5.3.2. Atom Waveguide with Evanescent Laser Fields

The use of a hollow optical fiber (HOF) with an evanescent wave on its inner surface for a guiding of atoms was first considered by Savage, Marksteiner, and Zoller [163]. Atom can propagate inside a hollow fiber provided that an evanescent surface field keeps atoms away from the inner wall, (Fig. 23). It was shown the feasibility of HOF as new element of atom optics. Later it was developed the theory of quantum motion of atoms in a hollow fiber and analyzed in detail the different mechanics losses in HOF [135, 136, 142, 164].

5.3.2.1. Evanescent Wave as an Atom Mirror Evanescent field above a surface can be used as a mirror to reflect atoms. The first scheme for atom mirror with an evanescent wave was suggested by Cook and Hill [86]. In their scheme, a plane traveling light wave is totally reflected internally at the surface of a dielectric in a vacuum: a thin evanescent light wave is generated on the surface. It is this surface wave that can be served as an atomic mirror for atoms running into it.

Simple evanescent wave. Suppose we have a laser light incident on a smooth interface of a dielectric in a vacuum and its electrical field has a form

$$E_i = E_{0i} e^{i(\omega t - \mathbf{k}_i \mathbf{r})} \tag{72}$$

The transmitted field through the interface is

$$E_t = E_{0t} e^{i(\omega t - \mathbf{k}_t \mathbf{r})} \tag{73}$$

For incident angles greater than so-called critical angle $\Theta_c (\sin \Theta_c = n$, n is a refractive index of the dielectric), there is a total internal reflection of the incident wave.

Using Shell's law we can write for the perpendicular k_{ty} and parallel k_{tx} components of wave vector of the transmitted wave:

$$k_{ty} = +ik_t \left(\frac{\sin^2 \Theta_i}{n} - 1 \right)^{1/2} \equiv i\beta \tag{74}$$

$$k_{tx} = \frac{k_t}{n} \sin \Theta_i \tag{75}$$

Then the electrical field of the transmitted field has the form

$$E_{ev} = E_t = E_{0t} e^{-\beta y} e^{i(\omega t - k_t x \sin \Theta_i / n)} = E_{0t} e^{-\beta y} \cos(\omega t - qx) \tag{76}$$

$$q = k_t \sin \Theta_i / n \tag{77}$$

and it is called the surface or evanescent wave. Its amplitude decays at a distance

$$\beta^{-1} = \omega (n^2 \sin^2 \Theta_c - 1)^{-1/2} / c \simeq \lambda / 2\pi \tag{78}$$

where λ is the wavelength of the incident wave, and c is the speed of light in vacuum. The evanescent wave propagates parallel to the dielectric vacuum interface with the wave

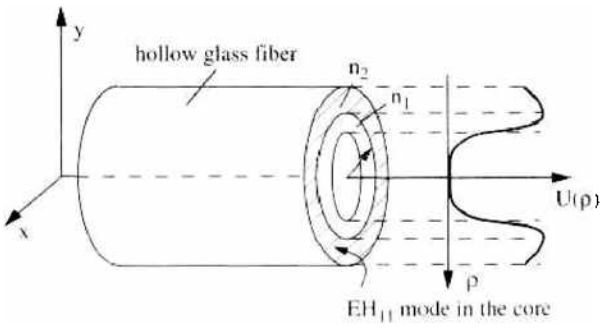


Figure 23. Atom waveguide based on hollow dielectric waveguide having two cylindrical dielectric sheaths differing in refractive index (*left*) and the optical potential (*right*).

vector of the evanescent wave $q = k_i \sin\Theta_i / n$. The amplitude of the evanescent wave at the dielectric interface is greater than the electrical field in the incident laser field E_{ins} (outside of the dielectric) and the relationship between them is given by the following equation [165, 166]:

$$E_{ev}^2 = E_{ins}^2 \left\{ \frac{4n^2 \cos^2\Theta_i}{(n^2-1)[(n^2+1)\sin^2\Theta + 1]^m} \right\} \tag{79}$$

where $m = 0$, 1 for TE or TM polarizations, respectively.

Surface plasmon-enhanced evanescent wave. Surface plasmons are electromagnetic charge-density waves propagating along a metal-dielectric interface. A metal-dielectric interface is prepared usually by evaporating a thin metal layer on the surface of a prism. For a surface plasmon, the wave vector component k_x along the interface is

$$k_x = \frac{\omega}{c}\sqrt{\frac{\varepsilon_1'}{\varepsilon_1' + 1}} \tag{80}$$

where $\varepsilon_1 = \varepsilon_1' + i\varepsilon_1''$ is the relative permittivity of the metal of the surface layer. Surface plasmons can be excited by the attenuated total reflection methods. In this method, a laser beam, polarized parallel to the plane of incidence, undergoes total internal reflection on the interface with the metal layer. If the wave vector k_{ix} of the evanescent wave propagating along the interface matches the wave vector k_x of the plasmon wave (Eq. [80]), then there is a resonant coupling into the plasmon wave.

An attractive feature of a plasmon wave in atom optics is the large-field enhancement of the initial laser beam intensity. The field enhancement factor η, defined as the ratio between the maximum intensity of the evanescent wave with and without plasmon excitation, is

$$\eta = \left(\frac{\varepsilon_1'^2}{2\varepsilon_2\varepsilon_1''}\right)\frac{\sqrt{-\varepsilon_1'(\varepsilon_2 - 1) - \varepsilon_2}}{1 - \varepsilon_1'} \tag{81}$$

A quartz prism ($n = \sqrt{\varepsilon_2} = 1.45$) with a silver layer ($\varepsilon_1' = -30$) and at optimum angle of incident ($\Theta_{i,opt} = 44.6°$) has the maximum enhancement factor between 15 and 100 [168].

There are a number of limitations in the use of the plasmon wave as an atom mirror. First, the field enhancement is limited by the surface imperfections such as metal surface corrugations. Second, the resonance condition of plasmon excitation depends on the thickness of the metal layer, which results in a thickness-dependence resonance of plasmon excitation. This could cause a strong fluctuation of the field intensity. Third, if the film layer on the prism surface is rather rough (larger than 5 nm), plasmon waves are strongly scattered. This scattering leads to accumulation of electromagnetic field density, which is locally higher than that of an extended surface plasmon wave on a smooth surface.

The surface plasmon technique has the advantage in its simplicity over other techniques of the enhancement of the evanescent wave [166, 167, 169].

Enhancement with dielectric waveguide. Another approach to enhancing an evanescent field is to use a multilayer dielectric structure on the prism surface [166, 170]. The dielectric structure consists of a layer with a refractive index n_1, which is separated from the prism surface by a small gap of low refractive index n_2. The layer with a refractive index n_1 plays a role of a thin-film waveguide. The incident laser light is totally reflected at the dielectric interface and the laser light in the prism and a mode in the waveguide are coupled through the evanescent field in the gap. When a mode of the waveguide is excited, even a moderate incident laser power gives a large light intensity in the waveguide.

The theoretically predicted enhancement factor is about several thousands [171, 172]. The realized enhancement of the evanescent wave was about 1700 [173].

Reflection of atoms by evanescent wave. If the evanescent wave is created by a total internal reflection of the Gaussian laser beam at a dielectric vacuum interface, then the electrical field in the vacuum above the surface takes the form

$$E(x, y, z) = E_0 \exp\left[-\beta z - \left(\frac{x}{w_x}\right)^2 - \left(\frac{y}{w_y}\right)^2\right] e^{-ik_x x} \tag{82}$$

where w_x and w_y are the beam waists along the surface and x and z are the coordinates parallel and perpendicular to the surface. The corresponding Rabi frequency

$$\Omega(t) = \Omega_0 \exp\left[-\beta z - \left(\frac{x}{w_x}\right)^2 - \left(\frac{y}{w_y}\right)^2\right] \tag{83}$$

For a two-level atom, the radiative force has two components: a normal to the dielectric surface and a component parallel to the surface. The normal component is associated with a gradient of the field amplitude perpendicular to the surface [86, 167, 174, 175] and it is equal to

$$\mathbf{F}_\perp = -\beta \frac{\hbar \Omega_R^2}{2\Omega} \mathbf{e}_x \tag{84}$$

where Ω_R is the Rabi frequency, $\Omega = \sqrt{\Delta^2 + \Omega_R^2}$, and $\Delta = \omega - \omega_0 - k_x v$. The parallel component of the radiative force consists of the dipole force associated with a gradient of the field amplitude along the surface and the spontaneous light-pressure force

$$\mathbf{F}_\parallel = -\frac{\hbar \Omega_R^2}{\Omega}\left(\frac{x}{w_x^2}\mathbf{e}_x + \frac{y}{w_y^2}\mathbf{e}_y\right) + \hbar k_x \gamma e_z \tag{85}$$

Consideration of the geometry of the atomic initial and final velocity vectors shows that the law of atom reflection of atoms from the evanescent wave [86] is

$$\tan\varphi_r = \tan\varphi_i - 2R \tag{86}$$

where φ_r and φ_i are the angles of reflection and incidence and R is the ratio of the parallel to the normal radiation force $R = F_\parallel / F_\perp$.

When the parameter R is negligibly small, Eq. (86) reduces to the equation of the specular reflection law: $\varphi_r = \varphi_i$ [176]. It occurs when the spontaneous force and the parallel component of the dipole force (Eq. [85]) are considerably smaller then the normal component of the dipole force (Eq. [84]). Because the spontaneous force is negligible for a large laser intensity ($\Omega_R \gg \gamma$), and for a large laser detuning ($\Delta \gg \gamma$), it is evident that a specular reflection of atoms can be expected only for a large laser detuning and a large diameter of laser beam.

Reflections of atoms by the evanescent waves were reported in many papers [138, 166, 167, 177].

Figure 24 shows the atomic intensity distribution produced by the reflection of a metastable argon beam from a simple evanescent wave [167]. The peak at $\varphi_i = 0$ mrad corresponds to the initial atomic beam. The peaks to the right correspond to the atoms that were reflected by the evanescent wave. The position of the reflected peaks is equal to the position expected for specular reflection that indicated by the dashed line. In the experiment, the near specular reflection was expected because the Rabi frequency was much larger than the spontaneous decay rate and the laser beam waist on the surface was much larger the evanescent wave decay length. The broadening of the reflected atomic beam was caused mainly by the defocusing effect and by fluctuation in the dipole force.

Electromagnetic field in optical hollow fiber. We consider a simple cylindrical-core hollow dielectric waveguide as shown in Fig. 25. The refractive index of the cylindrical core and cladding are denoted by n_1 and n_2. The inner diameter and the thickness of the waveguide are 2ρ and $\Delta\rho$, respectively.

The mode picture of a hollow fiber depends on the geometrical parameters of the waveguide. Decreasing the radius of the hole and the thickness of the core leads to disappearance of high-order modes and under certain conditions all modes except the fundamental mode HE_{11} are cut off. Figure 26 shows the map of mode structure regime of a dielectric waveguide in the $\rho - \Delta\rho$ plane where ρ is the radius and $\Delta\rho$ is the thickness of the core. The refractive index of the core and the cladding are chosen as $n_1 = 1.5$ and $n_2 = 1.497$, respectively. In the case of large radius of the hole the fiber can be considered as a planar dielectric waveguide. At a small thickness of the cladding, it is still possible to excite only a fundamental mode. For a larger hollow region and a cladding thickness, there is only a multimode regime. When the radius of the hole is on the order of the optical wavelength, the hollow fiber is a

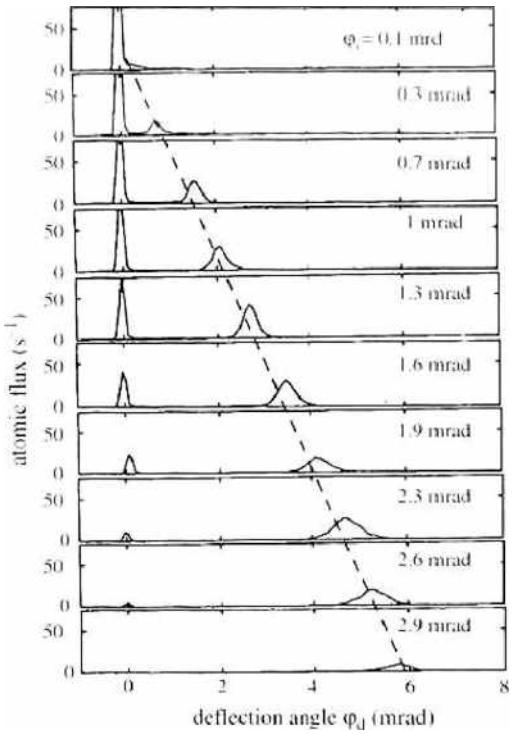


Figure 24. The atomic intensity distribution produced by the reflection of metastable argon atomic beam from a single evanescent wave. The peak at $\varphi_i = 0$ corresponds to the position of the initial atomic beam. The peaks to the right correspond to the atoms that were reflected by the evanescent wave. The position of the reflected peaks was equal to the position expected for specular reflection that indicated by the dashed line. Reprinted with permission from [167], W. Seifert et al., *Phys. Rev. A* 49, 3814 (1994). © 1994, American Physical Society.

single-mode fiber. The electrical field strength in the single-mode fiber has the form [135]

$$\mathbf{E}(r, \theta, z) = E(r)\mathbf{u}(r, \theta)e^{i(\beta z - \omega t)} + \text{c.c.} \tag{87}$$

where $E(r)$ is the electrical field strength and $\mathbf{u}(r, \theta)$ is a complex unit vector. The electrical field Eq. (87), is the field of HE_{11} mode with azimuth mode number $n = 1$. The dependence of the electrical strength on radius for this mode is shown in Figure 27. From the expression (87), it follows that the electrical field has angular dependence however for a circular polarized light field the time average field is circularly symmetric on the slow time scale of the atomic motion.

The most interesting case for the guiding of atoms in a hollow fiber corresponds to the fundamental mode excitation in the fiber: the single-mode fiber guarantees that the electrical field inside the fiber has no zero value anywhere on the wall. In the multimode case the

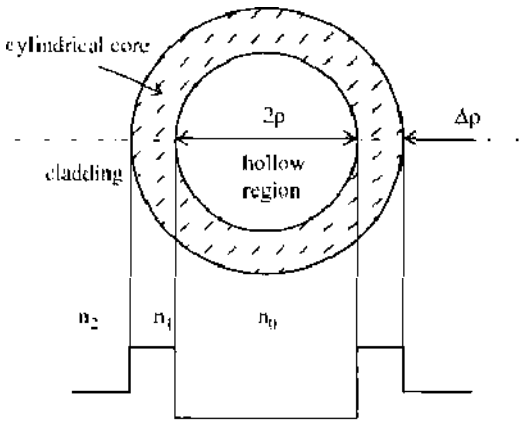


Figure 25. Diagram of a cylindrical-core hollow fiber. The refractive index of the cylindrical core and cladding are denoted by n_1 and n_2. The inner diameter and the thickness of the fiber are 2ρ and $\Delta\rho$, respectively.

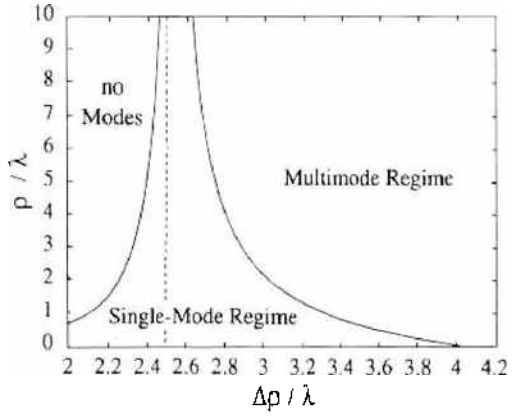


Figure 26. The map of mode structure regime of dielectric waveguide in the $\rho - \Delta\rho$ plane, where ρ is the radius of the hollow region and $\Delta\rho$ is the thickness of the core. In the case of large radius of the hole, the fiber can be considered as a planar dielectric waveguide. At a small thickness of the cladding, it is still possible to excite only fundamental mode that, in this case, corresponds to the TE_{01} and TM_{01} modes of a slab waveguide. For a larger hollow region and a cladding thickness, there is only multimode regime. When the radius of the hole is on the order of the optical wavelength, the hollow fiber is a single mode fiber. [135].

electromagnetic field inside the fiber is a superposition of all allowed eigen modes. The amplitude and the phase of the resulting field depend strongly on the way the fiber is coupled to the external laser source.

5.3.2.2. Cylindrical Hollow Fiber as Atom Waveguide Consider the motion of atoms in the hollow cylindrical fiber with the following assumptions. A laser light is coupled to the fiber and the fundamental mode HE_{11} is excited. The evanescent field of the HE_{11} mode in the hollow region produces the repulsive potential over the surface on the inner wall. Near the surface atoms are attracted to the dielectric by the long range van der Waals force. The resulting potential remains repulsive in a few wavelength distances from the surface and becomes attractive close to the surface. The van der Waals force lowers the effective potential barrier seen by the atoms. Spontaneous emission is one of the loss mechanisms in the atom guiding. Spontaneous emission during the guiding of atoms leads to a recoil heating of atoms and the fluctuation of the dipole force of the optical potential. Other loss mechanisms is the tunneling of atoms to the surface and the nonadiabatic transitions.

We are interested in the atom wave model structure and the atom energy eigenvalues. The stationary Schrödinger equation for an atom in the waveguide is

$$\left[-\frac{\hbar^2}{2m}\Delta + V(\mathbf{r})\right]\varphi(\mathbf{r}) = E\varphi(\mathbf{r}) \tag{88}$$

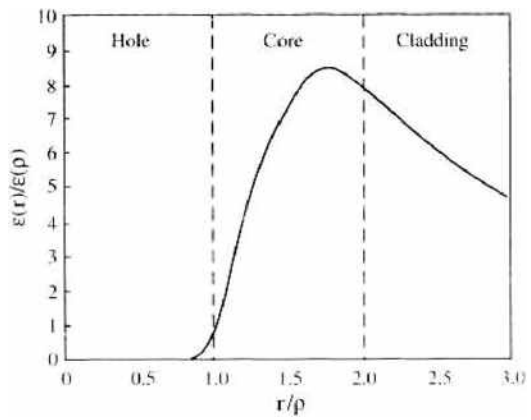


Figure 27. Electrical-field amplitude for the lowest-order HE_{11} mode in the annular core of a hollow fiber. The hollow radius and the core thickness equal 2.889 λ [135].

where $V(\mathbf{r})$ is the combined evanescent wave—van der Waals potential in the waveguide. Because the potential $V(\mathbf{r})$ is axially symmetric and depends only on the distance r of the atom from the surface, the cylindrical coordinates (ρ, θ, z) are best adapted to the problem. Then the Laplacian in cylindrical coordinates [165] is

$$\Delta = \frac{\partial^2}{\partial \rho^2} + \frac{1}{\partial}\frac{\partial}{\partial \rho} + \frac{1}{\rho^2}\frac{\partial^2}{\partial \theta^2} + \frac{\partial^2}{\partial z^2} \tag{89}$$

and the wavefunction $\varphi(\mathbf{r})$ is a function of the variables (ρ, θ, z). The separation of variables leads to the Schrödinger equation in the form [181]

$$-\frac{\hbar^2}{2m}\left(\frac{\partial^2}{\partial \rho^2} + \frac{1}{\rho}\frac{\partial}{\partial \rho} - \frac{l^2}{\rho^2} - k_{dB}^2 - \frac{2m}{\hbar^2}V(\rho)\right)\varphi(\rho, \Theta, z) = E\varphi(\rho, \theta, z) \tag{90}$$

with the wavefunction as

$$\varphi(\rho, \theta, z) = R(\rho)\exp[\pm il\theta]\exp(\pm k_{dB}z) \tag{91}$$

where $k_{dB} = 2\pi/\lambda_{dB}$, and λ_{dB} is the atom de Broglie wavelength. The solution of the Schrödinger equation (88) with the combined potential $V(\mathbf{r})$ is a rather complicated problem. The comprehensive theory of atom guiding with only the evanescent wave optical potential can be found in [135, 163, 182].

However, main features of the atom guiding can be earned by considering a more simple case of an infinite step potential on the wall of the waveguide

$$V(\rho) = \begin{cases} 0, & 0 \le \rho < a \\ \infty, & \rho \ge a \end{cases} \tag{92}$$

For the lowest-order atom wave modes the transverse de Broglie wavelength is much larger than the penetration length of the evanescent wave and the simplified step potential model fits the guiding process adequately. Such a model leads to considerable simplification of the picture of the guiding of atoms.

From Eqs. (90) and (91) with the potential $V(\rho)$ in the form of Eq. (92), the wave function takes the form [181]

$$\varphi_{lq}(\rho, \theta, z) = \text{const}\, J_l\left(\frac{j_{lq}\rho}{a}\right)e^{ik_{lq}z}e^{il\theta} \tag{93}$$

where l the number of azimuth mode and q is the number of radial mode in the atom wavefunction. The boundary condition on the wall of the fiber gives the quantization conditions on the transverse motion

$$\sigma_{lq}a = j_{lq} \tag{94}$$

which can be rewritten for total atom energy as

$$\varepsilon_{lq} = k_{lq}^2 + \frac{j_{lq}^2}{a^2} \tag{95}$$

where j_{lq} is the qth root of the Bessel function (Eq. [93]). The transverse quantization condition of Eq. (95) couples the longitudinal velocity of the atoms $v_{lq} = \hbar k_{lq}/m$ with the transverse velocity $v_{lq}^{\perp} = \hbar\sigma_{lq}/m$. For the basic atom wave mode φ_{01}, the transverse velocity atom in the fiber is of the order

$$v_{lq}^{\perp} = \frac{\hbar\sigma_{lq}}{m} \approx \frac{1.2}{\pi}\frac{\lambda}{a}v_r \tag{96}$$

and considerably smaller than recoil velocity v_r. For the hollow fiber parameters of [135] the transverse atom velocity of the basic mode equals $v_{01}^{\perp} \approx 0.12v_r$. Such small-atom velocity can be reached, for instance, by realizing atoms from a magneto-optical trap and by a following spatial selection of atoms with small transverse velocities. The scheme of laser cooling of atoms to subrecoil temperatures [183] can also be used for a loading of the basic mode of atom waveguides.

5.3.2.3. Losses in Atom Waveguide There are mainly three loss mechanisms for atoms guided in the hollow fiber [135, 136, 140, 164]: (1) spontaneous emission, (2) tunneling of atoms to the dielectric surface, and (3) nonadiabatic transitions between states corresponding to the attractive and repulsive potential of the surface wave. Each of these processes limits the lifetime of atoms in the basic mode and, finally, the length of the atom waveguide.

Spontaneous emission. An atom in a waveguide moves in two different regions: in the surface light wave and in the light free space near the center of the waveguide. When an atom moves in the guiding field, it may re-emit a spontaneous photon. For a single-mode atom waveguide the transverse velocity of the guided atom is smaller than its recoil velocity. That means that a single spontaneous emission event transfers the atom from the basic modes to the higher-order modes and the atom lifetime in the fundamental mode will be determined by the photon re-emission rate.

Tunneling to dielectric surface. When the transverse kinetic energy of an atom in the waveguide is close to the potential barrier of the combined evanescent wave plus the van der Waals potential, the atom may undergo tunneling over the resulting potential. In that case, the atom reaches the dielectric surface and it is either scattered in a diffuse manner (with a considerable gain of its transverse kinetic energy) or absorbed on the surface. In both cases, the atom is lost from the waveguide.

Nonadiabatic transitions. An atom in a waveguide moves in two different regions: in the surface light wave and in the light-free space near the center of the waveguide. In the case of a single-mode atom waveguide with a small hollow region such a picture is oversimplified. A more appropriate physical picture will be that the atom moves in the adiabatic confining state, which converges to the atom ground state at the limit of a zero guiding electrical field [188]. The adiabatic confined state has admixtures of the excited internal state and there is nonzero probability for emission of a spontaneous photon for the atom in this state. This loss mechanism is due to the transition from the confined state to the nonconfined states. The nonadiabatic transitions appear if the atom cannot adiabatically follow the local changes of the confining field intensity. For the typically used parameters of atom waveguides, the nonadiabatic transition loss rate is negligible in comparison with the spontaneous emission loss rate [55, 135, 184].

5.3.2.4. Horn Shape Hollow Fiber The scheme of exposing the fiber's input hole to an atom vapor is probably the most simple scheme of a loading of atom waveguides. In this scheme, the input fiber face is in one vapor cell (which is considered as a source of atoms), and the flux of atoms is directed into another empty cell (where the atoms can be detected). One of the main parameters of such a scheme is the total flux of atoms through the fiber [141, 145, 163]. Suppose that in the first cell we have the Maxwell velocity distribution of atoms. In the case of a single-mode guiding regime, the fiber selects from the Maxwell distribution the atoms with velocities that fulfilled the transverse quantization conditions (Eq. [95]). In a multimode regime of atom guiding, the fiber is coupled to the atoms' having transverse velocity less than the maximum reflected by the surface wave velocity $v_{\max}$.

The total flux of atoms Q coupled to the fiber equals to the number of atoms striking the entrance surface of the fiber per second and having the transverse velocity less then $v_{\max}$:

$$Q = S\int_0^{\infty} dv_z \int_{-v_{\max}}^{v_{\max}} dv_x \int_{-v_{\max}}^{v_{\max}} dv_y\, n(\mathbf{v}) \cong \frac{S n_0 v_{\max}^2}{\sqrt{\pi}} \langle v \rangle \tag{97}$$

where $\langle v \rangle$ is the average velocity of atoms in the source cell and S is the area of the entrance surface of the fiber. For the room temperature atom cell, Eq. (97) gives $Q_1 = 10^{-17} n_0$ atoms/s [140]. Actually the flux will be even less if we take into account the transverse quantization condition (Eq. [95]). It is clear that for a single-mode fiber the considered scheme of the loading of atoms from a vapor cell is not practical.

If we consider a MOT as the initial source for a loading of a single-mode fiber, then the longitudinal velocity of atoms is considerably smaller than in the previous case, but the maximum achievable atom density is also smaller than in a vapor cell and, once more, the flux is discouragingly small.

To overcome the loading problem of the atom waveguide, Balykin et al. [185] proposed using a special type of atom waveguide: a horn-shape hollow fiber. Their waveguide has a shape of a horn with a large entrance diameter (see Fig. 28). The entrance diameter is about 1 mm to fit the typical size of a magneto-optical trap. At the end of the horn fiber, its diameter is decreased to the diameter of a single-mode atom waveguide. As in the case of a cylindrical hollow fiber, the internal wall of the horn fiber is supposed to be covered by an evanescent light wave. A magneto-optical trap is considered an external source to load the horn fiber.

A simple geometrical consideration of a motion of an atom from a MOT source into the horn fiber shows that after several specular reflections from the inner fiber wall the atom will be directed out of the fiber. To overcome this particular loading problem, it was proposed, to use at the initial loading—guiding stage the dissipative reflection of atoms from the evanescent wave [186, 187]. The role of dissipative reflection is not only to load atoms into the horn fiber. The dissipative reflection also implements (a) the cooling of the transverse velocities of atoms in the fiber (and also the longitudinal and the azimuth velocity components through coupling all degrees of freedom); (b) the accumulation of atoms in the narrow end of the fiber; and (c) a population of the base mode of the final cylindrical part of the fiber.

The presence of a mechanism by which an atom in the waveguide loses some of its kinetic energy in inelastic reflections from the evanescent light wave causes the equilibrium position of the atom to move gradually downward. Assume that atoms are being continuously injected from a magneto-optical trap into the waveguide cavity and their velocity distribution corresponds to a Maxwell one with a temperature of $T = \hbar\Gamma/k_B$ that can easily be attained in the trap. The entrance inside the diameter of the waveguide corresponds to the characteristic size of the cloud of atoms in the trap and amounts to 500 μm. The average transverse atomic velocity in the ensemble at the exit from the waveguide amounts to some 10 cm/s, while the mean absolute velocity is around 20 cm/s. However, at the expected high densities of atoms in such a waveguide, there will take place the equalization of their kinetic energy distributions among all their degrees of freedom because of collisions and a long channeling time. It was shown [185] that, while the atoms channel in the waveguide over a distance of $L \cong 1$ cm, their space phase density is increased by five orders of magnitude.

The achievement of extremely cold and dense samples of weekly interacting bosons gives an opportunity to use the horn fiber for investigating the quantum statistics phenomena and wave propagation of matter. Figure 29 presents the relative population $G = N_0/N$ of the minimum-energy mode of the horn-shape waveguide as a function of the number of particles in the magneto-optical trap, N_{MOT}, where N_0 is the number of atoms in the fundamental waveguide mode. One can see that a sharp increase in the proportion of atoms in the

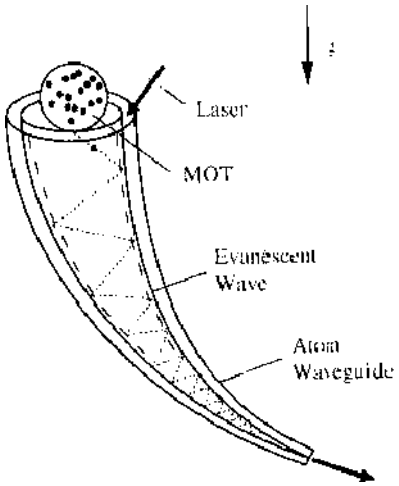


Figure 28. Hornfiber: atoms are injected continuously from a magneto-optical trap into a hollow waveguide with an evanescent light wave formed on its inside surface. The atoms channel over the waveguide while undergoing reflections from the evanescent wave [186, 187].

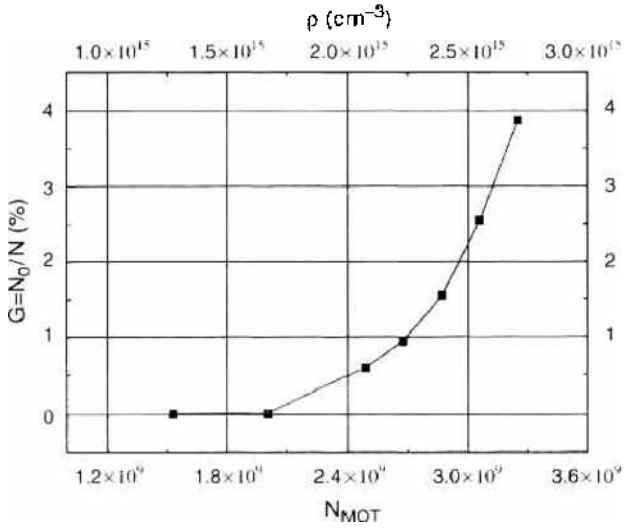


Figure 29. Relative population of the fundamental mode of a cylindrical waveguide with a diameter of $d = 4$ μm, in which the hornfiber terminates, as a function of the number of atoms confined in a magneto-optical trap, of which 10% are being injected every second into the hornfiber. Reprinted with permission from [185], Subbotin et al., *Laser Physics* 7, 358 (1997). © 1997, Maik Nauka/Interperiodica Publishing.

fundamental waveguide mode occurs, and the relative population of the fundamental mode amounts to a few percent.

5.3.2.5. Planar Waveguides A two-dimensional atom waveguide permits us to prepare an atomic sample as two-dimensional gas whose behavior at small temperature and sufficient atom density could be quite different from three-dimensional atomic ensembles [189–191]. Several schemes were proposed for a two-dimensional waveguide. In one of them [87, 175], the waveguide is formed with two different evanescent waves at one dielectric interface. The evanescent waves are different in a sense of frequency detuning with respect to atomic transition and the depth of penetration in the vacuum. One evanescent wave is obtained from the total internal reflection of a red-detuned laser beam at a very small angle of incidence only slightly greater than the critical value. The decay length may in this case be as great as a few wavelengths of light. The second evanescent wave is produced by a blue-detuned laser beam incident upon the interface at a much larger angle of incidence. The decay length for such a wave may be only of the order of one-tenth of the laser wavelength. If the atom is placed close enough to the interface, the first wave will tend to push the atom away from the surface, while the second one will tend to pull it toward the interface. If the condition $\Delta \gg \gamma$ is satisfied, the guiding potential may be represented in the form [63]

$$V = \hbar\gamma\left(\frac{\gamma G_1}{2\Delta_1} + \frac{\gamma G_2}{2\Delta_2}\right) \tag{98}$$

where Δ_1 and Δ_2 are the frequency detuning of the two evanescent waves and G_1 and G_2 are their local saturation parameters. The resulting potential is a Morse potential. The corresponding potential for the laser parameters $G_1 = G_1 = 4 \cdot 10^7$, $\Delta_1 = 5 \cdot 10^5\gamma$, $\Delta_2 = -10^6\gamma$, and the angle of incidence $\Theta_1 = 47°$, and $\Theta_2 = 45.7°$ ($n = 1.4$) has the potential minimum at a distance of $x_{\min} \cong \lambda$ from the dielectric surface and the depth of the well is $\Delta U = 7\hbar\gamma$. The width of the well at the $\Delta V/2$ level is of the same order of magnitude as the wavelength of light.

The guiding time of atoms is determined mainly by their heating as a result of impulse diffusion: the atoms spend a lot of time in the high-intensity red-detuned laser light. For the previously chosen laser parameters and for sodium atoms, the guiding lifetime is of about 1 s [87].

In an other scheme [192], a quasi-two-dimensional gas of atoms was created with the (anti-) nodes of a far-off resonant standing light wave in front of the surface providing a confining potential for the atomic motion normal to the surface. A cloud of cold atoms that approaches

the surface is compressed by deceleration in an evanescent field and loaded predominantly into one single-potential well close to the surface by optical pumping. The spatial density achieved is more than 100 times higher than the density of the approaching atoms.

In the paper Barnett et al. [164] suggested a two-color waveguide based on the evanescent wave fields above a single-mode, submicron optical "channel" waveguide. The waveguide provides tight confinement in *two* dimensions and allows atoms propagation in the third. Their proposal is to use the differing vertical evanescent decay lengths of the two *polarizations* carried in the single-mode optical guide. The physical origin of the decay length difference is the fact that the TM mode is closer to optical *cutoff* than the TE mode at the same frequency.

5.3.2.6. Experiments with Evanescent Wave in Hollow Fiber Experimental demonstration of atom guiding with an evanescent wave were performed in (1) a hollow fiber with relatively large hollow diameter ~20 μm [137] and 100 μm and (2) in a micron-sized hollow fiber [143–145]. In all experiments, Rb or Cs was used as a guided atom. Atom guidance with a fiber of large diameter shows a number of limitations of the principal character. During launching light in a core of the fiber, there is inevitable excitation of the grazing-incidence modes in the hollow region besides the main guiding mode in the core and some fraction of the laser light scatters in the fiber and couples to grazing incidence modes, in particular, EH_{11} mode. The propagation attenuation length of a grazing mode depends on the inner diameter as a^3 and for the case of a relatively large fiber diameter the grazing modes accompany the evanescent mode. However, because they are now bluedetuned, they push atoms to the wall through the weaker evanescent wave.

Renn et al. [137] found that an effective guiding is possible when the light intensity of evanescent wave exceeds the intensity of the basic grazing-incidence mode by a factor of 10. To achieve this ratio, the scattering of the laser light on the fiber wall must be less than 0.05%. It is rather difficult to fulfill in a real experiment this lower-level scattering condition.

Another limitation of large-sized fiber comes from inevitable multimode excitation in the glass-core region of the fiber. The interference between these modes give rise to an optical speckle pattern on the inner glass wall and, as a result, the modulation of the intensity of the evanescent wave. In the "dark" region on the wall of the fiber atoms are attracted by van der Waals force and may be lost from guided atomic flux.

To circumvent the problem of a large-sized fiber, Renn et al. [137] used an additional red-detuned laser beam. The second laser coupled mainly into the lowest-order grazing incidence EH_{11} mode that has the maximum intensity on the fiber axis, and the potential of this mode attracts atoms to an axial region of the fiber. The attractive potential of the additional mode compensates the repulsive potential of the scattering modes at the initial stage of atom guiding until the atoms will not come to the region of a pure evanescent wave potential.

In Renn et al. [137] experiment is used a 6-cm-long fiber with a 20-μm-core diameter. To create the evanescent field, a laser beam of 500-mW power was focused into the annual region of the fiber facet. It was coupled mainly to the evanescent field, but other modes were also excited. The additional laser beam was focused into the hollow region of the fiber, where it was mainly coupled to the EH_{11} grazing-incidence mode. It was sufficient 10-mW of the second laser power to escort atoms at the initial launching stage of their guiding through the fiber until the evanescent potential begins to dominate. When only the red-detuned "escort" laser was used, the guided flux through the fiber was around 200 s^{-1} that is by a factor 500 less than the initially launched flux. Addition of the evanescent field in the fiber enhances the flux by a factor of 3 at the optimal detuning of both lasers. The measurements of guided atom flux as a function of evanescent wave detuning show the dispersive character as expected from the conservative component of dipole force and the flux was increased to its maximum value at the positive detuning of ~2 GHz, which corresponds to the maximum of the dipole potential of the evanescent wave.

The character of evanescent wave atom guiding is qualitatively different than for grazing incidence mode case. In grazing-incidence guiding, the atoms are concentrated near the axis of the fiber and the influence of the van der Waals force is not significant. In evanescent

wave mode guiding, the van der Waals force plays an essential role, especially when the evanescent wave intensity is relatively low. Renn et al. [137] observed the threshold intensity behavior for guiding atoms: for the evanescent wave intensity below 6 MW/m^2, there was no optical guiding of atoms; above this threshold, the ejected flux of guided atoms increases linear with laser intensity. At the threshold intensity, the dipole force from evanescent wave exceeds the van der Waals force.

The detail treatment of the influence of the van der Waals force and the cavity QED effect on the atom guiding was done by the Japanese-Korean group [144, 145] with a micron-size diameter fiber. A great advantage of the use of a micron-size diameter fiber for atom guiding is due to the intrinsic ability of that kind of fiber to support only a desirable guiding mode and strongly suppress all another parasitic modes. An additional attractive feature of a small fiber is that even a small coupled laser power can produce a sufficiently high-potential barrier for atom guiding. Ito et al. [143] demonstrated the guiding of Rb atoms by a cylindrical-core hollow fiber with 7- and 2-μm-hollow diameter. They manage to reach a high-coupling efficiency: about 40% of laser power was coupled into the core of the fiber. The measured light pattern at the exit facet of the 3-cm-long fiber shown that an effectively dark-spot mode could be excited. In the experiment [143], a well-collimated atomic beam was used as the atom source for the fiber. A straight section of an optical fiber was aligned with respect to the atomic beam. The atoms that did not enter the fiber were blocked by a fiber holder. The collimated atomic beam and its alignment with respect to the fiber provided a transverse velocity of atoms up to 0.3 m/s. With the used laser power (several hundred multiwatts) the maximum transverse atom velocity that can be reflected is around 2-4 m/s. Therefore, most atoms impinging on the entrance facet of the fiber are expected to be guided. The hollow fiber of 7 μm diameter was coupled with a laser beam of 130 mW power. Figure 30 shows the transmitted ^{85}Rb flux in the sublevel $F = 3$ as a function of the frequency detuning of the guide laser. In the red-detuned region, the atomic flux is decreased even below the background level (the curve *b*), which testifies to the action of the attractive dipole force with the result of absorption of atoms on the fiber wall. The maximum transmitted flux was $3 * 10^4$ s^{-1} and a comparison with the background transmitted flux gives a rather high enhancement factor of 20.

The same group also demonstrated a first application of atom guidance: it performed an online isotope separation for two stable ^{85}Rb and ^{87}Rb isotopes. The quantum state-selective character of the atom-mirror reflection was demonstrated before in a reflection of sodium atoms [193]. That the atomic reflection is quantum-state and isotope selective follows from the character of the relationship between dipole force and a laser detuning with respect to atomic transition. When the detuning is positive, the gradient force repels the atom from the fiber surface and thus the atom guiding is effected. With a negative detuning, the force

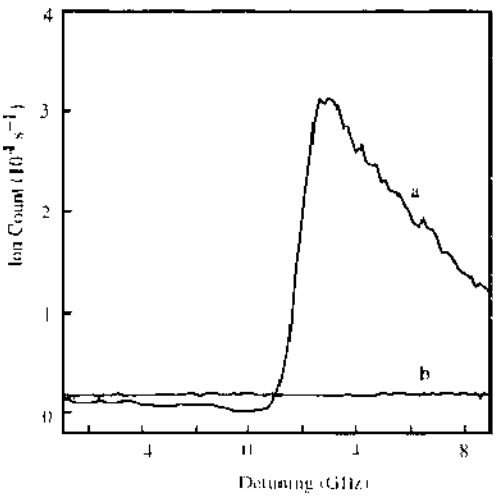


Figure 30. Atom guiding with an evanescent wave in hollow optical fiber. Figure shows the transmitted ^{85}Rb flux in the sublevel $F = 3$ as a function of the frequency detuning of the guide laser [143].

attracts the atoms to the surface where the atoms are absorbed and lost from the guided flux. Ito et al. [143] select a specific isotope by adjusting the guide-laser frequency. Figure 31 shows a demonstration of two isotopes ^{85}Rb and ^{87}Rb separation by the 7-μm-hollow fiber. In the experiment, the atomic transmission flux was recorded as a function of the guided laser frequency. The upper trace of Fig. 31 corresponds to a large blue detuning for both Rb isotopes, and the transmission flux contains both isotopes. When the laser frequency was blue detuned for ^{87}Rb atoms but red detuned for ^{85}Rb atoms, the transmitted flux of the ^{85}Rb isotopes is strongly suppressed (the lower trace on Fig. 31).

5.3.3. Guiding with Evanescent Wave in Solid Glass Fiber

As we have seen in previous section, *inside* the fiber the evanescent wave decays exponentially away from the wall, producing a repulsive potential that guides atoms along the center axis. Alternatively, a red-detuned light in the hollow center of the fiber can also be used to guide atoms. Here we consider a method for trapping and guiding neutral atoms around a thin optical fiber with evanescent wave *outside* of the fiber (Fig. 32). This scheme is based on the use of a subwavelength diameter silica fiber with a red-detuned light launched into it. The light wave decays away from the fiber wall and produces an attractive potential for neutral atoms. The atom trapping and guiding occur in the outside of the fiber. To sustain a stable trapping and guiding, the atoms have to be kept away from the fiber wall. This can be achieved by the centrifugal potential barrier. The centrifugal potential barrier can compensate all the potentials that diverge less rapidly than $1/r^2$. This can be achieved only when the fiber diameter is subwavelength [194, 195]. Nowadays, thin fibers can be produced with diameters down to 50 nm.

Consider an atom moving around an optical fiber (see Fig. 32). The potential V of the atom is cylindrically symmetric, that is, V depends only on the radial distance r from the atom to the fiber axis z. Because of this symmetry, the component L_z of the angular momentum of the atom is conserved. In the eigenstate problem, we have $L_z = \hbar m$, where m is an integer called the rotational quantum number. The centrifugal potential of the atom is repulsive and is given by

$$V_{cf} = \frac{\hbar^2(m^2 - 1/4)}{2Mr^2} \tag{99}$$

The radial motion of the atom can be treated as the one-dimensional motion of a particle in the effective potential $V_e = V_{cf} + V$. There are stable bound states for the atom if V_e has

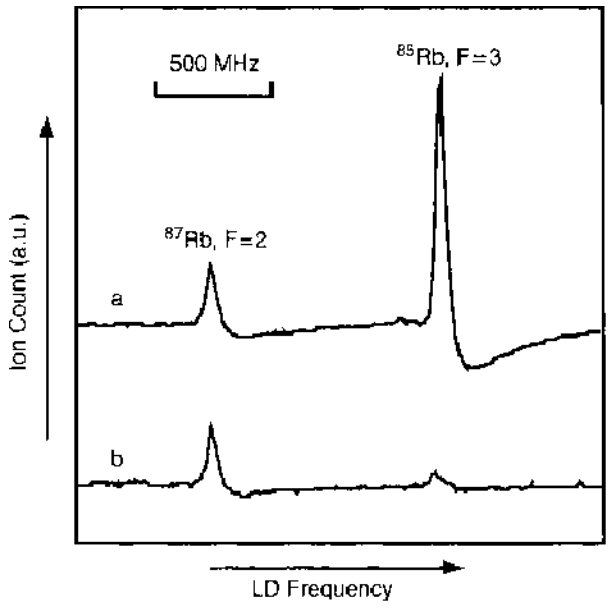


Figure 31. Separation of two isotopes ^{87}Rb and ^{85}Rb through a guiding of atoms by hollow fiber. (a) The laser frequency is blue detuned for both isotopes. (b) The laser frequency is blue detuned for ^{87}Rb and red detuned for ^{85}Rb atoms [143].

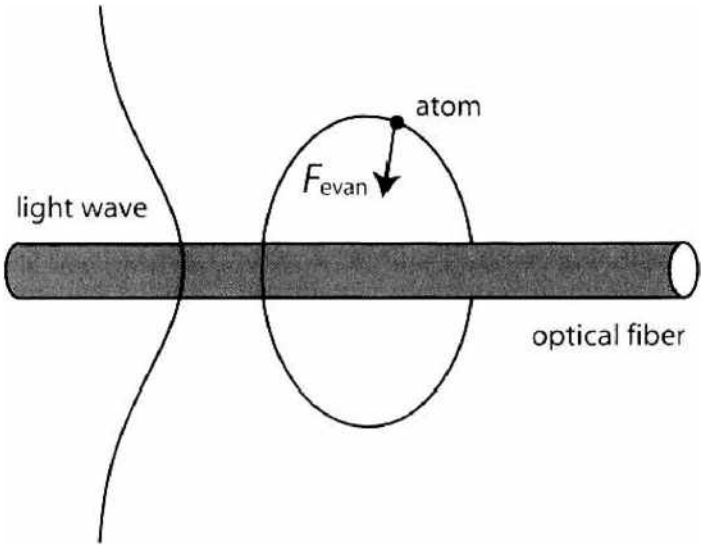


Figure 32. Schematic of atom trapping and guiding around a thin optical fiber with evanescent wave *outside* of the fiber.

a local minimum at a distance $r = r_m$ outside the fiber. This may happen only if the potential V is attractive, opposite to the centrifugal potential V_{cf}. To produce such a potential V, it is necessary to send an optical field through the fiber. This field generates an evanescent wave around the fiber, whose steep variation in the transverse plane leads to a gradient force on the atom. If the atom is initially in the ground state and the detuning of the field from the dominant atomic line is large compared with the Rabi frequency and the line width γ then, the optical potential of the gradient force is given by $V = \hbar\Omega_R^2/\Delta$. We assume that the fiber is sufficiently thin; that it has a vanishing core; and that it can support only a single, fundamental mode HE_{11}. The fiber mode characteristics are then determined by the fiber radius a, the light wavelength λ, the fiber refractive index n_f, and the refractive index n_0 of the surrounding environment. In the linear-polarization approximation, the spatial dependence of the amplitude of the field outside the fiber is described by the modified Bessel function $K_0(qr)$. Here, the parameter $q = 1/\Lambda$ is the inverse of the characteristic decay length Λ of the evanescent-wave field and is determined by the fiber eigenvalue equation [194, 195]. Then, the optical potential outside the fiber can be written as $V = -GK_0^2(qr)$, where $G = \hbar\Omega_R^2/\Delta GK_0^2(qa)$ is the coupling constant for the interaction between the evanescent wave and the atom. Here, Ω_R is the Rabi frequency of the field at the fiber surface. It should be noted here that the field distribution $E(r)$ corresponding to the fundamental mode HE_{11} of the fiber has three nonzero components E_r, E_φ, and E_z, which have azimuthal variation and therefore are not cylindrically symmetric. A simple general way to produce a cylindrically symmetric optical potential is to use a circularly polarized light. The time average of the potential of such a field is cylindrically symmetric on the slow time scale of the atomic center-of-mass motion. Consequently, in the case where the polarizability tensor of the atom is isotropic in the fiber cross-section plane, the optical potential is cylindrically symmetric. The effective potential for the radial motion of the atom in the optical potential V can be written in the form [194, 195]

$$V_{\text{eff}} = R\left[\frac{m^2 - 1/4}{k^2 r^2} - gK_0^2(qr)\right] \tag{100}$$

where $R = (\hbar k)^2/2m$ is the recoil energy and $g = G/R$ is the normalized coupling parameter. Here, m is the mass of the atom and k is the wave number of the field. It was shown [194, 195] that the gradient force of a red-detuned evanescent-wave field in the fundamental mode of a silica fiber can balance the centrifugal force when the fiber diameter is about two times smaller than the wavelength of the light and the component of the angular momentum of the atoms along the fiber axis is in an appropriate range. As an example, the system should be

realizable for Cesium atoms at a temperature of less than 0.29 mK using a silica fiber with a radius of 0.2 μm and a 1.3-μm-wavelength light with a power of about 27 mW.

The existence of a deep local minimum of the potential V_{tot} leads to the existence of bound states of the atom. In Fig. 33, the wave functions u_n are plotted for the first five energy levels of the radial motion of a cesium atom in the potential V_{tot}. Since the potential is deep, the lower levels practically do not depend on the van der Waals potential. They are mainly determined by the effective optical potential V_{eff}. However, the upper levels, which are not shown in the figure, are sensitive to the van der Waals potential and also to the boundary condition at the fiber surface. Tunneling from such highly excited levels into the narrow potential well between the repulsive hardcore potential and the attractive van der Waals potential may occur.

5.3.4. Atom Guiding with a Standing Light Wave

In this section, we consider a standing wave of laser light as an atom-guiding configuration. Intensity variation within a standing wave creates a series of parallel dipole-force potential wells, spaced by half of the wavelength. A single potential well of a standing wave is an almost ideal atom waveguide. Indeed, for blue detuning of laser frequency with respect to the atomic transition, the potential minimum lies along the standing wave node. It means that during their guiding the atoms experience a minimum of impulse diffusion. The shape of the potential well is perfectly defined by a mirror setup (used to form the standing wave) and the laser-light phase. Both parameters can be controlled in the guiding experiments with very high precision. A spatial imperfection of the mirror and a spatial variation of the laser beam profile have an insignificant influence on the shape of the individual well. The bottom of the individual well can lie along a straight line (with a flat mirror setup) or along a curved line (with a curved mirror setup). The shape of the potential well can be even a more complicated form with a use of computer-generated optical elements. There are no spatial spreading and variations of the potential well formed by a standing wave. A transverse size of the potential well is suited to a single-mode regime of atom guiding.

There is one drawback of a standing wave as an atom waveguide: it is almost impossible to localize all laser power in one or several individual potential wells. As a consequence, the standing wave atom waveguide needs either a high-laser power or its length is rather limited.

5.3.4.1. Atom Potential in a Standing Wave An appropriate description of atom motion in far frequency detuned from the atomic resonance standing wave is a "dressed" state approach [78]. Let us consider a two-level atom with a ground state $|g\rangle$ and an excited state $|e\rangle$ in a laser standing wave $E(z) = 2E_0 \cos kz \cos \omega_L t$, where E_0 is the maximum field strength of one traveling wave component, and k is the wave vector that is along the z-axis. The uncoupled states of "atom + photons of laser field" can be written as $|g, n+1\rangle$ and $|e, n\rangle$. The first state corresponds to the atom in the initial state $|g\rangle$ in the presence of $n+1$ photons and the second state corresponds to atom in the excited state $|e\rangle$ and the presence of n photons. These states are bunched in a manifold separated by energy $\hbar\Delta$. When the coupling is taken into account, the two unperturbed states transformed into perturbed states $|1, n\rangle$ and $|2, n+1\rangle$. Bunches of these states are called "dressed" states. Each dressed state is a linear superposition of the unperturbed state $|g, n+1\rangle$ and $|e, n\rangle$ [78]

$$|1, n\rangle e^{i\hbar/4} \cos\theta |e, n\rangle + e^{-i\hbar/4} \sin\theta |g, n+1\rangle \tag{101}$$

$$|2, n\rangle e^{i\hbar/4} \cos\theta |e, n\rangle + e^{-i\hbar/4} \cos\theta |g, n+1\rangle \tag{102}$$

The eigen energies of the atom in the dressed states are

$$E_1 = \hbar\omega(n+1) - \frac{\hbar\Delta}{2} + \frac{\hbar\Omega(z)}{2} \tag{103}$$

$$E_2 = \hbar\omega(n+1) - \frac{\hbar\Delta}{2} - \frac{\hbar\Omega(z)}{2} \tag{104}$$

where $\Omega(z)$ is off-resonance Rabi frequency, $\cos 2\Theta = -\Delta/\Omega$, $\sin 2\Theta = \Omega_R \sin kz/\Omega$, $\Omega \equiv \Omega(z) = [\Omega_R^2(z) + \Delta^2]^{1/2}$, and $\Omega_R(z) = \Omega_R \sin kz$. The last terms in Eqs. (103) and (104)

are related to an interaction of the atom with the laser field and can be interpreted as a potential energy of the atom in the laser field. The potential energy function $U_1(z)$ and $U_2(z)$, corresponding to the states $|1, n\rangle$ and $|2, n\rangle$, respectively, can be derived from the total energy E_1 and E_2 (Eqs. [103] and [104]) [196]:

$$V_1 = \frac{\hbar}{2}\left[(\Omega_R^2 \sin^2 kz + \Delta^2)^{1/2} - |\Delta|\right] \tag{105}$$

$$V_2 = \frac{\hbar}{2}\left[\Omega_m - (\Omega_R^2 \sin^2 kz + \Delta^2)^{1/2}\right] \tag{106}$$

The depth of the potential wells is $V_{\max} = (\hbar/2) * (\Omega_m - |\Delta|)$, where $\Omega_m = [\Omega_R^2 + \Delta^2]^{1/2}$. If we take into account the coupling of the dressed states with a vacuum field, then through spontaneous emission both states $|1, n\rangle$ and $|2, n+1\rangle$ can decay to the state $|1, n-1\rangle$ and $|2, n\rangle$; that is, the atom changes its energy from E_1 to E_2. Spontaneous emission transfers the "atom + laser field system" from one dressed state to another one. When spontaneous emission rate is sufficiently high that the atom does not move a significant fraction of a wavelength between spontaneous events, the atom can be described as moving in a "mean" potential

$$V = V_1\rho_1 + V_2\rho_2 \tag{107}$$

where ρ_1 and ρ_2 are the relative populations of the dressed states. This expression for the potential energy can be written in the more familiar form

$$V = \hbar\Delta \ln(1 + s) \tag{108}$$

where s is a saturation parameter. From Eq. (108) it is clear that the "mean" potential is less than the potentials of the atom in certain dressed states.

5.3.4.2. Guiding Time in a Single Potential Well The dressed-state picture is quite appropriate for an estimation of the atom-guiding time in a single-potential well of a standing wave. For a blue-detuned standing wave, atoms being in the state $|1, n\rangle$ move near the nodes if their maximum transverse energy is less than the potential depth $V_{\max}$. The transverse atomic motion is periodic with interruption caused by spontaneous emission. For sufficiently large detuning, the rate of spontaneous emission is very small compared with the atomic oscillation frequency. When spontaneous emission occurs, the atom from state $|1, n\rangle$ can go to state $|2, n-1\rangle$, where the atom is not localized. The rate of spontaneous transition from state $|1, n\rangle$ to state $|2, n-1\rangle$ averaged over an oscillation period equals [78, 196].

$$\Gamma_2 = 2\gamma\langle\cos^4\Theta\rangle \tag{109}$$

For sufficiently large detuning ($\Delta \gg \Omega_R$) $\cos\Theta \approx \langle\Omega_R\rangle/\Delta$ and the mean lifetime of the guided atom is

$$\tau_{\text{guid}} = \Gamma_2^{-1} = \tau_{\text{sp}}\left(\frac{2\Delta}{\langle\Omega_R\rangle}\right)^4 \tag{110}$$

where $\langle\Omega_R\rangle$ is the Rabi frequency average over an oscillation period. From Eq. (110), we can conclude that the guiding time can be easily made sufficiently long by increasing the detuning. However, at the same time, the potential depth is also decreased.

It is rather interesting to estimate the guiding time for a single-mode guiding regime since it requires the minimum potential well depth. For the quantum mechanical ground state with spatial spot size $z_0 = \lambda/2\pi$, the transverse energy spot is about the recoil energy $R = (\hbar k)^2/2m$. Then, at the potential depth comparable with recoil energy $V_{\max} \sim R$, the guiding time becomes

$$\tau_{\text{guid}} \approx \tau_{\text{sp}}\left(\frac{\Omega_R}{\omega_r}\right) \tag{111}$$

where $\omega_r = R/\hbar$ is the recoil frequency. At the rather low Rabi frequency $\Omega_R = 10^8\omega_r$, the guiding time can be of the order of 1 s. For deeper localization of atoms near the

node of the standing wave (to avoid atom tunneling between the neighboring potential wells), it is necessary to increase the Rabi frequency with an appropriate increasing of the detuning.

5.3.4.3. Experiments with a Standing Wave Three different methods were used to investigate the guiding of atoms in a standing wave. When an atomic beam crosses a standing wave and the atoms have a low enough transverse kinetic energy, they are guided into the channels where they move along the channel and oscillate in the transverse direction. Prentiss and Ezekiel [131] detected an increase of atomic concentration in the vicinity of nodes of the wave by measuring the fluorescence line shape of a beam of sodium atoms that crossed a plane standing wave at a right angle. The detected asymmetry in the fluorescence line shape was attributed to the action on the atoms of the gradient force that caused the concentration of atoms near the nodes of the standing wave. Salomon et al. [132] used absorption of the additional weakly resonant wave to measure the atomic density distribution in a standing wave. The atoms inside the standing wave were chosen as probes of their position. Because of spatial varying of the laser field, the light shift depends on the position of atoms in the standing wave: atoms at a node have no light shift; elsewhere the absorption line is shifted. The calculation showed that the absorption spectrum of uniformly distributed atoms in a standing wave is quite different from the spectrum of atoms with a periodic spatial distribution of atoms near the nodes. The density of atoms was found to increase near the nodes or loops of the standing wave, depending on whether the light-frequency detuning was positive or negative with respect to the atomic transition frequency. The experiment has been performed with a Cs atomic beam. From the experimental absorption spectra have been deduced to correspond with the spatial distributions of atoms that have shown the concentration of atoms near the nodes of the standing wave (Fig. 33).

A clear demonstration of guiding of atoms was obtained using a curved standing wave formed by a spherical laser wavefront (Fig. 34) [133, 197]. The atomic beam transverses the spherical standing wave at a point far from the beam waist. In the polar system of coordinates related to the spherical wave, the effective potential has the form

$$V(z) = V_0 \cos^2 kz + fr \tag{112}$$

where the first term is the atomic potential in the laser field, the second one is an inertial potential, f is the centrifugal force, r is the wavefront radius at the point of atom-wave interaction. The inertial field gives rise to the force averaged over the standing wave period. This force accelerates the nonlocalized atoms and hence causes their spatial separation from the localized atoms. This makes it possible, first, to measure the atomic localization effect itself by observing the spatial separation of the atoms and second, to isolate cold (localized) atoms from nonlocalized ones.

A very successful experiments with a standing wave as microlenses will be discussed in Section 6. Each potential well of a standing wave can produce focusing of atoms in the near field (i.e., within the standing wave) [198–202].

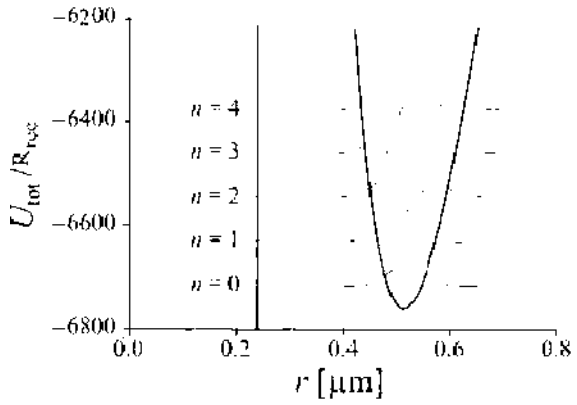


Figure 33. Atom trapping around a thin optical fiber. Figure shows the bound states n_r for the first five levels of the radial motion of cesium atom in the total effective potential V_{tot} [133, 197].

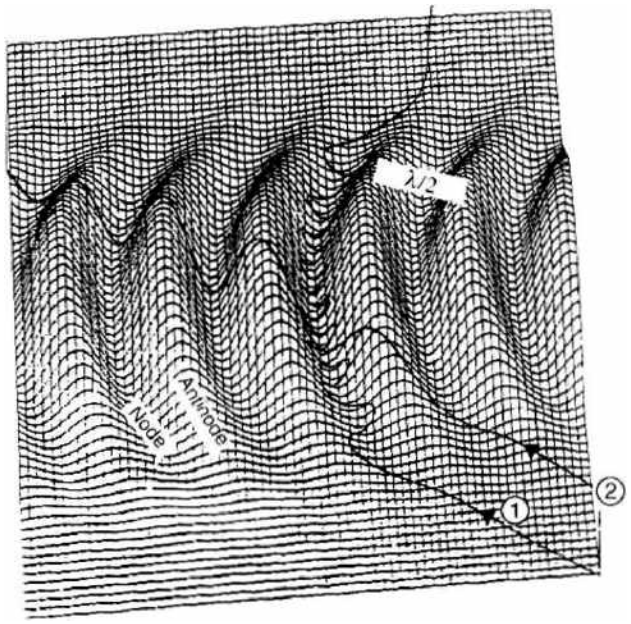


Figure 34. The potential energy of the localized (1) and nonlocalized (2) atoms in a standing spherical light wave, which reflects their trajectories in the laser field [133, 197].

6. FOCUSING OF ATOMS BY OPTICAL FIELDS

6.1. Basic Atomic Lens Idea

As in any other type of optics, the principal element of atomic optics and nanooptics is a lens. As a first approximation, the motion of atoms in the field of a gradient force or a light pressure force can be treated in the classical approximation, that is, as the problem on the motion of a massive particle in a force field (Fig. 35).

For a beam of particles issuing from point 1 to be focused, it must get collected, after traversing the focusing region, at point 2. To this end, the beam must deflect through the angle

$$\Delta\varphi = \varphi_1 + \varphi_2 \approx \rho\left(\frac{1}{z_1} + \frac{1}{z_2}\right) \tag{113}$$

from its original direction: ρ is the radial coordinate in Eq. (113).

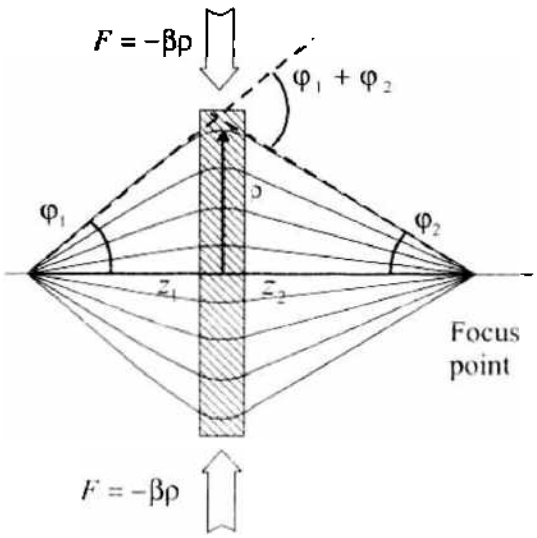


Figure 35. Geometry of main idea of atom focusing.

The deflection through the angle $\Delta\varphi$ is due to the change of the transverse velocity of the particle:

$$\Delta v_{\perp} = v_{\parallel} \Delta\varphi \tag{114}$$

However, the change of the transverse velocity is due to the effect of the transverse force $F_{\perp}$ acting on the particle while it traverses the focusing region:

$$\Delta v_{\perp} = \frac{F_{\perp} t_{\text{int}}}{M} = \frac{F_{\perp} L}{M v_{\parallel}} \tag{115}$$

Equating expressions (114) and (115), one can find the well-known expression for the force capable of focusing an atomic beam:

$$F_{\perp} = \frac{M v_{\parallel}^2}{L} \left(\frac{1}{z_1} + \frac{1}{z_2} \right) \rho = \beta \rho \tag{116}$$

This force must be proportional to the distance from the beam axis to the point of passage of the given atom, and the respective potential (if any) must be quadratic in this distance (i.e., it must be a harmonic oscillator potential).

However, the quadratic potential can be derived from the condition that atoms entering the optical field at different points must spend the same time to reach the focal point. It is only when the atomic motion takes place in a harmonic potential that the oscillation period is independent of the amplitude (i.e., the distance from the axis of the system to the point of entry of the given atom into the focusing region).

Thus, to create a lens for atomic beams, it is necessary to produce such optical fields as give rise, at least in a single direction, to a force proportional to the distance from the focusing axis. And if such dependence holds for one direction only, one can speak of a cylindrical lens. But if such dependence proves valid for any direction, one then can speak of an ordinary lens.

In the next section, we will consider optical field configurations that can, in principle, form the basis for the development of atomic lenses. Before moving on to the discussion of such configurations, we would like to mention briefly one more aspect of atomic lenses. Because the forces acting on atoms are generally weak, atomic lenses are often produced in practice by fields that extend over some distance in the axial direction (i.e., along the z-axis). In this situation, lenses are thick more often than not, and in many instances, multiple crossovers can occur within an atomic lens. In such a situation, the concept of focal length is less useful. The atoms are essentially "channeled" by the focusing field, and the lens acts more as a concentrator than as a true lens, in the sense that a small spot can be produced, but no image can ever be formed at the focus. However, if the purpose of a lens is to produce a very small spot of atoms, this is not necessarily a disadvantage, and in many cases, it can even be an advantage. If the atoms are channeled through a lens, one is not relying here on focusing at a particular focal point but rather on the average effect of many oscillations within the lens. Thus the effects of the velocity spread in the atomic beam are drastically reduced, and the tolerance for the focal spot location is greatly increased. The final spot size is, of course, not as small as can be achieved with true focusing, but nevertheless this technique can be used to compress atomic beams into nanometer-size spots.

6.2. General Equations

For one to be able to implement a precision control over an atomic beam and to focus it into nanometer-size regions, one should know how to calculate exactly its dynamics in light fields differing greatly in geometry and at different values of the parameters, too.

The dynamics of an atomic beam in the field of an electromagnetic wave is an extremely involved matter, and to adequately calculate it one should simultaneously take account of the dynamics of the internal degrees of freedom of the atom, the dynamics of its translational degree of freedom, and radiation processes. The general approaches to the solution of this problem can be found in the literature [54, 55].

However, faraway from resonance the rate of spontaneous decays becomes low so that the atomic momentum and dipole force fluctuations can in the first approximation be disregarded. As a result, the effect of the dipole force becomes predominant, and the quantum-mechanical dynamics of the atomic beam can be described by a Schrödinger equation, wherein the role of the potential is played by the potential of the optical gradient force:

$$\left[-\frac{\hbar^2}{2m}\Delta + V(\mathbf{r})\right]\Psi(\mathbf{r}) = E\Psi(\mathbf{r}) = \frac{\hbar^2 k^2}{2m}\Psi(\mathbf{r}) \tag{117}$$

Here $\Psi(\mathbf{r})$ is the atomic beam wavefunction, k is the wave vector of the incident atom at infinity, and $V(\mathbf{r})$ is the potential of the optical gradient force (see Section 3). By solving Schrödinger equation (117) by some method or other, we get a spatial description of the atomic beam and in its focal region as well. One then can find the necessary corrections to this description associated with the transition of the atom to its excited state and the spontaneous decays of the latter.

To obtain qualitative (and even quantitative) results, there is frequently no need for one to solve Schrödinger equation (117). If all the characteristic dimensions of the focusing field are great in comparison with the de Broglie wavelength of the atom, then to estimate the width d of the focal spot at half-maximum intensity, use can be made of the usual expression

$$d = \frac{0.61\lambda_{\mathrm{dB}}}{\alpha} \tag{118}$$

where α is one-half of the angle at which the aperture is viewed from the focal point.

To trace the trajectories of atoms in a lens and find the focus position, the starting point is with the basic equations of motion derived from the classical mechanics [201]. In a cylindrically symmetric potential, these reduce to

$$\frac{d^2\rho}{dt^2} + \frac{1}{m}\frac{\partial V(\rho, z)}{\partial \rho} = 0 \tag{119}$$

$$\frac{d^2 z}{dt^2} + \frac{1}{m}\frac{\partial V(\rho, z)}{\partial z} = 0 \tag{120}$$

where ρ is the radial coordinate, z is the axial coordinate, m is the mass of the atom, and $V(\rho, z)$ is the potential of the optical force. We note that Eqs. (119) and (120) are also applicable in a one-dimensional focusing geometry, such as the one found in a one-dimensional laser-standing wave, with the substitution of the coordinate x for ρ. Thus, all the following discussion also applies to this geometry.

One approach to analyzing an atom optical lens is simply to integrate numerically Eqs. (119) and (120). This approach certainly gives useful information [199], but for a motion that is generally axial, it often proves useful to eliminate time from these equations and write them down as a single equation for ρ as a function of z. This is done by using the law of conservation of energy to reduce Eqs. (119) and (120) to

$$\frac{d}{dz}\left[\left(1 - \frac{V(\rho, z)}{E_0}\right)^{1/2}(1+\rho'^2)^{-1/2}\rho'\right] + \frac{1}{2E_0}\left[1 - \frac{V(\rho, z)}{E_0}\right]^{-1/2}(1+\rho'^2)^{1/2}\frac{\partial V(\rho, z)}{\partial \rho} = 0 \tag{121}$$

where $\rho' = \partial\rho/\partial z$ and E_0 is energy of incoming atom.

To simplify Eq. (121), it is very useful to make the paraxial approximation. This approximation considers the trajectories that are not affected too greatly by the potential, that is, those that are near the axis, and is made by taking the limit $\rho' \ll 1$ and $V(\rho, z)/E_0 \ll 1$. In the paraxial limit, Eq. (121) reduces to

$$\rho'' + \frac{1}{2E_0}\frac{\partial V(\rho, z)}{\partial \rho} = 0 \tag{122}$$

Equation (122) is very simple and can often be solved analytically or at least with minimal numerical assistance. This allows one to derive the first-order lens properties, such as the

focal lengths and the principal plane locations, if the lens is thick. Such an analysis is invaluable in determining the basic behavior of the lens in terms of the external parameters, such as the laser intensity and frequency detuning [85, 201].

Once the paraxial approximation is made, it is then possible to determine the spot-size limitations imposed by the lens aberrations. These aberrations originate from the higher-order terms in the expansion of Eq. (121) and also from any spread that may be present in the velocities of the atoms entering the lens (chromatic aberration). The effects of aberrations can be analyzed by taking the next-order terms in the expansion of Eq. (121), as done in the conventional aberration theory [85]. Alternatively, it might prove more straightforward to solve numerically Eq. (121). This can be done by introducing the slope of the trajectory, $\alpha = \rho' = \partial\rho/\partial z$, as an independent variable and separating the equation into two first-order equations:

$$\rho' = \alpha \tag{123}$$

$$\alpha' = \frac{1+\alpha^2}{2(E_0 - V)}\left(\alpha\frac{\partial V}{\partial z} - \frac{\partial V}{\partial \rho}\right) \tag{124}$$

Equations (123) and (124) can readily be solved by conventional numerical integration techniques. Without going any further into the details of aberrations, we note only that in those types of atomic lenses that have been analyzed so far the spherical aberration (which results from the higher-order terms in the expansion of the potential about the axis) tends to be relatively minor. Thus, in the absence of other aberrations, a diffraction-limited spot size can often be achieved. However, the chromatic aberration arising from the velocity spread of the incident atoms tends to be rather significant. One way to see this is to solve the paraxial equation of motion for a particular lens and derive the velocity dependence of the focal length. For an immersion lens, the focal length is proportional to the velocity, and for a thin lens, it is proportional to v^2 [201]. Since atomic beams tend to have relatively broad velocity spreads, this velocity dependence can lead to large chromatic aberration effects. For this reason, research has been carried out on the possibility of an achromatic lens for atoms [203].

The correctness of these simplified approaches Eqs. (117) and (119), (120) based on the dipole force potential was analyzed in [204, 205] as applied to the focusing of an atomic beam by a one-dimensional standing light wave. The dynamics of a two-level atom was in that case described by the time-dependent Schrödinger equation:

$$i\hbar\frac{\partial}{\partial t}\Psi(x,t) = \left[\frac{p_x^2}{2m} - \hbar\Delta S_z + \hbar\Omega(x,t)S_x\right]\Psi(x,t), \quad \Psi(x,t) = \begin{bmatrix}\psi_e(x,t)\\ \psi_g(x,t)\end{bmatrix} \tag{125}$$

where $\psi_g(x,t)$ and $\psi_e(x,t)$ denote the wave functions of the center of mass, corresponding, respectively, to the lower state $|g\rangle$ and the upper state $|e\rangle$ of the atom along the x-axis, p_x is the atomic momentum in the transverse direction, m is the atomic mass, $\Delta = \omega - \omega_0$ is the detuning between the laser frequency and the atomic transition frequency ω_0, and $S_\alpha(\alpha = x, y, z)$ are the spin-1/2 operators for the internal atomic states. The atom-laser coupling was expressed in terms of the local Rabi frequency defined by $\Omega(x,t) = -2\frac{(\mathbf{d}\cdot\boldsymbol{\epsilon})E(t)}{\hbar}\cos kx$, where $\mathbf{d}$ is the dipole moment of the atom relevant to the transition $|g\rangle \Leftrightarrow |e\rangle$, $\boldsymbol{\epsilon}$ is the polarization vector, and $E(t)$ is the electric field at the position of the moving atom. If the atom evolves adiabatically in the lower state, Eq. (125) reduces to the scalar Schrödinger equation (117).

The numerical solution of Eq. (125) was found by the Monte-Carlo method with

$$\Psi(x, t+\delta t) = \exp\left(\frac{-iH\delta t}{\hbar}\right)\Psi(x,t) \tag{126}$$

where

$$H = \frac{p_x^2}{2m} - \hbar\Delta S_z + \hbar\Omega(x,t)S_x - i\hbar\frac{\gamma}{2}S_+S_- \tag{127}$$

where γ is the spontaneous decay rate and $S_\pm = S_x \pm iS_y$.

Spontaneous emission interrupts the evolution and causes a quantum jump to occur. The probability for such a jump to occur in the interval $(t, t + \delta t)$ is

$$p_e(t) = \gamma \delta t \rho_e(t) \tag{128}$$

where $\rho_e(t)$ is the population of the upper state.

Numerical modeling by this scheme has shown that the final spot size width comes mostly from the diffraction aberration. Spontaneous emission broadens the linewidth further.

In Ref. [206] it was estimated the effect of deviations from the conservative gradient potential on the process of focusing of an atomic beam with a standing light wave in a different way. They investigated the classical dynamics of an atom, but with due consideration given for the action of the velocity-dependent optical force [62]. Numerical modeling results showed that the velocity-dependent term has a negligibly small effect on the focusing process but could be quite noticeable in the case of channeling of atoms (i.e., in the case of their repeated reflection from the walls of the atomic waveguide).

Thus, the numerical solutions showed the conservative adiabatic approach to the analysis of the dynamics of atoms in focusing fields (Schrödinger equation [117] and classical Newton's equations [119], [120]) to be highly accurate. This circumstance makes it possible to study in detail only the process of focusing by the gradient force, with the unaccounted factors (aberrations) being subsequently evaluated on the basis of the solutions found.

6.3. Focusing by a Running Light Wave

6.3.1. Focusing by a Gaussian Beam (Coaxial Laser Lens)

The focusing of an atomic beam by means of the gradient force was demonstrated first at Bell Labs [37, 207]. In their scheme, the atomic lens was produced by a CW dye laser beam 200 μm in diameter superimposed upon a beam of sodium atoms (Fig. 36). The laser power was 50 mW, and the frequency detuning $\Omega = -2$GHz. The atomic beam propagated along and inside the narrow near-Gaussian laser beam. The laser frequency was tuned below the atomic transition frequency, so that the gradient force was directed toward the laser beam axis. The radial potential is determined by the saturation parameter

$$s(\rho, z) = \frac{I_0}{I_s}\left[\frac{\gamma^2}{\gamma^2 + \Delta^2}\right] \exp\left[\frac{-2\rho^2}{w^2(z)}\right] \tag{129}$$

where γ is the laser linewidth, Δ is the laser frequency detuning, $w(z)$ is the waist of the atomic beam, and I_0 and I_s are the incident and the saturation intensities, respectively.

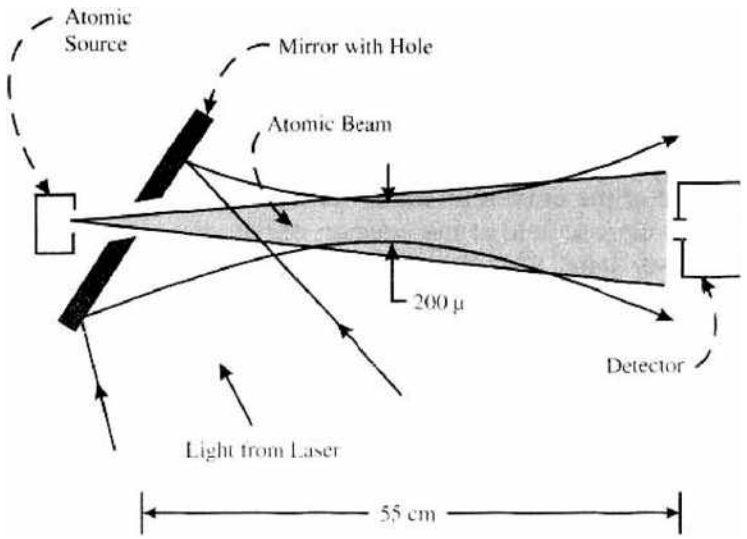


Figure 36. Coaxial laser lens [37].

In the above experiment, the atomic beam was focused into a spot 28 μm in diameter (Fig. 37). The minimum achievable spot diameter in the experiment was determined by the atomic momentum fluctuations due to spontaneous emission.

It is worth noting that every atom in this experiment scattered but a small number of the optical-pump photons, so that the duration of the resonant atom-field interaction was shorter than the time of flight of the atoms through the laser beam. If the atoms interacted with the field all the time, their transverse motion would be alternating focusing and defocusing.

6.3.2. Focusing by the "Doughnut" Laser Mode

The main limitation on the spot size for atoms focused by a co-propagating Gaussian laser beam is imposed by the diffusion of the atomic trajectories caused by spontaneous emission. More preferable for focusing purposes are light beams with a low-intensity region (i.e., dark regions). With the laser-frequency detuning relative to the atomic transition frequency being positive, atoms are expelled from regions of high-field intensity (i.e., light regions) into minimum-intensity regions (dark regions). The atoms in that case move for most of the time in the dark region where the effects of spontaneous decays are negligible.

The best-known light-field configuration of this type is the TEM_{01}^{*} or doughnut mode [85, 153, 208–210]. The intensity distribution in this configuration is defined by the expression

$$I(\rho, z) = 4I_0 \frac{2w_0^2\rho^2}{w^4(z)} \exp\left(-\frac{2\rho^2}{w^2(z)}\right)$$
$$w^2(z) = w_0^2\left(1 + \frac{z^2}{z_R^2}\right); \quad z_R = \frac{\pi}{\lambda} w_0^2 \tag{130}$$

Near the axis of the system, distribution Eq. (130) assumes the parabolic form:

$$I(\rho, z) = 4I_0 \frac{2w_0^2\rho^2}{w^4(z)} \tag{140}$$

Figure 38 presents intensity distributions Eqs. (130) and (140) and Fig. 39 schematically illustrates the focusing of atoms by the TEM_{01}^{*} laser radiation mode.

As noted earlier, the TEM_{01}^{*} optical field mode is very attractive because the motion of atoms along its axis occurs in a low-intensity region, which allows one to speak of the smallness of aberrations due to the fluctuations of the dipole force. The focusing of thermal

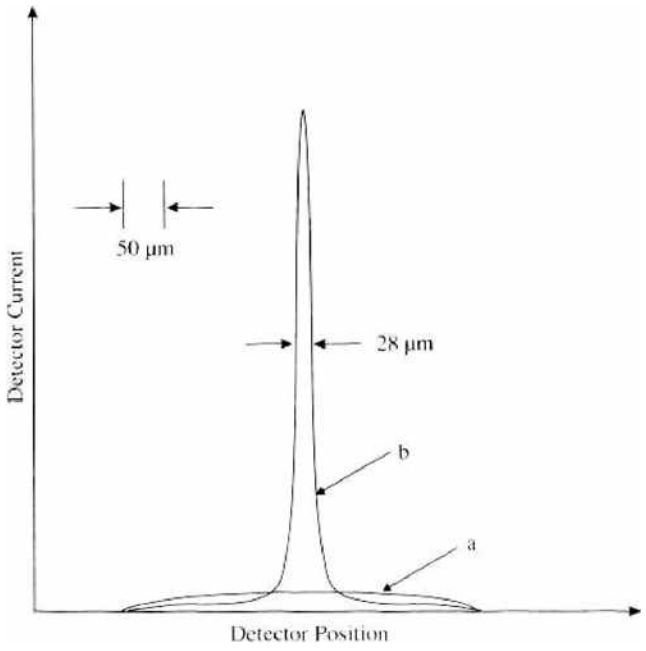


Figure 37. Spatial profile of sodium atomic beam focused by gradient force [207].

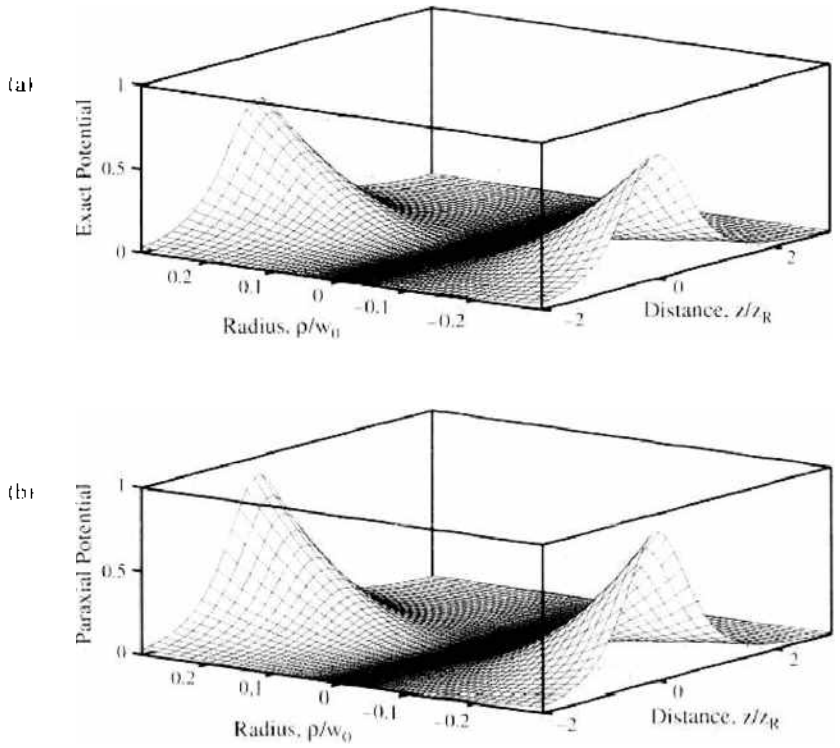


Figure 38. Distribution of intensity for TEM_{01}^{*} mode. (a) (6.18), (b) (6.19).

atomic beams with the aid of this mode was considered in [85, 153, 209]. In Ref. [85] the solutions of the classical equations for the paraxial rays in the field of the gradient force were found and the spherical, chromatic, diffusion, and diffraction aberrations were estimated on their basis. It follows from these estimates that the main contribution to the focal spot width is, as a rule, from the diffraction aberrations; the contribution from the dipole force fluctuations sometimes being also substantial.

The process of focusing of an atomic beam by means of a path integral with a conservative gradient force potential was analyzed numerically in Ref. [153]. By and large, the results obtained in Refs. [85, 153] agree well enough. However, as is the case with any approaches based on numerical schemes, the approaches used by these authors are not quite suitable for the purposes of operative evaluation of the focusing process. What is more, these approaches are poorly suited for the description of the ultracold atomic beams ("atomic laser") that

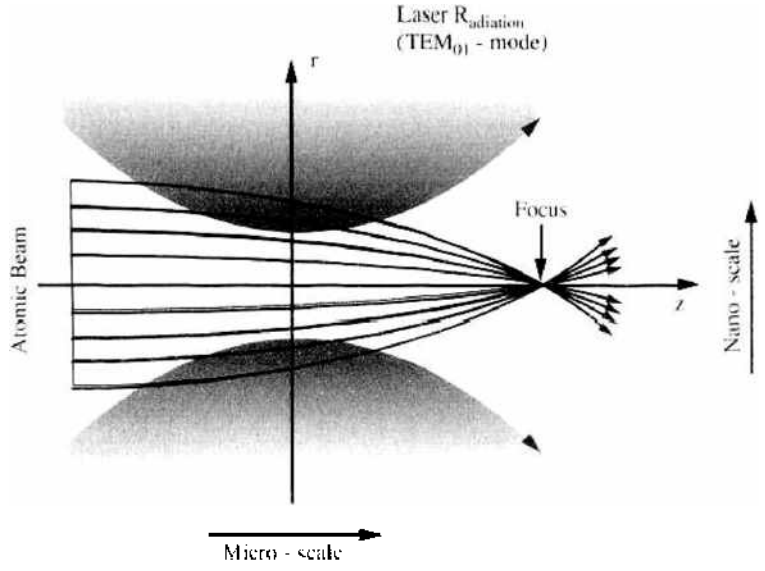


Figure 39. Focusing by TEM_{01}^{*} mode (Scheme).

have recently found widespread application (see, e.g., Refs. [211, 212]), for the de Broglie wavelength of atoms can in that case be comparable with both the laser wavelength and the size of the focusing region.

A theoretical analysis of the focusing process, applicable to the case of ultracold atoms as well, was suggested in Refs. [210, 213] who found an analytical solution of the parabolic equation [214] describing the focusing of paraxial beams:

$$ik\frac{\hbar^2}{m}\frac{\partial}{\partial z}\psi(\rho, z) = \left[-\frac{\hbar^2}{2m}\Delta_\perp + V(\rho, z)\right]\psi(\rho, z) \tag{141}$$

where $\Delta_\perp = (\partial^2/\partial x^2) + (\partial^2/\partial y^2)$ is the Laplacian over the transverse coordinates and $\rho = \sqrt{x^2 + y^2}$. Of course, the assumption of the smallness of the de Broglie wavelength in comparison with the longitudinal gradients, necessary for Eq. (141) to hold true, limits the domain of applicability of the solutions of Eq. (141). However, even in the case of Bose-Einstein condensate ("atomic laser") [211, 212], the de Broglie wavelength for Rb atoms moving with a velocity of 10 cm/s amounts to 0.1 μm, which allows one to speak of the applicability of Eq. (141) in this case. However, parabolic equation (141) has a wider applicability region than the geometrical optics method, especially where the potential depends on the longitudinal atomic coordinate [215].

Thus, in realistic situations, the dynamics of an atomic beam can adequately be described by Eq. (141). Moreover, to obtain exact analytical results, one should confine oneself to paraxial atomic beams for which the relationship between the gradient force potential and the transverse atomic coordinates is of quadratic character:

$$V(\rho, z) = \frac{C}{2}\frac{\rho^2}{w(z)} \tag{142}$$

which is the first term of expansion of potential $V(\rho, z)$ in terms of the transverse coordinate ρ, where C is a constant proportional to the intensity of the laser field and $w(z)$ is the beam waist.

As demonstrated in [210, 213], parabolic equation (141) with potential Eq. (142) can be solved analytically for a variety of beam-width functions $w(z)$. Moreover, when an atomic beam is focused by the TEM_{01}^* laser mode, the Green function, $G(\rho_2, z; \rho_1, z_1)$, of parabolic equation (141) can be expressed by analytical formulae [210, 213]. One can use it as a basis for finding the solutions of a great variety of problems. For example, if the wave function $\varphi_0(\rho)$ is specified at $z = z_1$, one then can use the formula

$$\psi(\rho_2, z) = \int d^2\rho_1\, G(\rho_2, z; \rho_1, z_1)\varphi_0(\rho_1) \tag{143}$$

to find the atomic probability density distribution at any point in space. In that case, by virtue of the Gaussian dependence of the Green function on the transverse atomic coordinates, the integral in expression (143) can be analytically calculated for a wide class of beams.

Specifically, of practical interest is the choice of $\varphi_0(\rho)$ in the Gaussian axially symmetric form

$$\varphi_0(\rho) = \exp\left(\frac{-\rho^2}{\sigma_0^2}\right) \tag{144}$$

where σ_0^2 describes the initial width of the atomic beam.

Substituting expression (144) into (143) and integrating, one can obtain the following expression for the wave function of the atomic beam at any point in space:

$$\begin{aligned}
\psi(\tilde{\rho}_2, \tau_2) &= \frac{iA\sin\alpha_1\sin\alpha_2}{\sin A(\alpha_2 - \alpha_1)(p_1/\tilde{\sigma}_0^2 + iD_1)}\, e^{\frac{i\rho_2^2 D_2}{\tilde{\sigma}^2}} \\
\frac{1}{\tilde{\sigma}^2} &= \frac{1}{p_1}\left[\frac{A^2\sin^2\alpha_1\sin^2\alpha_2}{\sin^2 A(\alpha_2 - \alpha_1)(p_1/\tilde{\sigma}_0^2 + iD_1)} - iD_2\right] \\
D_1 &= \sin^2(\alpha_1)[A\cot(A(\alpha_2 - \alpha_1)) + \cot\alpha_1] \\
D_2 &= \sin^2(\alpha_2)[-A\cot(A(\alpha_2 - \alpha_1)) + \cot\alpha_2]
\end{aligned} \tag{145}$$

where $\tilde{\rho}_{1,2} = \rho_{1,2}/w_0$, $\tilde{\sigma}_0^2 = \sigma_0^2/w_0^2$, $\tau_{1,2} = z_{1,2}/z_R = \cot\alpha_{1,2}$, $A = \sqrt{1 + \tilde{C}p_1^2 p_2/2}$, $p_1 = \lambda_{dB}/\lambda$, $p_2 = \gamma m w_0^2/(2\hbar)$, and $\tilde{C} = \frac{I_0}{I_s}\frac{2\gamma}{\Delta}$.

This expression becomes substantially simplified on the axis of the system ($\tilde{\rho}_2 = 0$):

$$\psi(\rho_2 = 0, \tau_2) = \frac{iA}{(p_1/\tilde{\sigma}_0^2 + i\sin^2\alpha_1(A\cot A(\alpha_2 - \alpha_1) + \cot\alpha_1))}\frac{\sin\alpha_1\sin\alpha_2}{\sin A(\alpha_2 - \alpha_1)} \tag{146}$$

Note that, generally speaking, the wave function within the framework of parabolic approximation Eq. (145) (in contrast to geometrical optics) has no singularities outside of the sources and allows one to describe the atomic intensity distribution near the focal point.

In the case of *thermal beam*, the maxima $|\psi(\rho_2 = 0, \tau_2)|^2$ determine the positions of the focal points $z_c = z_R\tau_c = z_R\cot\alpha_c$. The de Broglie wavelength is small, or to be more exact, $\frac{p_1}{\tilde{\sigma}_0^2} = \frac{\lambda_{dB}w_0^2}{\lambda\sigma_0^2} \ll 1$. As a result, the position of the focal points can be found from the equation

$$D_1 = \sin^2(\alpha_1)[A\cot(A(\alpha_c - \alpha_1)) + \cot\alpha_1] = 0 \tag{147}$$

whose solution has the form

$$z_c = z_R\cot\left(\alpha_1 + \frac{1}{A}\left[\operatorname{arc\,cot}\left(-\frac{\cot\alpha_1}{A}\right) - \pi n\right]\right) \tag{148}$$

where $z_1 = z_R\cot\alpha_1$, and $n = 1$ at $1 < A < 2$, $n = 1, 2$ at $1 < A < 3$, and so on. Thus, depending on the magnitude of the focusing potential, there may be several foci in the system.

The minimum width of the atomic beam at the focal point, $\tilde{\sigma}_c^2 = \tilde{\sigma}^2(\alpha_c)$ can easily be found from expression (145) by means of expression (147):

$$\sigma_c^2 = \lambda_{dB}^2\frac{z_c^2 + z_R^2}{\pi^2\sigma_0^2}\frac{z_1^2 + z_R^2}{z_1^2 + z_R^2A^2} \tag{149}$$

In the case of source located at $z_1 = -\infty(\alpha_1 = \pi)$, the positions of the focal points and the beam widths at them are defined by the expressions

$$z_c = -z_R\cot\frac{\pi n}{A} \tag{150}$$

$$\sigma_c^2 = \lambda_{dB}^2\frac{z_c^2 + z_R^2}{\pi^2\sigma_0^2} \tag{151}$$

Naturally the focal point positions defined by expressions (148) and (150) coincide with those in the geometrical optics approximation [85, 153, 216].

The analysis of aberrations in the case where geometrical optics is applicable [85, 153] showed spherical aberrations and aberrations due to spontaneous decays to be small compared with the diffraction focal spot width (Eq. [151]).

Using expression (145), one can easily find the atomic beam intensity at any point in space. To illustrate, Fig. 40 presents the spatial intensity distribution of a relatively fast beam (whose parameters correspond to those in [153]) in the neighborhood of the focal spot. It is clearly evident from this figure that it is quite possible to focus atoms into a nanometer-size region.

Figure 41 shows the atomic beam width σ as a function of the longitudinal distance z at the same beam parameters as in Fig. 40. One can see from this figure that the position of the focus and the minimum beam width found in [153] within the framework of their numerical approach agree well enough with the exact solution results. It should be stressed that the solid curve in Fig. 41 is described by the following simple analytical formula following from expression (145):

$$\frac{1}{\tilde{\sigma}^{*2}} = \mathrm{Re}\left(\frac{1}{\tilde{\sigma}^2}\right) = \frac{A^2\sin^2\alpha_1\sin^2\alpha_2}{\tilde{\sigma}_0^2\sin^2A(\alpha_2 - \alpha_1)(p_1^2/\tilde{\sigma}_0^4 + D_1^2)} \tag{152}$$

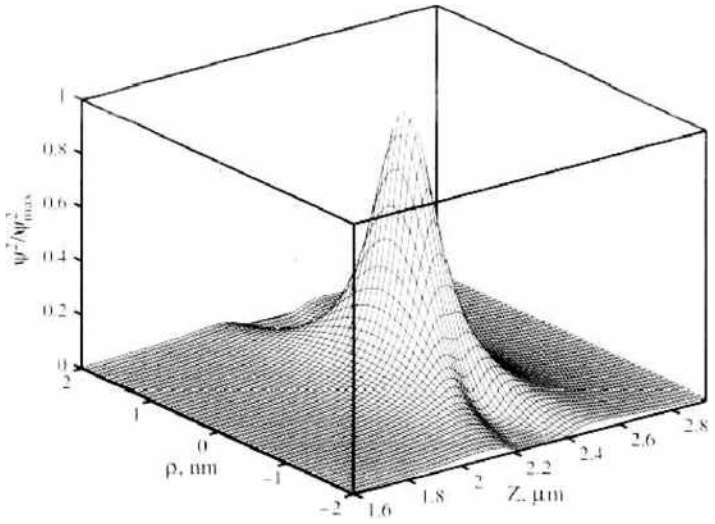


Figure 40. Three-dimensional intensity distribution of a beam of Na atoms in the focal region ($\lambda = 0.59$ μm, $\gamma/2\pi = 10$ MHz, $I_s = 10$ mW/cm^2, $I_0 = 10^5/2\pi$ W/cm^2, $w_0 = 1$ μm, $\sigma_0 = 0.07 w_0$, $v_z = 500$ m/s, $\Delta/\gamma = 40{,}000$).

Formulas (148) through (151) enable one to find easily the main characteristics of lenses, no matter what the parameters of sufficiently fast atomic beams.

In the region of strong focusing and great diffraction effects on the atomic beam, that is, in the case of low velocities of the incident atoms, the approaches used in Refs. [85, 153] do not work, whereas analytical result Eq. (145) still remain to be applicable. To illustrate, Figs. 42 and 43 show the three-dimensional probability density distribution in the case of focusing of a cold atomic beam (with parameters close to those in Ref. [211]) and the corresponding beam trajectories, respectively. One can see from Fig. 42 that the intensity distribution is generally similar to that in the case of fast atoms (Fig. 40). A specific feature of this case is that the de Broglie wavelength of the incident atoms becomes comparable with the laser radiation wavelength. The question arises in this connection as to the minimum possible focal spot size. It is seen from Fig. 42 that the focal spot diameter at the $1/e^2$ level ($2\sigma_z \approx 0.1$ μm) is approximately equal to the de Broglie wavelength ($\lambda_{dB} \approx 0.08$ μm).

Note that the case in point above has been the description of the focusing of an atomic beam by the TEM_{01}^* laser mode. However, the analytical approach [210] described can also be used for the operative description of any other cases of focusing of paraxial beams,

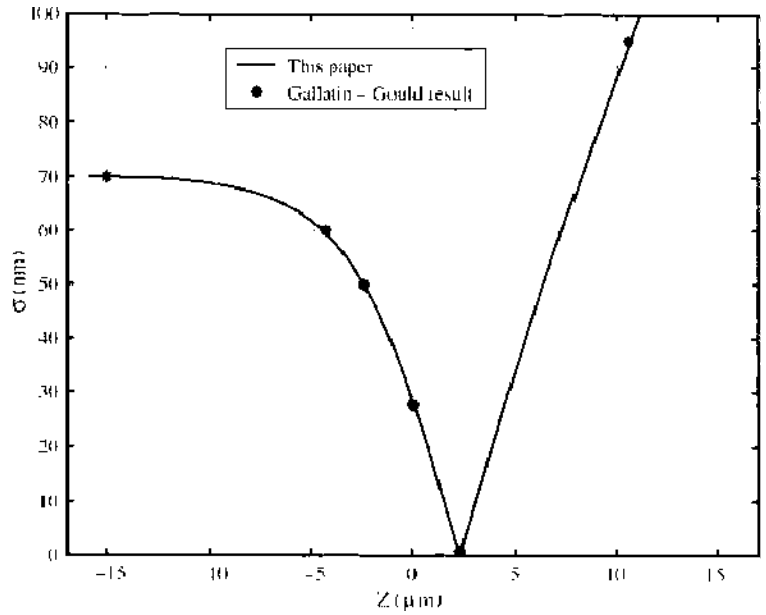


Figure 41. Width σ of a beam of Na atoms as a function of the longitudinal distance z (Eq. [152]). The atomic beam parameters are the same as in Fig. 7. Stars correspond to results of Ref. [153].

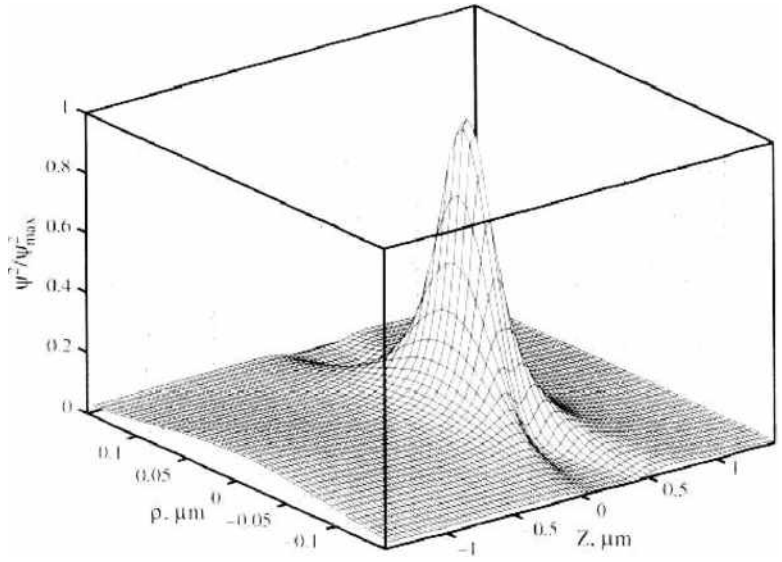


Figure 42. Spatial distribution of the probability density $|\psi|^2$ for cold Na atoms ($\lambda = 0.59$ μm, $\gamma/2\pi = 10$ MHz, $I_1 = 10$ mW/cm^2, $I_0 = 180$ W/cm^2, $w_0 = 0.3$ μm, $r_0 = 0.6\ w_0$, $v_z = 20$ cm/s, $\Delta/\gamma = 20{,}000$).

including those of focusing by photon dots and photon holes [217] considered elsewhere in the text.

The influence of the small effects due to the spontaneous emission processes can be evaluated on the basis of the above solution by using the approaches expounded, for example, in Refs. [85, 153].

6.3.3. Other Light Field Configurations

6.3.3.1. Conical Lens for Atoms In Ref. [218] it was suggested to use a conical lens (Fig. 44) to produce a focusing field for atoms. Near the axis of the system, the focusing electric field is described by the expression

$$E_z(\rho, \varphi) = -\sqrt{\pi i k a}\, E \exp(ika + i\varphi) J_1(k\rho) \tag{153}$$

that is, it is a Besselian beam. Here E is the incoming field amplitude, k is the wave vector, a is the beam radius, and (ρ, φ) are the transversal cylindrical coordinates. Like the TEM^*_{01}

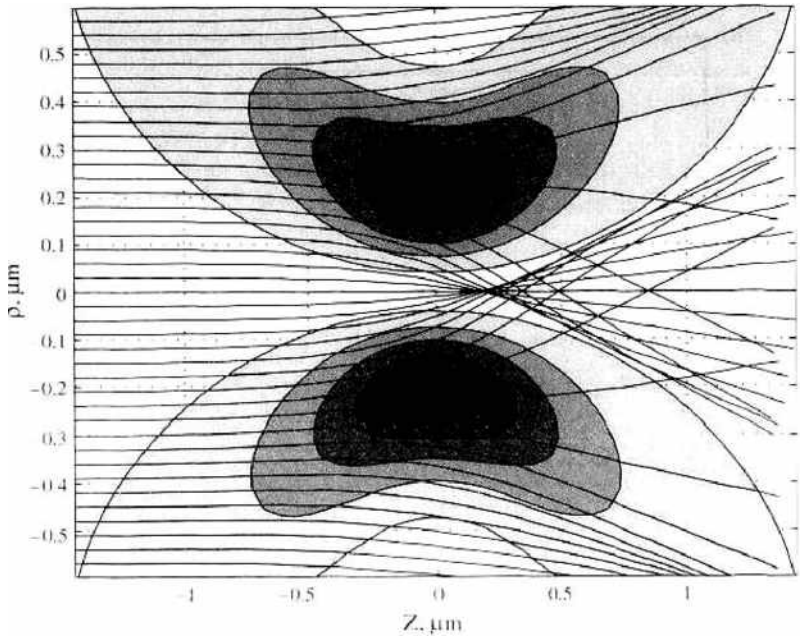


Figure 43. Beam trajectories and potential isolines of the TEM^*_{01} laser wave at the same atomic beam parameters as in Fig. 42.

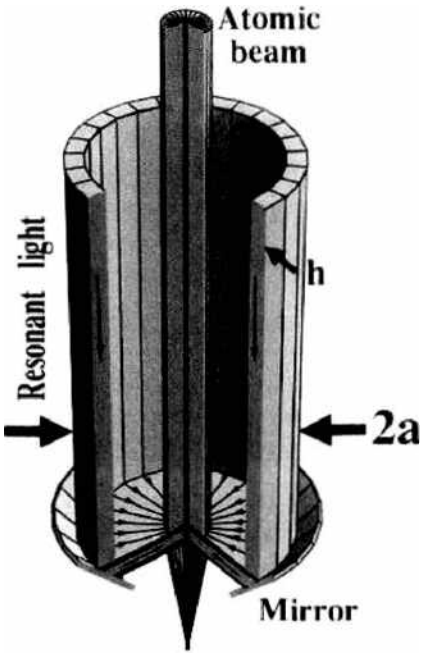


Figure 44. Schematic diagram of a conical lens. Light propagating along a cylindrical shell reflects from a conical mirror and forms a lens that focuses an atomic beam onto one spot [218].

laser mode, such a beam has a minimum at the axis, and in the case of positive laser frequency detuning, it can be used as a lens for atoms.

It was assumed in the conical-lens focusing theory suggested by the above authors that it was only the atomic motion in the (x, y)-plane that was quantized. It is convenient to carry out calculations in an atomic rest frame moving with a velocity of u along the z-axis. In this frame, the atoms are subject to a potential pulse, centered at $t = 0$, given by

$$V(\rho, t) = \pi k a \left(\frac{|dE|^2}{4\hbar\Delta} \right) J_1^2(k\rho)\Theta\left(\frac{2t}{\tau} \right) \tag{154}$$

where $\Theta(t)$ is Heaviside step function. $\tau = h/u$ is the potential pulse duration, d is the dipole moment operator matrix element, and Δ is large positive detuning. The atomic wave function $\psi(\mathbf{r}_\perp, t)$ obeys the Schrödinger equation:

$$i\hbar \frac{\partial}{\partial t} \psi(\mathbf{r}_\perp, t) = [H_0 + V(\rho, t)]\psi(\mathbf{r}_\perp, t) \tag{155}$$

where H_0 is the Hamiltonian of free motion in the (x, y)-plane, $t = z/u$ and $\mathbf{r}_\perp = (\rho, \varphi)$.

It is assumed that the atomic matter wave is uniform over the optical potential and that its initial wave function is $\psi(\mathbf{r}_\perp, -\tau/2) = 1$. Following the atom-field interaction, the wave function becomes

$$\psi\left(\mathbf{r}_\perp, \frac{\tau}{2} \right) = \exp(-i\tilde{A}(\rho)) \tag{156}$$

where $\tilde{A} = AJ_1^2(k\rho)$ and $A = \pi k a \tau(|dE/2\hbar|^2/\Delta)$. We will refer to the quantities A and $\tilde{A}$ as the pulse field areas. In the case of linearly polarized wave, the field area is proportional to the light intensity. Using the two-dimensional propagator for the particles' free motion, one finds that the atomic wave function evolves as

$$\psi(\mathbf{r}, t) = \left(\frac{m}{i\hbar t} \right) \exp\left(i\left[\frac{m}{2\hbar t} \right] \mathbf{r}^2 \right) \int_0^\infty \rho' d\rho' J_0\left(\frac{m\rho\rho'}{\hbar t} \right) \exp\left\{ i\left[\left(\frac{m}{2\hbar t} \right) \rho'^2 - AJ_1^2(k\rho') \right] \right\} \tag{157}$$

where the time origin has been shifted by $\tau/2$ (i.e., t is now reckoned from the end of the potential pulse).

Expression (157) was used to calculate the atomic density $|\psi(\mathbf{r}, t)|^2$. Let us consider the time evolution of the atomic density at the atom lens axis ($\rho = 0$). When the field area is

large ($A \gg 1$), the main contribution to the integral Eq. (157) comes from small values of ρ', where

$$J_1^2(k\rho') \approx \left(\frac{k\rho'}{2}\right)^2 \tag{158}$$

One finds that $|\psi(0, t)|^2 \approx (\omega_R t - A^{-1})^{-2}$. This expression contains a singularity at $t_f = (A\omega_R)^{-1}$, which determines the position of the focal plane. The focal distance in the laboratory frame, $f \equiv ut_f$, is approximately given by

$$f = (2\pi A)^{-1} L_T \tag{159}$$

where $L_T = 2\lambda^2/\lambda_{dB}$ is the so-called Talbot distance and λ_{dB} is the atomic de Broglie wavelength.

This result has a simple interpretation. The central minimum ($\rho = 0$) of the conical lens potential Eq. (154) focuses atoms at a distance of f given by Eq. (159). Other minima focus atoms at distances $z > f$ and do not significantly modify the primary focus Eq. (159) for $f \ll L_T$. This is in contrast to a standing-wave field, where each minimum of the potential focuses atoms in the same focal plane. Moreover, the central potential of the conical lens is around πka times as high as the potential of the standing-wave field near its minimum, and its second derivative is $\pi^2 a/2\lambda$ times as great.

This ray picture of focusing by a harmonic potential is helpful, but detailed calculations must be carried out using the wave approach (Eq. [157]). The focal distance is still defined as $f = ut_f$, but now t_f is the instant of time when the atomic density at the axis, $|\psi(0, t)|^2$, reaches its maximum.

The spatial atomic distribution in the focal plane is illustrated in Fig. 45 for a pulse area of $A = 30$. It can be seen from this figure that, given these beam parameters, the beam can be focused into a spot around 10 nm in radius. It is extremely important that to achieve the same focal spot radius in focusing by a standing light wave it is required that the field intensity should be higher by two orders of magnitude!

6.3.3.2. Laguerre-Gaussian and Besselian Light Beams In Ref. [219] the optical potentials of the more complex Laguerre-Gaussian and Besselian light beams were examined. They demonstrated that Laguerre-Gaussian beams of high-azimuthal index were advantageous for focusing atoms, while high-order Besselian beams offered significant advantages for transporting them over long distances.

6.4. Focusing by a Standing Light Wave

6.4.1. Single Standing Light Wave

6.4.1.1. Large-Period Standing-Wave Lens The problem associated with the great thickness of the coaxial laser lens (severe aberrations) was overcome by a group of

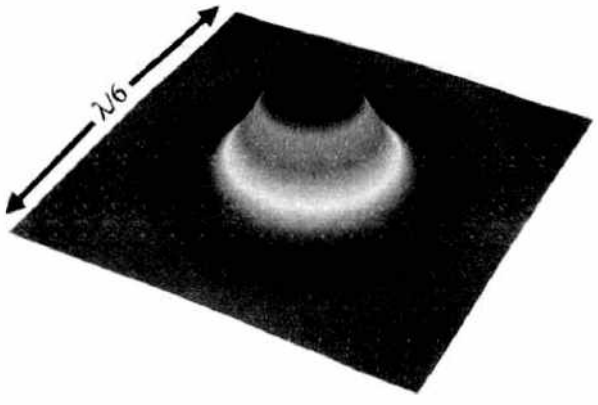


Figure 45. Atomic spatial distribution at the focal plane of conical lens for ^{52}Cr atomic beam obtained with the use of Eq. (157) [218]. Parameters are taken close to those in experiments [246, 224]: $\lambda = 425.5$ nm, $\Delta/2\pi = 350$ Mhz, $u = 8.1 \times 10^2$ cm/s, $P = 89$ μW, $a = 1$ cm, field area $A = 30$.

investigators at Constanz University [220]. In this work, the atom lens was based on a large-period (45 μm) standing light wave produced by bouncing a laser beam, tuned just below the $2^3S_1 \rightarrow 2^3P_2$ transition at 1083 nm, off a glass surface at a small angle (Fig. 46). One period of such standing wave was used as a cylindrical lens for atoms. A 25-μm-diameter, beam of metastable He^* atoms crossed the standing wave at right angles at its antinode. The thickness of the atomic lens was 80 μm, so that the atom-field interaction time $t_{int} = 40$ ns was much shorter than the natural lifetime of the excited state 2^3P_2 of the metastable He^* atoms ($\tau_{sp} = 100$ ns). This lens focused the atomic beam into a spot 4 μm across. The major contribution to this spot size was considered to be of diffraction origin, because of the long focal length (28 cm) and the small lens aperture (25 μm). Chromatic aberrations were held to a minimum because the atomic beam in this case was produced in a supersonic expansion. An additional interesting feature of this lens is that it was formed under conditions of relatively high intensity and small laser frequency detuning. Spontaneous emission would usually be a major effect under such conditions, but in this case, the transit time through the lens was too short for any significant effect to occur. This lens was also used to image the atomic source, the image being produced by passing the atoms through a microfabricated transmission grating with a period of 8 μm.

6.4.1.2. Standing-Wave Lens Array The field of an ordinary standing light wave (Fig. 47[a]) can serve as an example of an array of cylindrical lenses for an atomic beam. The principle of this approach is to make use of each node of a near-resonant, blue-detuned standing laser–light wave as an individual lens, so that the entire standing wave acts as a large lens array. Near the center of the standing-wave nodes, the intensity increases quadratically as a function of the distance from the node center. This intensity variation gives rise to a quadratically varying potential (as long as the proportion of the excited-state atoms remains small), and hence the force on the atom is linear. As the intensity gradient inside the nodes is high (the intensity goes from zero to a maximum over a quarter of the optical wavelength), with a standing-wave lens it is relatively easy to attain quite short focal lengths (on the order of a few tens of micrometers), hence small focal spot sizes, down to the nanometer-scale ones.

This idea was first developed in Ref. [198] where an optical standing wave was used as an array of cylindrical lenses (each period of the standing wave served as a lens) to focus a perpendicular beam of sodium atoms. The atomic beam was focused onto a grating with a period of $\lambda/2$ on a substrate.

The next experimental work was aimed at applying the principle of laser focusing to the deposition of chromium atoms on a surface [200]. Figure 47(a) is a schematic illustration of this experiment. A collimated, uniform beam of chromium atoms is directed onto a silicon surface. Standing wave positioned immediately above silica substrate acts as a cylindrical lens for the atoms. The atoms are thus focused into a series of lines spaced half a wavelength

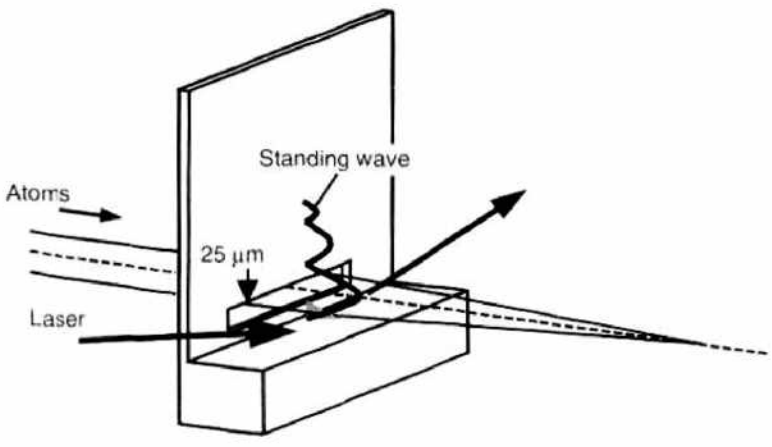


Figure 46. Large-period standing wave lens. A below-resonance laser beam reflects at grazing incidence from a substrate, creating a standing wave with a 45-μm-wide antinode. Atoms, apertured by a 25-μm slit aligned with the peak of the antinode, feel a dipole force toward the highest intensity, resulting in focusing [220].

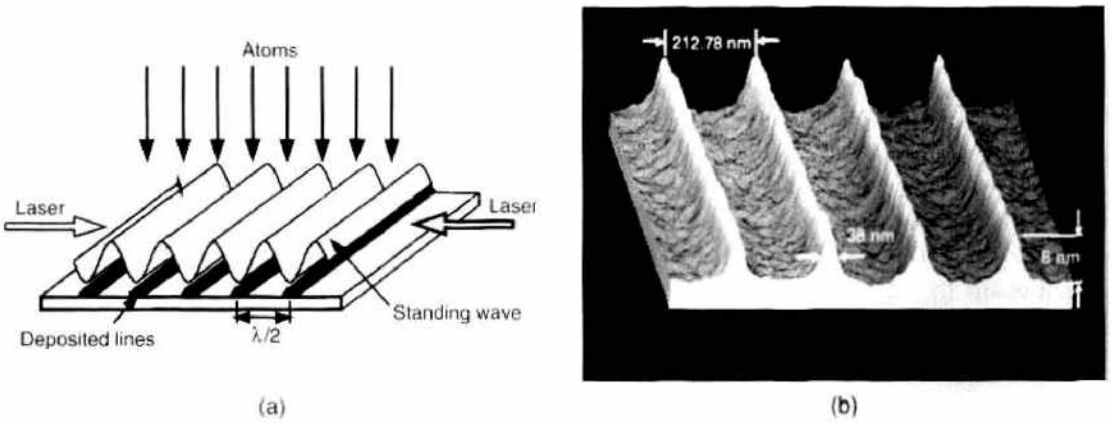


Figure 47. (a) Schematic representation focusing of chromium in standing wave. (b) Image of Cr lines on Si substrate. Lines are 50 nm wide, spread 212 nm [200].

apart. One-dimensional optical molasses is produced in the atomic beam before it crosses the standing wave. In the molasses region, the atoms are cooled transversely to a temperature of 76 μK and made to have an angular divergence of 0.3 mrad. This small angular divergence permitted a sharp focusing of the atoms. The chromium atoms were deposited on a silicon substrate and observed with either a scanning electron microscope or an atomic force microscope. Figure 47(b) shows the image of Cr lines. The widths of the lines are about 50 nm. Authors of Ref. [221] managed to attain a much smaller line width (20 nm, contrast 10 : 1) in their experiment on focusing a beam of sodium atoms onto a silicon substrate.

The detailed theoretical standing-wave focusing analysis carried out in Ref. [201] demonstrated that the principal factor responsible for the impairment of the quality of focusing was the angular divergence of the incident atomic beam and that atoms could be focused into nanometer-size strips, some 3 nm wide, provided that the angular divergence was small enough. Similar results (line width around 6.5 nm) were obtained in modeling in [222].

6.4.2. Combinations of Standing Light Waves

The superposition of several standing light waves of the same frequency gives rise to a more complex distribution of the optical-field intensity maxima and minima. Where the number of the laser beams used is relatively small, the field pattern and its symmetries remain simple and easy to recognize. For instance, an orthogonal arrangement of two counterrunning pairs of laser beams creates a square array of dots [223] (Fig. 48), and a 120-degree configuration of three laser beams produces a honeycomb-type structure of sixfold symmetry [224]

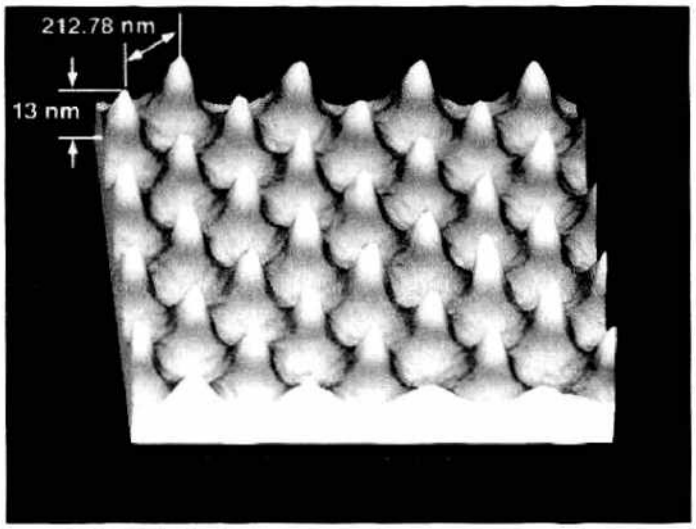


Figure 48. Square array of dots obtained by focusing of atoms [223].

(Fig. 49). By using a still greater number of laser beams, one can achieve an even more complex field intensity distribution, which will in turn lead to the focusing of the atomic beam into a specified set of dots with a micron-high resolution [225]. In Ref. [226] is analyzed the possibility of using arrays of microlenses to produce optical fields for focusing and confining atoms.

In recent years, many experimental works have been conducted on the focusing of atomic beams by means of various standing light-wave modifications. These works were aimed at improving resolution and finding the conditions necessary for the focusing of various atoms [221, 228, 229].

6.4.2.1. Superposition of Several Standing Light Waves Differing in Frequency The possibility of focusing atoms into narrow strips a few nanometers wide [201, 221, 222] makes it possible to pose the question about the formation of a denser, compared with the optical wavelength, pattern of strips of focused atoms. In Ref. [230] it was theoretically demonstrated that the use of three standing light waves differing in frequency detuning relative to the atomic transition frequency could allow an atomic beam to be focused into a set of strips spaced $\lambda/(2n)$ apart, where λ is the optical radiation wavelength and n is an integer.

6.4.2.2. Use of the Talbot Effect To focus atoms into nanometer-size regions regularly spaced at distances less than the wavelength of light, use can also be made of the Talbot effect [231]. This effect was discovered in 1836 and suggests that light passing through a coarse diffraction grating with a spacing greater than the wavelength of light has a very complex structure containing periodic structures with periods less than the spacing of the grating.

The first experiments on the study of the Talbot effect, as applied to atom optics, showed it to be possible to produce, with a microfabricated grating, periodic atomic structures with a spacing equal to that of the grating [232, 233] and even smaller than the length of the standing light wave [234].

In Ref. [235] the possibility of using *only* optical fields to make use of the Talbot effect in atom optics was analyzed. They suggested using only two standing light waves to focus metastable atoms into strips spaced at intervals much shorter than the length of the standing optical wave. One standing wave focuses the atoms into narrow strips spaced $\lambda/2$ apart. In the focusing plane, the atoms pass through the nodes of another standing light wave tuned to resonate with the atomic transition in the metastable atom. The atoms remain in their metastable state only if they pass through the nodes of the latter standing wave, which thus serves as an amplitude grating with a low aperture-to-spacing ratio. The result of the Talbot effect is that the atoms, after passing through the second standing wave, gather in narrow strips spaced at intervals of λ/n at a distance of L_T/n (where L_T is the Talbot distance) from the wave (Fig. 50).

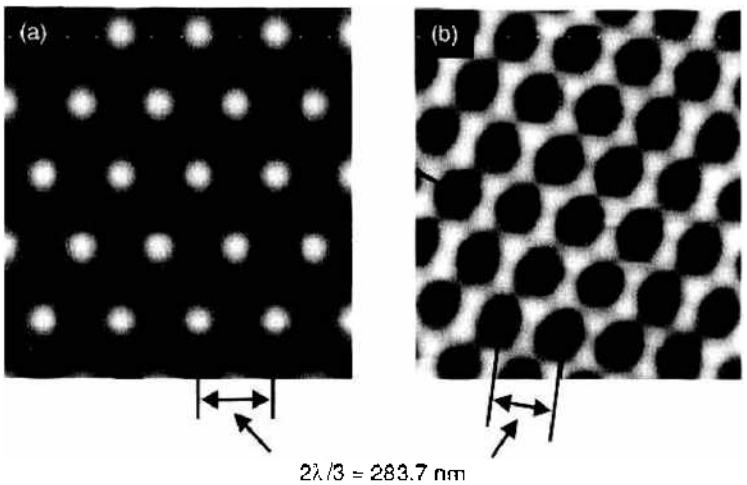


Figure 49. Honeycomb-type structures obtained by focusing of atoms [224].

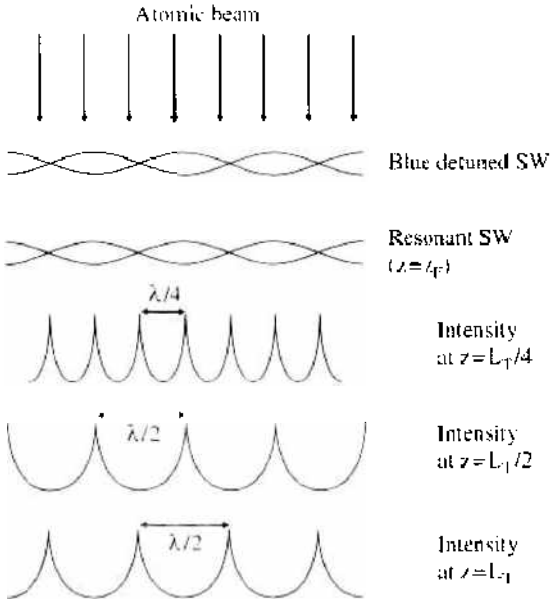


Figure 50. Scheme of using of Talbot effect for atom focusing.

To focus atoms into arbitrary nanometer-size structures, it is apparently necessary, after all, to use light beams with however great wave vectors, which is impossible to achieve with the aid of optical fields in free-space regions. Nevertheless, the necessary field configurations can be produced in the vicinity of material objects of subwavelength size (i.e., by means of near fields).

6.5. Focusing by Laser Near Fields

6.5.1. Near Field of a Nanotip

Optical fields with nanometer-size localization regions can be produced by means of material bodies whose size is smaller than the optical wavelength. A most simple configuration of such a *nanofield* is the field formed near a pointed tip diffracting a plane wave (Fig. 51). This field is localized in the vicinity of the tip and has an amplitude much in excess of that of the incident field. The enhancement originates from the electrostatic lightning-rod effect due to the geometric singularity of sharply pointed structures. Such fields are being widely used in scanning microscopy of nanometer-high resolution [91, 89, 236]. Specifically it was suggested using optical fields of this kind to produce nanopincers for individual atoms [237].

In the case of negative field frequency detuning, a beam of atoms passing through the nanometer-size region of high-field intensity will be pulled into it so that the focusing of the atoms may result. The focusing element, although small on the spatial scale, suffers from a shortcoming due to the spontaneous decays of the atoms passing through the region occupied by the incident radiation and the region of high-field strength in the vicinity of the tip. Moreover, the focal structure may prove to be complicated because of the diffraction of the atoms by the tip.

6.5.2. Near Field of a Nanohole

Another optical near field configuration capable of focusing atomic beams originates on the diffraction of light by a round nanohole [238–242]. Figure 52 shows the intensity distribution of the optical field produced on the diffraction of a circularly polarized plane standing light wave by a small hole. It can be seen from this figure that in the region before the hole an intensity maximum is formed at the antinode of the wave, which will act, in the case of negative optical field frequency detuning, as an atomic nanolens with a radial size of the same order of magnitude as that of the hole (Fig. 52). This configuration is again disadvantageous in that the atoms will propagate, before or after being focused, in the field of the

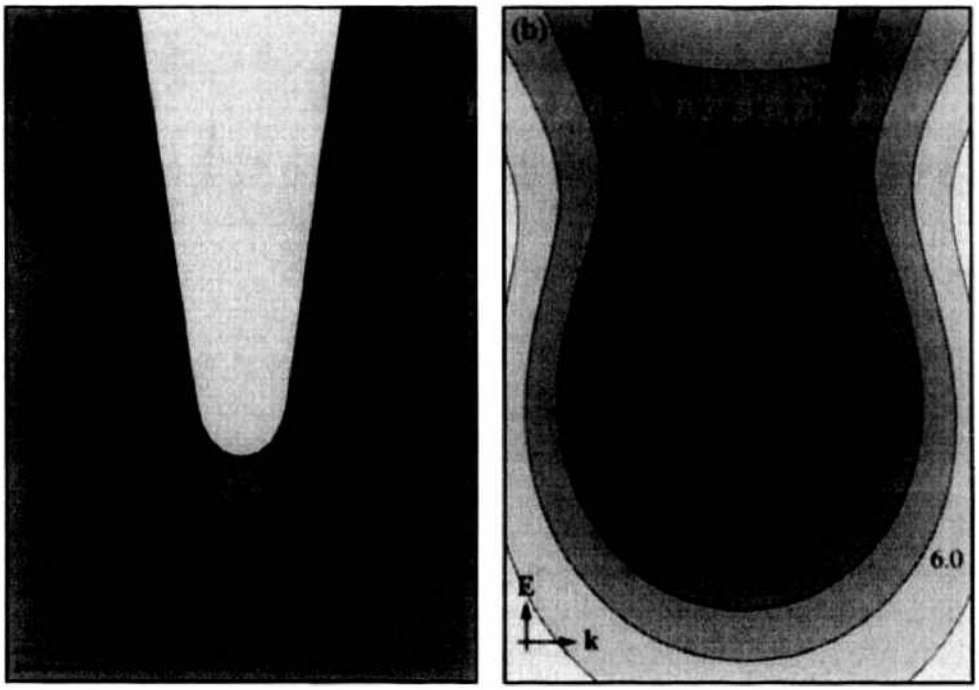


Figure 51. Near field of a gold tip in water illuminated by two different monochromatic waves at $\lambda = 810$ nm. Direction and polarization of the incident wave are indicated by the k and E vectors. The figures show contours of E^2 (factor of 2 between successive lines). The scaling is given by the numbers in the figures (multiples of the exciting field). No enhancement at the tip in (a); enhancement of about 3000 in (b). The field in (b) is almost rotationally symmetric in the vicinity of the tip [237].

standing light wave, where spontaneous decay processes, undesirable in many cases from the standpoint of atom optics purposes, will take place.

The near field in the vicinity of a hole features a true three-dimensional maximum (see Fig. 53) that in the case of negative optical field frequency detuning will focus atomic beams. A remarkable feature of this near field is that its potential can be described with a fairly high accuracy by a quadratic law, which bears witness to the smallness of spherical aberrations and a high quality of focusing. The detailed analysis of the focusing of an atomic beam by such a near field was performed in Refs. [241, 242] within the framework of the classical dynamics

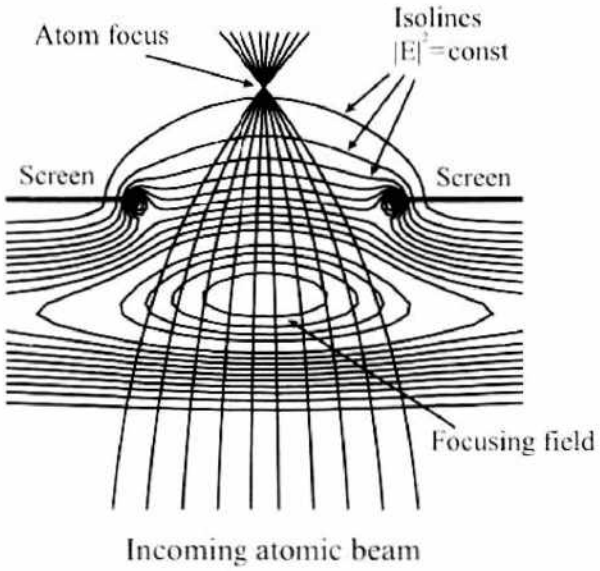


Figure 52. Distribution of the intensity of a circularly polarized wave normally incident on the nanoaperture and focusing scheme.

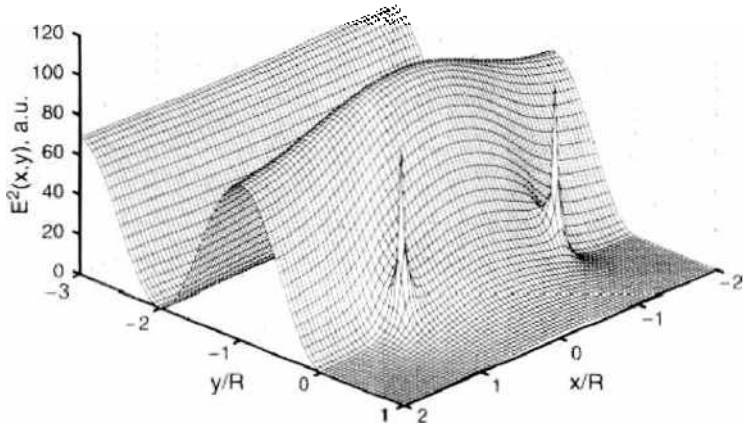


Figure 53. Spatial distribution of the intensity of a circularly polarized wave normally incident on the nanoaperture with radius R.

and in Ref. [239] within the framework of a stationary Schrödinger equation. Figure 54 shows the intensity distribution of a monochromatic beam of He atoms in the focal region. One can see from this figure that this distribution qualitatively resembles the Pearcy distribution [243]. The focusing quality can as usual be estimated by the radial width of the peak at half-maximum intensity. It is also seen from the figure that the radial width $\delta r_{1/2} \approx 0.0075a$ (around 2.585 nm at the parameter values used by Ref. [239]). With the solutions of the Schrödinger equations for monochromatic noncoherent beams known, it is not very difficult to take account of the effect of the nonmonochromaticity of the beam on the quality of focusing. If the atomic velocity distribution function has the form $f(v)$, the beam intensity in the focal spot region will be described by the superposition of the intensities of atoms moving with different velocities, that is, by the expression

$$P(r, z) = \int dv f(v) |\psi(v, r, z)|^2 \tag{160}$$

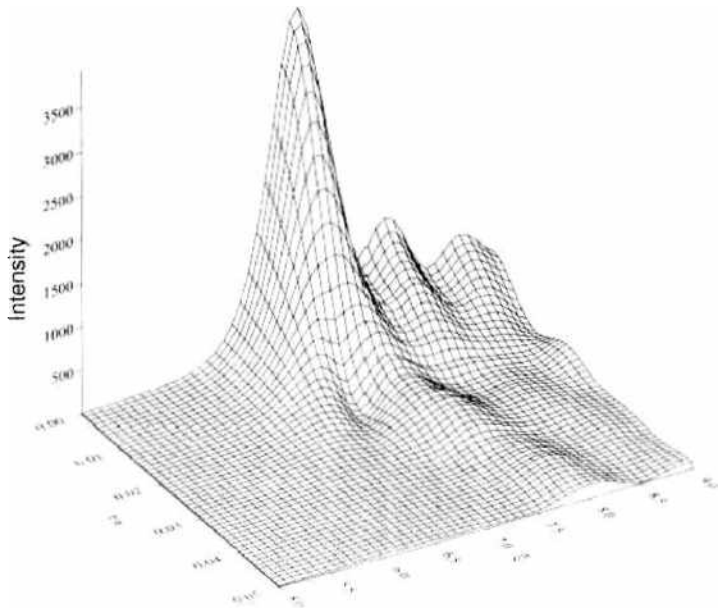


Figure 54. Three-dimensional distribution of the intensity of a monochromatic beam of He atom in the vicinity of the focal spot. Atomic and laser beams propagate in opposite directions ($kd = 2$, $G = 6 \cdot 10^5$, $2\Delta/\gamma = -2\sqrt{2} \cdot 10^5$, $r = 100$ m e).

To study chromatic aberrations, the atomic velocity distribution function was approximated by the expression

$$f(v) = (0.25\delta(v_z - v_0 + \delta v) + 0.5\delta(v_z - v_0) + 0.25\delta(v - v_0 - \delta v))\delta(v_x)\delta(v_y) \qquad (161)$$

where δv characterizes the degree of monochromaticity of the beam in the axial direction (the beam is taken to be monochromatic in the radial direction). The characteristic case considered was $2\delta v/v = 0.05$. The computation results [239] are presented in Fig. 55. Comparison between this figure and Fig. 54 shows that though chromatic aberrations substantially smooth out the diffraction picture the focal spot radius at half-maximum intensity remains practically unchanged: $\delta r_{1/2} = 0.075a$ (2.585 nm).

Thus it was found that an atomic beam could be focused, with reasonable parameter values, into a spot around 3 nm in radius at the half maximum intensity level. The main contribution to the focal spot size has been found to come from spherical aberrations and diffraction effects, the chromatic aberrations of the atomic beam with a 5% velocity distribution broadening having a lesser effect. Note that the above estimates were made for completely collimated beams. The angular divergence of actual beams can somewhat impair the quality of focusing.

6.5.3. Nanolocalized Fields in Waveguides (Photon Dots and Photon Holes)

The focusing schemes using the near fields produced by pointed tips and nanometer-size holes suffer from the disadvantage associated with the passage of the atoms being focused through high-intensity field regions, which gives rise to spontaneous decay processes and impairs the quality of focusing. A scheme providing for the production of a spatially localized optical *nanofield* free from this disadvantage [217] is presented in Fig. 56(a). Two flat conductive screens spaced at a distance of the order of or less than the wavelength of light, $d \leq \lambda$, form a plane two-dimensional waveguide for the laser radiation introduced into it from one side. It is well known [244] that for a waveguide consisting of two parallel ideally

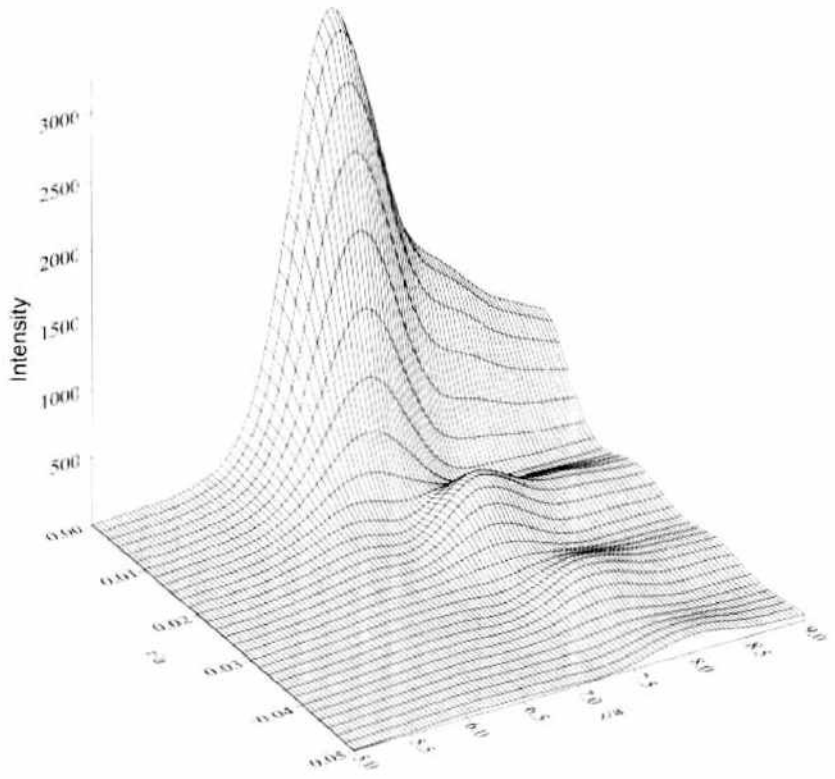


Figure 55. Three-dimensional distribution of the intensity of a nonmonochromatic beam of He atom in the vicinity of the focal spot. Atomic and laser beams propagate in opposite directions ($2\delta v/v = 0.05$, $ka = 2$, $G = 6 \cdot 10^8$, $2\Delta/\gamma = -2\sqrt{2} \cdot 10^4$, $v_z = 100$ m/s).

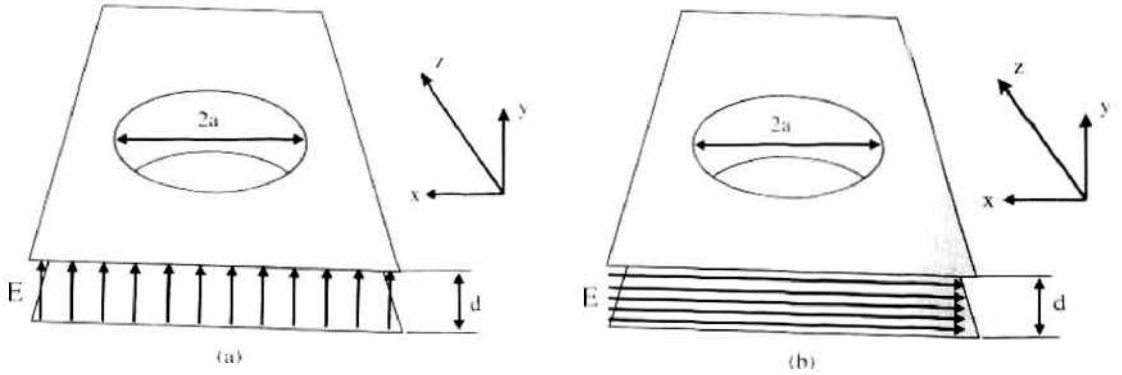


Figure 56. Geometry of a photon hole (a) and a photon dot (b).

conductive plates, there exist solutions of Maxwell's equations that permit the propagation of radiation through the waveguide, no matter how small its thickness d, be it even substantially smaller than the radiation wavelength. These solutions inside the waveguide correspond to a plane wave with its electric-field strength vector normal to the plates. Actually, this system is a two-wire line and provides for the two-dimensional nanometer-region localization of light [244].

Now let two small coaxial holes of radius $a \ll \lambda$ be made in the conductive screens (Fig. 56[a]). If the diameters of the holes are substantially smaller than the wavelength of the laser radiation introduced between the screens, the radiation will practically fail to issue from the holes but will be strongly modified near each of them while passing along the waveguide. In the vicinity of the holes, there actually takes place a reduction of the field strength in a region with a characteristic spatial size of the order of the hole diameter, that is, substantially smaller than the radiation wavelength λ. The volume of this region is $V \sim a^2 d \ll \lambda^3$. It is natural to call such a field modification a "photon hole."

To find the electric field distribution in the vicinity of the holes in the waveguide walls is a complex electrodynamics problem. In the particular case of nanohole ($a \ll \lambda$) under consideration, the problem reduces to a quasi-static one. The general solution of this problem reduces to the solution of an integral equation [245]. If the thickness d of the waveguide is much greater than the hole diameter, $d \gg a$, the mutual influence of the holes can be disregarded so that the problem is reduced to the superposition of the fields due to diffraction by a single hole. The problem on the modification of a uniform field by a conductive plate with a round hole has an analytical solution. As a result, the expression for the potential φ describing the electric field near the hole, $\mathbf{E} = -\nabla\varphi$, assumes the form [217]

$$\begin{aligned}
\varphi(\rho, y) &= \varphi^+\left(\rho, y - \frac{d}{2}\right) + \varphi^-\left(\rho, y + \frac{d}{2}\right) \\
\varphi^\pm(\rho, z) &= -\frac{E_0}{2}(z \mp d) \pm \frac{E_0}{\pi}|z|\left(\frac{1}{\mu(\rho, z)} + a\tan\mu(\rho, z)\right) \\
\mu(\rho, z) &= \sqrt{\frac{1}{2}\left(\frac{\rho^2}{a^2} + \frac{z^2}{a^2} - 1\right) + \frac{1}{2}\sqrt{\left(\left(\frac{\rho^2}{a^2} + \frac{z^2}{a^2} - 1\right)^2 + 4\frac{z^2}{a^2}\right)}}
\end{aligned} \tag{162}$$

where $\rho^2 = x^2 + z^2$.

Figure 57 shows the energy density distribution of the electric field near the holes. One can see from this figure that in the vicinity of the holes there actually forms a "photon hole" or, to be more exact, a "photon saddle" with a characteristic size determined by the size of the hole and the thickness of the waveguide.

Let us now consider one more method to localize fields in nanometer-size regions (Fig. 56[b]). Localization by this method is a generalization of the field localization near a nanohole [238–242], but is free from the shortcoming due to the presence of the standing light wave field. To this end, consider once more two ideally conductive plates with holes but

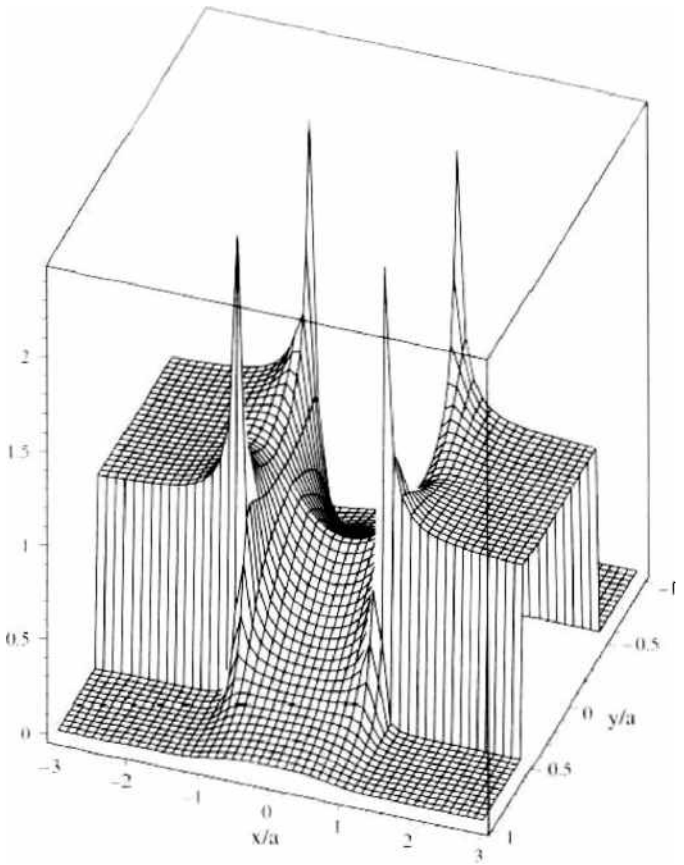


Figure 57. Spatial distribution of intensity of electromagnetic field near photon hole at $a/d = 1$.

now spaced at a distance of $d = \lambda/2$. Physically, the solution of Maxwell's equations in this case corresponds to a standing wave between the plates, whose wave vector is normal to the plates (directed along the y-axis). This is part of the standing wave formed upon reflection of a plane wave normally incident on one of the plates. By virtue of the condition $d = \lambda/2$, the other plate finds itself in a node of this standing wave, and so it has no effect on the latter.

If the holes are small in comparison with the wavelength, their effects are independent of each other, so that one can use, as a first approximation, the solution of the problem on diffraction by a single hole [238–242]. Figure 58 presents the field-intensity distribution near

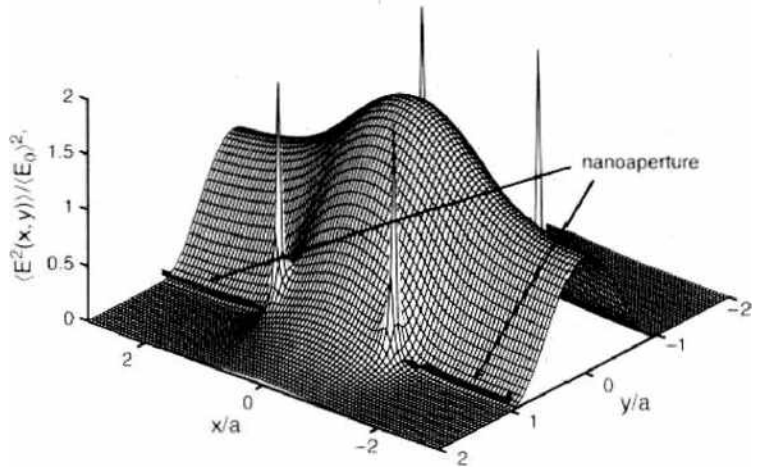


Figure 58. Spatial distribution of intensity of electromagnetic field near photon dot at $a/d = 0.5$.

the holes in a plane waveguide and inside it in the case where the thickness of the waveguide is equal to half the wavelength and the radius of the holes is $a = \lambda/4$. As can be seen from the figure, the field decreases rather rapidly outside of the waveguide in the direction normal to the waveguide plane and has a maximum in the center of the waveguide (i.e., there is formed a "photon dot"). The characteristic volume of such a "photon dot" is $V \sim \frac{\lambda}{2}a^2 \ll \lambda^3$. The sharp field-intensity peaks near the hole edges result from the assumption of the infinite conductivity of the waveguide walls. In waveguides with a finite wall conductivity, the amplitude of the peaks will be not so prominent. It is very important to note that the magnitude of the maximum (reckoned from that in the case where the holes are absent) at $x = y = 0$ is twice that in the case of a single hole. This circumstance is due to the structural interference of the optical fields scattered by the holes and makes it possible to use weaker fields as compared to those in the case of a single hole.

Let us now consider the possibility of using nanolocalized fields to focus atomic beams by the gradient force proportional to the intensity of the electric field. In the case of positive detuning of the laser radiation frequency relative to the atomic transition frequency, atoms are expelled into the weaker field regions, whereas in that of negative frequency detuning, they are pulled into the stronger field regions.

In the case of "photon hole," the region of the weaker fields is surrounded by the strong field inside the waveguide, and in the case of positive frequency detuning, the atoms flying through the holes in the waveguide walls will be attracted toward the axis of the system (i.e., their focusing will take place). As already noted, it is very important that the motion of the beams being focused occurs mainly in the region of weak fields, and so the probability of spontaneous decays causing the defocusing of the beams is, in that case, extremely low.

A "photon dot" in the case of negative frequency detuning pulls in the atoms, and there again occurs their focusing. In the case of "photon dot," atoms moves through a field region with an increased intensity, and so the probability of spontaneous decays here is higher than in the case of "photon hole." However, the time of flight of the atoms through nanometer-size regions is short, so that the effect of spontaneous decays on their focusing can again be neglected.

Note that thanks to the nanometer-scale size of "photon holes" and "photon dots," it is possible to produce their arrays of arbitrary geometry. Such arrays can find application in various schemes of nanofabrication with the aid of atomic beams.

7. CURRENT STATUS AND PROSPECTS OF ATOMIC NANOFABRICATION

Previous sections are largely devoted to the theoretical fundamentals of atom nanooptics that form a practically invariable basis for further development. The experimental implementation of atomic nanofabrication on the basis of atom optics is still at the initial stage of experimental demonstrations. But even now one can single out the main lines of successful experimentation and some prospects. And it is this matter that the concluding section of the present chapter is devoted to.

7.1. Deposition of Atoms for Physical Modification of Surface

The simplest version of atomic nanofabrication is the passing of a collimated atomic beam through a nonmaterial mask in the form of a standing light wave in the geometry shown in Fig. 47(a), that is, the channeling of atoms along the nodes or antinodes of the standing light wave [36]. The first experiment on the use of this effect was conducted with a beam of Na atoms incident upon a glass substrate [198]. The experiment was performed in a vacuum because of the chemical activity of sodium in the air.

The next important achievement was the experiment [200] performed with Cr atoms, wherein a standing light wave helped obtain surface structures (stripes) less than 100 nm in width (28 nm in the best experiments) spaced $\lambda/2 = 213$ nm apart.

This method was then extended to other elements, Al in particular [228, 247, 248], Yb [229]. Table I lists the best results obtained for various atoms by the direct deposition through

Table 1. Best feature values obtained for various atoms used in direct deposition (from Meschede and Metcalf [249]).

Atom	Substrate	Period (nm)	Width (nm)	Height (nm)	Contrast[a]	Ref.
Na	Si	295	20	1.5	~ 0.9	Timp et al. [198]; Natarajan et al. [262]
Cr	Si	213	29	47	~ 0.8	McClelland et al. [200]; Anderson et al. [263]
Al	Si	155	70	2	> 15	McGowan et al. [248]; Rehse et al. [228]

[a]Contrast = height/(height + pedestal).

a standing laser wave "mask," with the parameters of the stripe grating (Fig. 59) being indicated. The data were borrowed from latest review of the atomic nanofabrication by Ref. [249]. This technique is quite applicable to other atoms (Fe, Ga, Ag, In, Au, etc.) as well. We would like to emphasize that in the direct deposition method the atomic beam produces a physical change in the surface topography of the substrate without modifying it chemically.

7.2. Deposition of Metastable Atoms for Chemical Modification of Surface

Another trend in the atomic nanofabrication is based on the use of a beam of atoms excited to a long-lived metastable state. Such atoms (usually noble gas atoms) have a surplus electronic energy that may cause a chemical modification of a suitable substrate (an organic resist layer). This method resembles the ordinary photolithography technique because it uses the etching of the chemically modified sites. The substrates used are suitable, self-assembled monolayers. The first successful experiment along these lines was conducted with a beam of metastable argon atoms [250]. The release of around 12 eV of energy into a molecular layer of a hydrocarbon produced a durable carbonaceous material excellently amenable to wet or dry etching. In this experiment, nanopatterns with an edge resolution better than 100 nm were produced on gold films. Note that both in this and in the next experiment [251]. As distinct from the experiments with unexcited atoms, the standing light wave in this method does not change the trajectories of the atoms, but affects the de-excitation of the metastable atoms via the high-lying quantum states. With the intensity of the deactivating standing light wave being high enough, only those atoms remain in their metastable state, which fly along the nodes of the wave, where its intensity is very low. As a result, incident upon the self-assembled monolayer are narrow strips of metastable Kr* atoms that cause the appropriate chemical modifications. Experiments of this kind made possible the implementation of nanolithography at the Heisenberg limit (less than 100 nm) [251].

Later on similar experiments were performed with beams of metastable neon atoms [227] and metastable helium atoms [252–254]. The edge resolution attained in these experiments amounted to 50–40 nm.

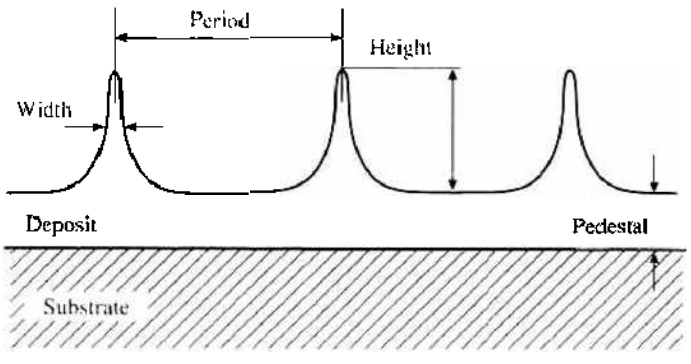


Figure 59. Characterization of deposited atomic stripes obtained by direct atomic-beam deposition through a standing laser light wave.

To carry out atomic nanofabrication with metastable triplet atoms, use is made of various resist/surface combinations. More detailed information on the experiments by the atomic and direct deposition nanofabrication techniques can be found in the comprehensive review by Meschede and Metcalf [249].

We would like to stress that to use standing light waves effectively as a "nonmaterial mask," use should be made of well-collimated atomic beams. The first experiments on the collimation of atomic beams by way of their transverse laser cooling had already been performed [255] before the first experiments on the atomic nanofabrication were staged. And the elements of these pioneering works have now become indispensable.

7.3. Complex Atomic Nanofabrication Patterns

The majority of experiments on the atom-optical nanofabrication were conducted with standing light waves producing grating patterns with a period of $\lambda/2$ (i.e., on the scale of a few hundred nanometers). Using crossing standing light waves can help obtain two-dimensional periodic structures (arrays) [223], and three standing light waves intersecting at an angle of 120° (a hexagonal optical lattice) make it possible to produce periodic honeycomb structures [224].

The structure periodicity associated with the half-wavelength $\lambda/2$ of laser light can be changed. By using a standing light wave produced by two counterrunning waves linearly polarized at right angles to each other (lin $\perp$ lin), Gupta and co-workers [256] obtained Cr dots with a period of $\lambda/8$. And a two-dimensional "light mask" helped Brezger and co-workers [257] to produce two-dimensional chromium patterns with a period of $\lambda/3$.

Combinations of several standing light waves intersecting at various angles and differing in polarization have considerable potential for atomic nanofabrication.

7.4. Atomic-Beam Holography for Nanofabrication

Optical holography that allows an optical wavefront to be reproduced by passing an optical beam through a hologram is well known. A hologram can be produced by photographically recording two interfering optical beams (a reference beam and a beam reflected from the object of interest). The optical holography principle has been known to be extended to electron holography [258]. The progress in the production of laser-cooled atomic beams made it possible to obtain a holographic image by means of Ne atoms and a computer-generated binary hologram [259, 260]. A laser-cooled (50 μK) beam of metastable Ne^* atoms was passed through a hologram produced on a 100-nm-thick silicon nitride membrane. The hologram comprised numerous holes (typically 200 nm across) in the membrane, their positions precisely specified by the computer, so that the diffracted Ne atoms generated the reconstructed pattern. Figure 60 presents the result of reconstructing the pattern of "atom

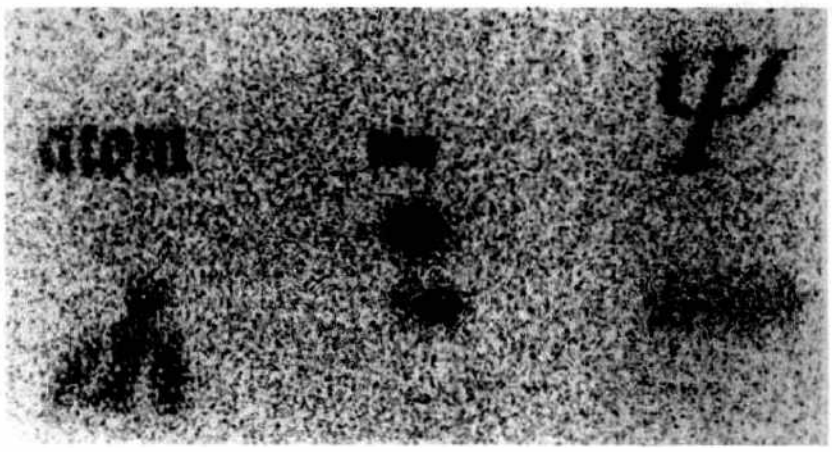


Figure 60. Reconstructed pattern of "atom Ne ψ" (?). The reconstructed pattern of Ne atoms is detected by using a microchannel plate (MCP) detector. The position of each atom detected by the MCP is recorded and accumulated by a computer. The length of ψ is about 4 nm [261].

Ne *et al.*" In this experiment, use was made of Ne atoms and a microchannel plate (MCP) detector.

In principle, many atoms can be used for atomic-beam holography purposes. It is possible to deposit these atoms directly on a substrate to produce a desired pattern with a theoretical resolution of about 200 nm under typical conditions. Atomic-beam holography has considerable potential for the production of patterns with a nanometer-scale resolution (i.e., for atomic nanofabrication).

7.5. Prospects

We are now at the initial stage of development of atom-optical nanofabrication techniques. Many interesting proposals have been put forward, but the main experiments have been conducted with standing laser light waves ("light masks"). In the preceding sections, we have described the prospects for the guiding of atoms, their near-field focusing, and so on. On the basis of these proposals, one can expect the development of an "atomic writing pencil" for atomic nanofabrication purposes. It is quite possible that nanostructures will in the future be formed in this way, atom by atom. There is a certain potential in using pulsed atomic beams in step with pulsed laser fields [264]. The rapid development of the technology of femtosecond laser pulses recurring with frequencies of a few hundreds of megahertz (i.e., multiple-frequency coherent laser fields) should also play a certain part in new atomic nanofabrication techniques. It might be expected that atom optical nanofabrication will reach maturity in the next decade.

REFERENCES

M. Born and E. Wolf, "Principles of Optics," Pergamon, Oxford, 1984.
V. I. Balykin and V. S. Letokhov, "Atom Optics with Laser Light," Harwood, 1995, p. 115.
P. Meystre, "Atom Optics," Springer, New York, 2001, p. 311.
P. Grivet, "Electron Optics," 2nd Edn. Oxford University, 1972.
E. Ruska, "The Early Development of Electron Lenses and Electron Microscopy," S. Hirzel, Stuttgart, 1980.
V. F. Sears, "Neutron Optics," Publ., New York, 1989.
F. L. Shapiro, "Neutron Research," Nauka, Moscow, 1976.
U. Bonze and H. Rauch, "Neutron Interferometry," Clarendon, Oxford, 1979.
F. Shutz, A. Steyerl, and W. Mampe, *Phys. Rev. Lett.* 44, 1400 (1980).
A. I. Frank, *Usp. Fiz. Nauk* 151, 229 (1987). [Sov. Phys.—Usp. 30, 110].
A. I. Frank, *Usp. Fiz. Nauk* 161, 95 (1991). [Sov. Phys.—Usp. 34, 980].
N. F. Ramsey, "Molecular Beams," Clarendon, Oxford, 1956.
F. Knamer and O. Stern, *Zs. Phys.* 53, 799 (1929).
I. Estermann and O. Stern, *Zh. Phys.* 61, 95 (1930).
V. U. Nayak, D. O. Edwards, and N. Masuhara, *Phys. Rev. Lett.* 50, 990 (1983).
A. Anderson, S. Haroche, E. A. Hinds, W. Jhe, D. Meschede, and L. Moi, *Phys. Rev. Lett. A* 34, 3513 (1986).
O. Stern, *Naturwiss* 17, 391 (1929).
O. R. Frish and O. Stern, *Zs. Phys.* 84, 430 (1933).
D. W. Keith, M. L. Schattenburg, H. I. Smith, and D. E. Pritchard, *Phys. Rev. Lett.* 61, 1580 (1988).
J. J. Berkhout, O. J. Luiten, I. D. Setija, T. W. Hijmans, T. Mizusaki, and J. T. M. Walraven, *Phys. Rev. Lett.* 63, 1689 (1989).
O. Carnal, A. Faulstich, and J. Mlynek, *Appl. Phys. B* 53, 88 (1991).
O. Carnal and J. Mlynek, *Phys. Rev. Lett.* 66, 2689 (1991).
D. W. Keith, C. R. Ekstrom, Q. A. Turchette, and D. E. Pritchard, *Phys. Rev. Lett.* 66, 2693 (1991).
H. Friedburg and W. Paul, *Naturwiss* 37, 20 (1950).
H. Friedburg and W. Paul, *Naturwiss* 38, 159 (1951).
M. I. Korsynskii and Y. M. Fogel, *Zh. Eksp. Teor. Fiz.* 21, 25 (1951); 21, 38 (1951).
R. Vanthier, *Cp. Rd.* 228, 1113 (1949).
G. Benewitz, W. Paul, and C. Schlier, *Zs. Phys.* 141, 6 (1955).
H. M. Goldenberg, D. Kleppner, and N. F. Ramsey, *Phys. Rev. Lett.* 8, 361 (1960).
H. M. Goldenberg, D. Kleppner, and N. F. Ramsey, *Phys. Rev. Lett.* 126, 603 (1962).
J. P. Gordon, H. J. Zeiger, and C. Townes, *Phys. Rev.* 95, 282 (1954).
J. P. Gordon, H. J. Zeiger, and C. Townes, *Phys. Rev.* 99, 1264 (1955).
R. B. Bernstein, "Chemical Dynamics via Molecular Beam and Laser Techniques," Clarendon, Oxford, 1982.
G. I. Opat, S. J. Wark, and A. Cimmino, *Appl. Phys. B* 54, 396 (1992).
G. A. Askarian, *Sov. Phys. JETP* 15, 1088 (1962).
V. S. Letokhov, *Pis'ma Zh. Eksp. Teor. Fiz.* 7, 348 (1968). [Sov. Phys. JETP Lett. 7, 272 (1968)].

37. J. E. Bjorkholm, R. E. Freeman, A. A. Askhin, and D. B. Pearson, *Phys. Rev. Lett.* 41, 1361 (1978).
38. A. Einstein, *Phys. Z.* 10, 185 (1909); 18, 121 (1917).
39. T. Hänsch and A. Schawlow, *Opt. Commun.* 13, 68 (1975).
40. D. J. Wineland and T. W. Dehmelt, *Bull. Amer. Phys. Soc.* 20, 637 (1975).
41. A. Ashkin, *Science* 210, 1081 (1980).
42. V. S. Letokhov and V. G. Minogin, *Phys. Rep.* 73, 1 (1981).
43. H. Dehmelt, in "Advances in Laser Spectroscopy" (F. T. Arrechi, F. Strumia, and H. Walther, Eds.), p. 153, Plenum, New York, 1983.
44. S. Stenholm, *Rev. Mod. Phys.* 58, 699 (1986).
45. D. J. Wineland and W. M. Itano, *Phys. Today* 6, 34 (1987).
46. V. I. Balykin and V. S. Letokhov, *Phys. Today* 4, 23 (1989).
47. C. Cohen-Tannoudji and W. D. Philips, *Phys. Today* 10, 33 (1990).
48. D. E. Pritchard and B. G. Oldaker, in "Coherence and Quantum Optics VI" (J. H. Eberly, L. Mandel, and E. Wolf, Eds.), p. 937, Plenum, New York, 1990.
49. S. Chu, *Science* 253, 861 (1991).
50. C. Cohen-Tannoudji, in "Fundamental Systems in Quantum Optics," Elsevier, B.V., 1991
51. P. Meystre and S. Stenholm, *J. Opt. Soc. Am. B* 2, 11 (1985).
52. S. Chu and C. Wieman, *J. Opt. Soc. Am. B* 6, 11 (1989).
53. J. Mlynek, V. I. Balykin, and P. Meystre, *Appl. Phys. B* 54, 5 (1992).
54. V. G. Minogin and V. S. Letokhov, "Laser Radiation Pressure on Atoms," Gordon and Breach, New York, 1987.
55. A. P. Kazantsev, G. I. Surdutovich, and V. P. Yakovlev, "Mechanical Action of Light on Atom," World Scientific, Singapore, 1991.
56. H. J. Metcalf and P. van der Straten, "Laser Cooling and Trapping," Springer, New York, 1999, p. 323.
57. A. Ashkin, *Phys. Rev. Lett.* 25, 1321 (1970).
58. V. S. Letokhov, V. G. Minogin, and B. D. Pavlik, *Zh. Eksp. Teor. Fiz.* 72, 1318 (1977). [Sov. Phys. JETP 45, 698 (1977)].
59. D. J. Wineland and W. M. Itano, *Phys. Rev. A* 20, 1521 (1979).
60. A. Ashkin, *Phys. Rev. Lett.* 40, 729 (1978).
61. A. P. Kazantsev, V. S. Smirnov, G. I. Surdutovich, D. O. Chudesnikov, and V. P. Yakovlev, *J. Opt. Soc. Am. B* 2, 1731 (1985).
62. V. G. Minogin and O. T. Serimaa, *Opt. Commun.* 30, 373 (1979).
63. J. P. Gordon and A. Ashkin, *Phys. Rev. A* 21, 1606 (1980).
64. A. P. Kazantsev, *Zh. Eksp. Teor. Fiz.* 66, 1599 (1974). [Sov. Phys. JETP 39, 784 (1974)].
65. A. P. Kazantsev, G. A. Ryabenko, G. I. Surdutovich, and V. P. Yakovlev, *Phys. Rep.* 129, 75 (1985).
66. J. Dalibard and C. Cohen-Tannoudji, *J. Opt. Soc. Am. B* 2, 1707 (1985).
67. A. Aspect, J. Dalibard, A. Heidmann, S. Salomon, and C. Cohen-Tannoudji, *Phys. Rev. Lett.* 57, 1688 (1986).
68. A. P. Kazantsev, D. O. Chudesnikov, and V. P. Yakovlev, *Sov. Phys. JETP* 63, 951 (1986).
69. C. Cohen-Tannoudji, *C.R. Acad. Sci.* 252, 394 (1961).
70. T. R. Carver, *Phys. Rev.* 124, 800 (1961).
71. M. Mizushima, *Phys. Rev. A* 133, 414 (1964).
72. A. M. Bonch-Bruevich and A. A. Khodovoi, *Sov. Phys. Usp.* 10, 637 (1968).
73. C. Cohen-Tannoudji and S. Raymond, in "Multiphoton Processes" (J. Eberly and P. Lambropoulos, Eds.), p. 103, Wiley, City, 1978.
74. A. Aspect, E. Arimondo, R. Kaiser, N. Vansteenkiste, and C. Cohen-Tannoudji, *Phys. Rev. Lett.* 61, 826 (1989).
75. G. Nienhuis, P. van der Straten, and S.-Q. Shang, *Phys. Rev. A* 44, 462 (1993).
76. P. Lett, R. Watts, C. Westbrook, W. D. Phillips, P. Could, and H. Metcalf, *Phys. Rev. Lett.* 61, 169 (1988).
77. Y. Shevy, D. S. Weiss, P. J. Ungar, and S. Chu, *Phys. Rev. Lett.* 62, 1118 (1989).
78. J. Dalibard and C. Cohen-Tannoudji, *J. Opt. Soc. Am. B* 6, 2023 (1989).
79. B. Sheehy, S.-Q. Shang, R. Watts, S. Hatamian, and H. Metcalf, *J. Opt. Soc. Am.* 6, 2165 (1989).
80. B. Sheehy, S.-Q. Shang, P. van der Straten, S. Hatamian, and H. Metcalf, *Phys. Rev. Lett.* 64, 858 (1990).
81. S.-Q. Shang, B. Sheehy, P. van der Straten, and H. Metchalf, *Phys. Rev. Lett.* 65, 317 (1990).
82. P. J. Ungar, D. S. Weiss, E. Riis, and S. Chu, *J. Opt. Soc. Am. B* 6, 2058 (1989).
83. E. L. Raab, M. Prentiss, A. Cable, S. Chu, and D. Pritchard, *Phys. Rev. Lett.* 59, 2631 (1987).
84. R. Grimm, Y. B. Ovchinnikov, A. I. Sidorov, and V. S. Letokhov, *Opt. Commun.* 84, 18 (1991).
85. J. J. McClelland and M. P. Scheinfein, *J. Opt. Soc. Am. B* 8, 1974 (1991).
86. R. J. Cook and R. K. Hill, *Opt. Commun.* 43, 258 (1982).
87. Y. B. Ovchinnikov, S. V. Shul'ga, and V. I. Balykin, *J. Phys. B* 24, 3173 (1991).
88. M. A. Ol'shanii, Y. B. Ovchinnikov, and V. S. Letokhov, *Opt. Commun.* 98, 77 (1993).
89. D. W. Pohl, "Scanning Near-Field Microscopy (SNOM)," Academic, London, 1991.
90. M. A. Paesler and P. J. Moyer, "Near-Field Optics: Theory, Instrumentation and Applications," John Wiley, New York, 1996.
91. M. Ohtsu, "Near Field Nano/Atom Optics and Technology," Springer, New York, 1998.
92. V. S. Letokhov, *Contemp. Phys.* 36, 235 (1995).
93. H.-R. Noh, X. Xu, and W. Jhe, *Adv. At. Mol. Opt. Phys.* 48, 153 (2002).
94. F. Shimizu, K. Shimizu, and H. Takuma, in "Proceedings of International Symposium on Quantum Optics," Shanghai, 1992, p. 193.

95. V. V. Vladimirskii, *Sov. Phys. JETP* 12, 740 (1961).
96. J. Schmiedmayer, *Appl. Phys. B* 60, 169 (1995).
97. J. Schmiedmayer, *Phys. Rev. A* 52, R13 (1995).
98. J. Denschlag, D. Cassettari, A. Chenet, S. Schneider, and J. Schmiedmayer, *Appl. Phys. B* 69, 291 (1999).
99. J. Denschlag, D. Cassettari, and J. Schmiedmayer, *Phys. Rev. Lett.* 82, 2014 (1999).
100. J. Denschlag, G. Umshaus, J. Schmiedmayer, *Phys. Rev. Lett.* 81, 737 (1999).
101. J. Denschlag, D. Cassettari, A. Chenet, S. Schneider, and J. Schmiedmayer, *Appl. Phys. B* 69, 291 (1999).
102. L. Hau, M. Burns, and J. Golovchenko, *Phys. Rev. A* 45, 6448 (1992).
103. R. Folman, P. Krüger, D. Cassettari, B. Hessmo, T. Maier, and J. Schmiedmayer, *Phys. Rev. Lett.* 84, 4749 (2000).
104. W. H. Wing, *Prog. Quant. Electr.* 8, 181 (1984).
105. V. I. Balykin, V. G. Minogin, and V. S. Letokhov, *Rep. Prog. Phys.* 63, 1429 (2000).
106. J. Schmiedmayer and A. Scrinzi, *Phys. Rev. A* 54, R2525 (1996).
107. G. P. Pron'kov and Y. G. Stroganov, *Sov. Phys. JETP* 45, 1075 (1997).
108. R. Blümel and K. Dietrich, *Phys. Rev. A* 43, 22 (1991).
109. A. I. Voronin, *Phys. Rev. A* 43, 29 (1991).
110. L. Hau, J. Golovchenko, and M. Burns, *Phys. Rev. Lett.* 74, 3138 (1995).
111. J. P. Burke, Jr., C. H. Greene, and B. D. Esry, *Phys. Rev. A* 54, 3225 (1996).
112. K. Berg-Sørenson, M. Burns, J. Golovchenko, and L. Hau, *Phys. Rev. A* 53, 1653 (1996).
113. S. Kuppens, M. Rauner, M. Schiffer, G. Wokurka, T. Slawinski, M. Zinner, K. Sengstock, and W. Ertmer, in "OSA TOPS on Ultracold Atoms and BEC 1996" (K. Burnett, Ed.), Vol. 7. Publ., City, 1997.
114. J. A. Sauer, M. D. Barrett, and M. S. Chapman, *Phys. Rev. Lett.* 87, 270401 (2001).
115. R. Folman, P. Krüger, J. Schmiedmayer, J. Denschlag, and C. Henkel, *Adv. At. Mol. Opt. Phys.* 48, 263 (2002).
116. M. Key, I. G. Hughes, W. Rooijakkers, B. E. Sauer, and E. A. Hinds, *Phys. Rev. Lett.* 84, 1371 (2000).
117. E. A. Hinds and C. Eberlein, *Phys. Rev. A* 61, 033614 (2000).
118. E. A. Hinds, M. G. Boshier, and I. G. Hughes, *Phys. Rev. Lett.* 80, 645 (1998).
119. A. I. Sidorov, R. J. McLean, W. J. Rowlands, D. C. Lau, J. E. Murphy, M. Walkiewicz, G. I. Opat, and P. Hannaford, *Quant. Semiclass. Opt.* 8, 713 (1996).
120. I. M. Roach, H. Abele, M. G. Boshier, H. L. Grossman, K. P. Zetie, and E. A. Hinds, *Phys. Rev. Lett.* 75, 629 (1995).
121. I. G. Hughes, P. A. Barton, T. M. Roach, M. G. Boshier, and E. A. Hinds, *J. Phys. B* 30, 2119 (1997).
122. C. V. Saba, P. A. Barton, M. G. Boshier, I. H. Hughes, P. Rosenbuch, and E. A. Hinds, *Phys. Rev. Lett.* 82, 468 (1999).
123. E. A. Hinds and I. G. Hushes, *J. Phys. D* 32, R119 (1999).
124. J. D. Weinstein and K. G. Libbrecht, *Phys. Rev.* 52, 4004 (1995).
125. J. H. Thywissen, R. M. Westervelt, and M. Prentiss, *Phys. Rev. Lett.* 83, 3762 (1999).
126. J. Fortagh, A. Grossmann, C. Zimmermann, and T. W. Hänsch, *Phys. Rev. Lett.* 81, 5310 (1998).
127. B. K. Teo and G. Raithel, *Phys. Rev. A* 63, 031402 (2001).
128. H. Ott, J. Fortagh, G. Schlottergeck, A. Grossman, and C. Zimmerman, *Phys. Rev. Lett.* 87, 230401 (2001).
129. W. Hansel, P. Hommelhoff, T. W. Hansch, and J. Reichel, *Nature* 413, 498 (2001).
130. A. E. Leanhardt, A. P. Chikkatur, D. Kielpinski, Y. Shin, T. L. Gustavson, W. Ketterle, and D. E. Pritchard, *Phys. Rev. Lett.* 89, 040401 (2002).
131. M. G. Prentiss and S. Ezekel, *Phys. Rev. Lett.* 56, 46 (1986).
132. C. Salomon, J. Dalibard, A. Aspect, H. Metcalf, and C. Cohen-Tannoudji, *Phys. Rev. Lett.* 59, 1659 (1987).
133. V. I. Balykin, Y. E. Lozovik, Y. B. Ovchinnikov, A. I. Sidorov, S. V. Shulga, and V. S. Letokhov, *J. Opt. Soc. Am. B* 6, 2178 (1989).
134. C. M. Savage, S. Marksteiner, and P. Zoller, in "Fundamentals of Quantum Optics III: Proceedings of the Fifth Meeting on Laser Phenomena, University of Innsbruck, 1992" (F. Ehlotzky, Ed.), p. 60. Springer, Berlin, 1993.
135. S. Marksteiner, C. M. Savage, P. Zoller, and S. L. Rolston, *Phys. Rev. A* 50, 2680 (1994).
136. J. P. Burke, Jr., S.-T. Chu, G. W. Bryant, C. J. Williams, and P. S. Julienne (2002).
137. M. J. Renn, D. Mantgomery, O. Vdovin, D. Z. Anderson, C. E. Wieman, and E. A. Cornell, *Phys. Rev. Lett.* 75, 3253 (1995).
138. V. I. Balykin, V. S. Letokhov, A. I. Sidorov, and Y. B. Ovchinnikov, *JETP Lett.* 45, 353; *Pis'ma Zh. Eksp. Teor. Fiz.* 45, 282 (1987).
139. J. P. Dowling and J. Gea-Banacloche, *Adv. At. Mol. Opt. Phys.* 34, 1 (1996).
140. V. I. Balykin, *Adv. At. Mol. Opt. Phys.* 41, 181 (1999).
141. M. J. Renn, E. A. Donley, E. A. Cornell, C. E. Wieman, and D. Z. Anderson, *Phys. Rev. A* 53, R648 (1996); M. J. Renn, A. A. Zozulya, E. A. Donley, E. A. Cornell, and D. Z. Anderson, *Phys. Rev. A* 55, 3684 (1997).
142. H. Ito, K. Sakaki, T. Nakata, W. Jhe, and M. Ohtsu, *Opt. Commun.* 115, 57 (1995).
143. H. Ito, T. Nakata, K. Sakaki, M. Ohtsu, K. I. Lee, and W. Jhe, *Phys. Rev. Lett.* 76, 4500 (1996).
144. H. Ito, K. Sakaki, W. Jhe, and M. Ohtsu, *Phys. Rev. A* 56, 712 (1997).
145. X. Xu, V. G. Minogin, K. Lee, Y. Wang, and W. Jhe, *Phys. Rev. A* 60, 4796 (1999).
146. A. Landragin, J.-Y. Courtois, G. Labeyrie, N. Vansteenkiste, C. I. Westbrook, and A. Aspect, *Phys. Rev. Lett.* 77, 1464 (1996).
147. S. Kuppens, M. Rauner, M. Schiffer, G. Wokurka, T. Slawinski, M. Zinner, K. Sengstock, and W. Ertmer, in "OSA TOPS on Ultracold Atoms and BEC 1996" (K. Burnett, Ed.), Vol. 7. Publ., City, 1997.

148. J. Yin, H.-R. Noh, K. I. Lee, K.-H. Kim, Y.-Z. Wang, and W. Jhe, *Opt. Commun.* 138, 287 (1997).
149. L. Pruvost, D. Marescaux, O. Houde, and H. Duong, *Opt. Commun.* 166, 199 (1999).
150. K. Szymaniec, H. J. Davies, and C. S. Adams, *Europhys. Lett.* 45, 450 (1999).
151. C. Henkel, K. Molmer, R. Kaiser, N. Vansteenkiste, C. I. Westbrook, and A. Aspect, *Phys. Rev. A* 55, 1160 (1997).
152. V. I. Balykin and V. S. Letokhov, *Opt. Commun.* 64, 151 (1987).
153. G. M. Gallatin and P. L. Gould, *J. Opt. Soc. Am. B* 8, 502 (1991).
154. X. Wang and M. G. Littman, *Opt. Lett.* 18, 767 (1993).
155. R. M. Herman and T. A. Wiggins, *J. Opt. Soc. Am. A* 8, 932 (1991).
156. H. S. Lee, B. W. Atewart, K. Choi, and H. Fenichel, *Phys. Rev. A* 49, 4922 (1994).
157. C. Paterson and R. Smith, *Opt. Commun.* 124, 121 (1996).
158. M. W. Beijersbergen, L. Allen, H. E. L. O. van der Veen, and J. P. Woerdman, *Opt. Commun.* 96, 132 (1992).
159. J. Yin, W. Gao, and Y. Zhu, *Prog. Opt.* 45, 119 (2003).
160. N. R. Heckenberg, R. McDuff, C. P. Smith, H. RubinszteinDunlop, and M. J. Wegener, *Opt. Quant. Electron.* 24, S951 (1992).
161. J. Nellessen, J. Werner, and W. Ertmer, *Opt. Commun.* 78, 300 (1990).
162. E. Riis, D. S. Weiss, K. A. Moler, and S. Chu, *Phys. Rev. Lett.* 64, 1658 (1990).
163. C. M. Savage, S. Marksteiner, and P. Zoller, in "Fundamentals of Quantum Optics III: Proceedings of the Fifth Meeting on Laser Phenomena, University of Innsbruck, 1993" (F. Ehlotzky, Ed.), p. 60. Springer, Berlin, 1993.
164. A. H. Barnett, S. P. Smith, M. Olshanii, K. S. Johnson, A. W. Adams, and M. Prentiss, *A* 61, 023608 (2000).
165. J. D. Jackson, "Classical Electrodynamics." Wiley, New York, 1975.
166. R. Kaiser, Y. Levy, N. Vansteenkiste, A. Aspect, W. Seifert, D. Leipold, and J. Mlynek, *Opt. Commun.* 104, 234 (1994).
167. W. Seifert, C. S. Adams, V. I. Balykin, C. Heine, Y. Ovchinnikov, and J. Mlynek, *Phys. Rev. A* 49, 3814 (1994).
168. P. B. Johnson and R. W. Christy, *Phys. Rev. B* 6, 4370 (1972).
169. C. R. Bennett, J. B. Kirk, and M. Babiker, *Phys. Rev. A* 63, 033405 (2001).
170. W. Seifert, R. Kaiser, A. Aspect, and J. Mlynek, *Opt. Commun.* 111, 566 (1994).
171. P. K. Tien and R. Ulrich, *J. Opt. Soc. Am.* 60, 1325 (1970).
172. R. Ulrich, *J. Opt. Soc. Am.* 60, 1336 (1970).
173. G. Labeyrie, A. Landragin, J. Von Zanthier, R. Kaiser, N. Vansteenkiste, C. Westbrook, and A. Aspect, *Quant. Semiclass. Opt.* 8, 603 (1996).
174. H. Wallis, *Phys. Rev. A* 56, 2060 (1997).
175. R. Cote, B. Segev, and M. G. Raizen, A 58, 3999 (1998); R. Cote and B. Segev, *Phys. Rev. A* 68, 041604R (2003).
176. D. Voigt, B. T. Wolschrijn, R. Jansen, N. Bhattacharya, R. J. C. Spreeuw, and H. B. van Linden van den Heuvell, A 61, 063412 (2000).
177. M. Kasevich, D. Weiss, and S. Chu, *Opt. Lett.* 15, 607 (1990).
178. T. Esslinger, M. Weidemuller, A. Hemmerich, and T. W. Hansch, *Opt. Lett.* 18, 450 (1993).
179. V. Savalli, D. Stevens, J. Estève, P. D. Featonby, V. Josse, N. Westbrook, C. I. Westbrook, and A. Aspect, *Phys. Rev. Lett.* 88, 250404 (2002).
180. H. Oberst, S. Kasashima, V. I. Balykin, and F. Shimizu, *Phys. Rev. A* 68, 013606 (2003).
181. J. P. Dowling and J. Gea-Banacloche, *Adv. At. Mol. Opt. Phys.* 34, 1 (1996).
182. M. W. Bromley and B. D. Esry, *Phys. Rev. A* 68, 043609 (2003).
183. J. Reichel, F. Bardou, M. B. Dahan, E. Peik, S. Rand, C. Salomon, and C. Cohen-Tannoudji, *Phys. Rev. Lett.* 75, 4575 (1995).
184. M. Wilkens, E. Goldstein, B. Taylor, and P. Meystre, *Phys. Rev. A* 47, 2366 (1993).
185. V. I. Balykin, D. V. Laryushin, M. V. Subbotin, and V. S. Letokhov, *JETP Lett.* 63, 802 (1996); V. I. Balykir, D. V. Laryushin, M. V. Subbotin, and V. S. Letokhov, *Laser Phys.* 7, 358 (1997).
186. Y. B. Ovchinnikov, D. V. Laryushin, V. I. Balykin, and V. S. Letokhov, *JETP Lett.* 62, 113 (1995).
187. Y. B. Ovchinnikov, I. Manek, and R. Grimm, *Phys. Rev. Lett.* 79, 2225 (1997).
188. A. P. Kazantsev, G. A. Ryabenko, G. I. Surdutovich, and V. P. Yakovlev, *Phys. Rep.* 129, 75 (1985).
189. J. M. Kosterlitz and D. J. Thouless, *J. Phys. C* 6, 1181 (1973).
190. B. V. Svistunov, T. W. Hijmans, G. V. Schlyapnikov, and J. T. Walraven, *Phys. Rev. B* 43, 13412 (1991).
191. H. T. Stoof and M. Bijlsma, *Phys. Rev. E* 47, 939 (1993).
192. H. Gauck, M. Hartl, D. Schneble, H. Schnitzler, T. Pfau, and J. Mlynek, 81, 5298 (1998).
193. V. I. Balykin, V. S. Letokhov, Y. B. Ovchinnikov, and A. I. Sidorov, *Phys. Rev. Lett.* 61, 902(E) (1988).
194. V. I. Balykin, K. Hakuta, F. L. Kien, J. Q. Liang, and M. Morinaga, *Phys. Rev. A* 70, 011401(R) (2004).
195. F. L. Kien, J. Q. Liang, K. Hakuta, and V. I. Balykin, *Opt. Commun.* 242, 445 (2004).
196. J. Chen, J. G. Story, and R. G. Hulet, *Phys. Rev. A* 47, 2128 (1993).
197. V. I. Balykin, V. S. Letokhov, Y. B. Ovchinnikov, A. I. Sidorov, and S. V. Shulga, *Opt. Lett.* 13, 958 (1988).
198. G. Timp, R. E. Behringer, D. M. Tennant, J. E. Cunningham, M. Prentiss, and K. K. Berggren, *Phys. Rev. Lett.* 69, 1636 (1992).
199. K. K. Berggren, M. Prentiss, G. L. Timp, R. E. Behringer, *J. Opt. Soc. Am. B* 11, 1166 (1994).
200. J. J. McClelland, R. E. Shelten, E. C. Alm, and R. J. Cellotta, *Science* 262, 877 (1993).
201. J. J. McClelland, *J. Opt. Soc. Am. B* 12, 1761 (1995).
202. J. J. McClelland, S. B. Hill, M. Pichler, and R. J. Celotta, *Sci. Tech. Adv. Mater.* 5, 575 (2004).

203. M. Drewsen, R. J. C. Spreeuw, and J. Mlynek, *Opt. Commun.* 125, 77 (1996).
204. C. J. Lee, *Phys. Rev. A* 61, 063604 (2000).
205. C. J. Lee, *Bulletin of the Korean Chemical Society* 24, 600 (2003).
206. S. J. H. Petra, K. A. H. van Leeuwen, L. Feenstra, W. Hogervorst, and W. Vassen, *European Physical Journal D* 27, 83 (2003).
207. J. E. Bjorkholm, R. E. Freeman, A. A. Askhin, and D. B. Pearson, *Opt. Lett.* 5, 111 (1980).
208. W. W. Rigrod, *Appl. Phys. Lett.* 2, 51 (1963).
209. V. I. Balykin and V. S. Letokhov, *Zh. Eksp. Teor. Fiz.* 94, 140 (1988).
210. V. V. Klimov and V. S. Letokhov, *Laser Physics* 13, 339 (2003).
211. I. Bloch, T. W. Hänsch, and T. Esslinger, *Phys. Rev. Lett.* 82, 3008 (1999).
212. I. Bloch, M. Köhl, M. Greiner, T. W. Hänsch, and T. Esslinger, *Phys. Rev. Lett.* 87, 030401 (2001).
213. V. V. Klimov and V. S. Letokhov, *JETP Lett.* 70, 654 (1999).
214. V. A. Fock, "Problems of Diffraction and Propagation of Electromagnetic Waves." Soviet Radio, Moscow, 1970, [in Russian].
215. F. D. Tappert, in "Wave Propagation and Underwater Acoustics." Springer, Berlin, 1977.
216. M. Szilagyi, "Electron and Ion Optics." Plenum, New York, 1988.
217. V. I. Balykin, V. V. Klimov, and V. S. Letokhov, *JETP Lett.* 78, 8 (2003).
218. B. Dubetsky and P. R. Berman, *Phys. Rev. A* 58, 2413 (1998).
219. J. Arlt, T. Hitomi, and K. Dholakia, *Appl. Phys. B* 71, 549 (2000).
220. T. Sleator, T. Pfau, V. Balykin, and J. Mlynek, *Appl. Phys. B* 54, 375 (1992).
221. V. Natarajan, R. E. Behringer, and G. Timp, *Phys. Rev. A* 53, 4381 (1996).
222. R. E. Behringer, V. Natarajan, G. Timp, and D. M. Tennant, *J. Vac. Sci. Technol. B* 14, 4072 (1996).
223. R. Gupta, J. J. McClelland, Z. J. Jabbour, and R. Celotta, *Appl. Phys. Lett.* 67, 1378 (1995).
224. U. Drodofsky, J. Stuhler, T. Schulze, M. Drewsen, B. Brezger, T. Pfau, and J. Mlynek, *Appl. Phys. B* 65, 755 (1997).
225. M. Mutzel, U. Rasbach, D. Meschede, C. Burstedde, J. Braun, A. Kunoth, K. Peithmann, and K. Buse, *Appl. Phys. B* 77, 1 (2003).
226. G. Birkl, F. B. J. Buchkremer, R. Dumke, and W. Ertmer, *Opt. Commun.* 191, 67 (2001).
227. S. J. Rehse, A. D. Glueck, S. A. Lee, A. B. Goulakov, C. S. Menoni, D. C. Ralph, K. S. Johnson, and M. Prentiss, *Appl. Phys. Lett.* 71, 1427 (1997).
228. S. J. Rehse, R. W. McGowan, and S. A. Lee, *Appl. Phys. B* 70, 657 (2000).
229. R. Ohmukai, S. Urabe, and M. Watanabe, *Appl. Phys. B* 77, 415 (2003).
230. P. R. Berman, B. Dubetsky, and J. L. Cohen, *Phys. Rev. A* 58, 4801 (1998).
231. H. F. Talbot, *Philos. Mag.* 9, 401 (1836).
232. J. F. Clauser and S. Li, *Phys. Rev. A* 49, R2213 (1994).
233. M. S. Chapman, C. R. Ekstrom, T. D. Hammond, J. Schmiedmayer, B. E. Tannian, S. Wehinger, and D. E. Pritchard, *Phys. Rev. A* 51, R14 (1995).
234. S. Nowak, C. Kurtsiefer, C. David, and T. Pfau, *Opt. Lett.* 22, 1430 (1997).
235. J. L. Cohen, B. Dubetsky, P. R. Berman, and J. Schmiedmayer, *Phys. Rev. A* 61, 033610 (2000).
236. J. P. Fillard, "Near Field Optics and Nanoscopy." World Scientfic, City, 1997.
237. L. Novotny, R. X. Bian, and X. Sunney, *Phys. Rev. Lett.* 79, 645 (1997).
238. V. V. Klimov and V. S. Letokhov, *Opt. Commun.* 106, 151 (1994).
239. V. V. Klimov and V. S. Letokhov, *J. Mod. Opt.* 42, 1485 (1995).
240. V. V. Klimov and V. S. Letokhov, *Laser Phys.* 6, 475 (1996).
241. V. I. Balykin, V. V. Klimov, and V. S. Letokhov, *JETP Lett.* 59, 219 (1994).
242. V. I. Balykin, V. V. Klimov, and V. S. Letokhov, *J. Phys.* 4, 1981 (1994).
243. T. Pearsey, *Philos. Mag.* 37, 311 (1946).
244. L. A. Vainshtein, "Electromagnetic Waves." Moscow Radio i Svyaz', 1988, [in Russian].
245. N. N. Lebedev, *Sov. Phys. Tech. Phys.* 2, 1943 (1957).
246. R. J. Celotta, R. Gupta, R. E. Scholten, and J. J. McClelland, *J. Appl. Phys.* 79, 6079 (1996).
247. L. Shifang and J. F. Clauser, *Phys. Rev. A* 49, 2702 (1994).
248. R. W. McGowan, D. M. Giltner, and S. A. Lee, *Opt. Lett.* 20, 2535 (1995).
249. D. Meschede and H. Metcalf, *J. Appl. Phys.* 36, R17 (2003).
250. K. K. Berggren, A. Bard, J. L. Wilbur, J. D. Gillaspy, A. G. Helg, J. J. McClelland, S. L. Rolston, W. D. Phillips, M. Prentiss, and G. M. Whitesides, *Science* 269, 1255 (1995).
251. K. S. Johnson, J. H. Thywissen, N. H. Dekner, K. K. Berggren, A. P. Chu, R. Younkin, and M. Prentiss, *Science* 280, 1583 (1998).
252. S. Nowak, T. Pfau, and J. Mlynek, *Appl. Phys. B* 63, 203 (1996).
253. W. Lu, K. G. H. Baldwin, M. D. Hoogerland, S. J. Buchman, T. J. Senden, T. E. Sheridan, and R. W. Boswell, *J. Vac. Sci. Technol. B* 16, 3846 (1998).
254. X. Ju, M. Kurahashi, T. Suzuki, and Y. Yamauchi, *J. Appl. Phys.* 42, 4767 (2003).
255. V. I. Balykin, V. S. Letokhov, V. G. Minogin, Y. B. Rozhdestvensky, and V. I. Sidorov, *J. Opt. Soc. Amer. B* 2, 1176 (1985).
256. R. Gupta, J. J. McClelland, P. Marte, and R. Celotta, *Phys. Rev. Lett.* 76, 4689 (1996).
257. B. Brezger, T. Schulze, P. O. Schmidt, R. Mertens, T. Pfau, and J. Mlynek, *Europhys. Lett.* 46, 148 (1999).
258. A. Tonomura, "Electron Holography." Springer, Berlin, 1994.
259. J. Fujita, M. Morinaga, T. Kishimoto, M. Yasuda, S. Matsui, and F. Shimizu, *Nature* 380, 691 (1996).

260. M. Morinaga, M. Yasuda, I. Kishimoto, F. Shimizu, J. Fujita, and S. Matsui, *Phys. Rev. Lett.* 77, 802 (1996).
261. J. Sone, J. Fujita, Y. Ochiai, S. Manako, S. Matsui, E. Nomura, T. Baba, H. Kawara, T. Sakamoto, C. D. Chen, Y. Nakamura, and J. S. Tsai, *Nanotechnology* 10, 135 (1999).
262. V. Natarajan, R. Behringer, and G. Timp, *Phys. Rev. A* 53, 4381 (1996).
263. W. R. Anderson, C. C. Bradley, J. J. McClelland, and R. J. Celotta, *Phys. Rev.* A 59, 2476 (1999).
264. X. D. Zhu, *Opt. Lett.* 22, 1890 (1997).

CHAPTER 2

Theoretical Investigation of Optical Properties of Single-Walled Carbon Nanotubes

Yang Zhao, XiuJun Wang, Chi-Chiu Ma, GuanHua Chen

Department of Chemistry, The University of Hong Kong, Hong Kong, People's Republic of China

CONTENTS

1. INTRODUCTION

Ever since their inception [1], carbon nanotubes (CNTs) have attracted a tremendous amount of intense experimental and theoretical interest. Synthesis of large quantities of CNTs [2] has been achieved through carbon-arc vaporization in a gas atmosphere or

ISBN: 1-58883-049-7

Handbook of Theoretical and Computational Nanotechnology
Edited By Michael Rieth and Wolfram Schommers
Volume 7: Pages (79–126)

transition-metal catalytic reaction; the latter method has been used to synthesize single-walled nanotubes (SWNTs) [3, 4]. Both open-ended and capped CNTs have been observed by high-resolution transmission electron microscope (TEM) [1, 5] and scanning tunneling microscopy [6] techniques. Because of their superb electronic and mechanical properties, many potential applications have been proposed for CNTs, such as single-electron transistors [7], tunneling-magnetoresistance devices [8], CNT diodes [9], intramolecular junctions [10], molecular bearings [11, 12], springs [13], hooks [14], and gigahertz oscillators [15–19]. Several theoretical studies revealed that CNTs would be either metallic or semiconducting, depending on their underlying structures [20–23].

The simplest among all CNTs, an SWNT, is an individual graphene sheet wrapped in the shape of a seamless cylinder, which can be characterized by its chiral vector (m, n) [22, 24]. As a simple, elongated constituent of the fullerene family, SWNTs are especially attractive to many applications thanks to their remarkable physical, chemical, and mechanical properties. In particular, their dimensions and electronic behaviors make them the ideal building blocks for molecular electronics, and recently, a self-assembled carbon nanotube field-effect transistor (FET) has been built at room temperature, using a scheme based on recognition between molecular building blocks [25]. Photoconductivity on infrared laser illumination has also been measured for individual CNTs that act as channels for an ambipolar FET [26].

It is difficult to exaggerate the importance of optical properties of SWNTs, which are vital to the development of SWNT photonic applications such as nanoscale integrated electroluminescent devices [27], to a variety of interdisciplinary applications. Optical absorption and fluorescence spectroscopy measurements provide evidence for individually dispersed carbon nanotubes, and therefore, optical properties of SWNTs have become an important tool for structure-based nanotube characterization in the exciting, emerging field of DNA-assisted manipulation of the carbon nanotubes [28]. Identification of spectroscopic features and correlations with nanotube structures also aids attempts to purify, separate, and sort SWNTs, and understanding their electronic structure helps their development as optical devices, sensors, and molecular electronics components.

Carbon nanotubes are quasi-one-dimensional crystals whose optical properties are dependent on chiralities and diameters, as well as orientations. Early measurements were often performed on multiwalled nanotubes or bundles of them and SWNTs [29, 30]. For instance, the bulk electronic properties of SWNT bundles have been studied by the high-resolution electron energy loss spectroscopy (EELS) in transmission [29]. Low-energy nondispersive features were attributed to the energy separations of DOS singularities in the nanotubes. The peak appearing in the optical conductivity at 1.8 eV was argued to originate from metallic nanotubes in the bundles. Optical absorption spectroscopy on SWNT-containing soot shows peaks between 0.6 and 3 eV that are interpreted as interband transitions between the van Hove singularities [31]. The absorption spectra of bundles of SWNTs of similar sizes have been measured, for example, on 4 Å SWNTs made by pyrolysis of tripropylamine molecules in the channels of porous zeolite $AlPO_4$-5 (AFI) single crystals [32, 33]. It is possible for only three chiralities to achieve diameters near 4 Å. When the light electric field is polarized parallel to the AFI crystal channel orientation, three low-frequency bands have been observed in the measured absorption spectra of bundles of carbon nanotubes, among which the two lowest bands were assigned to the semiconducting tubes and the third to the conducting nanotubes. When the electric field is perpendicular to the AFI crystal channel orientation, the nanotube is nearly transparent in the measured energy region 0.5–4.1 eV.

Suspension of individual SWNTs and removal of remaining bundles from solution have recently been achieved by encasing individual nanotubes in cylindrical micelles [34]. Spectrofluorimetric measurements on semiconducting SWNTs isolated in aqueous surfactant suspensions [35] delivered, for the first time, spectral information on nanotube chiralities in addition to that on diameters. Exploiting the band-gap fluorescence of these SWNTs as well as advances in solution-phase dispersion and processing of nanotubes led to a definitive (n, m) assignment to semiconducting features. Because metallic nanotubes do not fluoresce, assignment of the metallic species in a similar manner required the help of Raman spectroscopy, which was performed for laser excitations between 565 and 627 nm and also between 458 and 514.5 nm [36]. An (n, m) index has been provided to features in both optical

absorption and Raman spectra for these metallic SWNTs. In the absence of a perturbative environment of tubes and surfaces, SWNTs in aqueous micellar suspensions show much better resolved optical absorption spectra, and the one-dimensional semiconducting band gap was also found to fluoresce strongly in the 800–1600 nm range that is important to fiberoptic communications and bioimaging. Aggregation of nanotubes into bundles quenches fluorescence through interactions with metallic tubes and broadens the absorption spectra. In addition, polydispersity and poor solubility of SWNT bundles in both aqueous and nonaqueous solutions impose considerable challenges to various application-mandated separation and assembly attempts to use SWNTs as an individual macromolcular species. DNA-assisted dispersion and separation of carbon nanotubes have recently been reported [37] as an alternative to aqueous micellar suspension of SWNTs. Bundled SWNTs are effectively dispersed in water by their sonication in the presence of single-stranded DNA (ssDNA). Molecular modeling suggests that ssDNA can bind to carbon nanotubes through π-stacking, resulting in helical wrapping to the surface with a binding free energy of ssDNA to carbon nanotubes that rivals that between two nanotubes. Furthermore, wrapping of CNTs by ssDNA was found to be sequence dependent [28]. That DNA-coated SWNTs can be separated into fractions with different electronic structures by ion-exchange chromatography links one of the central molecules in biology to a technologically very important nanomaterial and opens the door to CNT-based applications in biotechnology. In this process, optical spectroscopy of SWNTs is indispensable to the characterization of nanotube geometrical, electronic, and various other properties.

Previously, it was known that ionic doping creates new metallic bands in semiconducting SWNTs. Only recently, it was shown that covalent chemistry can covert the metallic SWNTs to semiconductors, which may lead to efficient nanotube separation [38, 39]. Diazonium agents were shown to functionalize SWNTs suspended in aqueous solution with high selectivity and to display an autocatalytic effect that functionalizes the entire tube. In particular, metallic species are selected to react to the near exclusion of semiconducting SWNTs under controlled conditions [39]. Discovery of the selective functionalization of metallic nanotubes was made via the ultraviolet-visible-near-infrared absorption spectra and Raman spectroscopy. Absorption features allow for the monitoring of valence electrons in each distinct nantotube in transitions to conduction bands. A first theoretical study on the effect of functionalization on the SWNT optical spectra will be presented here. In addition, absorption spectra for double-walled carbon nanotubes (DWNTs) and CNT junctions are also discussed.

Part of the attractiveness of the CNTs has been their rich electronic properties, which may be altered via physical or chemical modifications to the CNTs. Significant progress has been made in filling the nanotubes with a range of materials. A large enhancement in conductivity is reported after doping with potassium and bromine separately [40]. Recently, potassium iodide has been successfully inserted into single-walled carbon nanotubes [41, 42]. We present here a first-principles DFT calculation of the electronic and optical properties of a potassium iodide intercalated (10, 10) nanotube.

This review is organized as follows. In Section 2, we introduce band structures of the SWNTs via the zone-folding approach from those of a graphene sheet. We discuss in detail the transition dipole moment of the SWNTs in the entire Brillouin zone when the external light field is aligned with the nanotube axis, which is followed by discussions on the absorption spectra as dictated intuitively by the transition dipole field lines and band structure contours in the framework of the simplified tight-binding model. In Section 3, optical properties of a series of finite size SWNTs including those with the smallest diameter (4 Å) are studied systematically. Their absorption spectra are calculated with the semiempirical localized-density-matrix (LDM) method based the time-dependent Hartree-Fock (TDHF) approximation. The finite optical gaps are predicted for the infinite long SWNTs. Strong anisotropy of the dynamic polarizabilities is found for 4-Å SWNTs. The compositions of the dipole-induced excitations are examined by projecting the corresponding density matrices onto the Hartree–Fock molecular orbital representation. Natures of optical excitations are investigated by examining the corresponding reduced single-electron density matrices. In Section 4, the application of the first-principles density-function theory (DFT) calculations to

SWNTs is discussed, and the focus is on the electronic and optical properties of a potassium iodide intercalated (10, 10) SWNT. In Section 5, low-energy absorption features calculated from the tight-binding model and the LDM method are compared with those from the DFT approach. Good agreements are obtained. A brief summary is given in Section 6.

2. TIGHT-BINDING MODEL

2.1. SWNTs As Quantum Confinements of Graphite

Conceptually, SWNTs are simply rolled-up graphene sheets. The diversity in the electronic structure of SWNTs arises from the quantization of the electronic wave vector of the one-dimensional systems classified by their chiralities. As shown in Fig. 1, the chiral vector $\mathbf{C}_h$ in units of hexagonal elements connecting two points on the plane defines a SWNT chirality, (n, m) [22, 24]

$$\mathbf{C}_h = n\mathbf{a}_1 + m\mathbf{a}_2 \tag{1}$$

with $\mathbf{a}_1$ and $\mathbf{a}_2$ being the real space unit vectors of the hexagonal graphite lattice. The integers n and m are such that $0 \leq |m| \leq n$. An armchair nanotube (3, 3) and a zigzag nanotube (6, 0) are shown in Fig. 1 as two examples. Because SWNTs are folded from graphene sheets, it is not surprising that their physical properties are closely related to those of the graphite, as shall be demonstrated in this review of their optical properties [22]. The chiral vector $\mathbf{C}_h$ is along the circumferential direction of the nanotube and, therefore, is orthogonal to the nanotube axis. The reciprocal lattice vectors $\mathbf{k}_2$ and $\mathbf{k}_2$ are defined as follows

$$\mathbf{k}_1 = N^{-1}(-t_2\mathbf{b}_1 + t_1\mathbf{b}_2), \qquad \mathbf{k}_2 = N^{-1}(m\mathbf{b}_1 - n\mathbf{b}_2) \tag{2}$$

where $\mathbf{b}_1$ and $\mathbf{b}_2$ are the unit vectors of the reciprocal lattice, $t_1 = d_R^{-1}(2m + n)$ and $t_2 = d_R^{-1}(2n + m)$, with d_R the greatest common divisor of $2n + m$ and $2m + n$. Spatially, $\mathbf{k}_2$ is along the tubule axis, and $\mathbf{k}_1$ is in the circumferential direction. The translational vector $\mathbf{T}$ is a unit vector for the one-dimensional carbon nanotube. It is parallel to the nanotube axis and can be expressed as

$$\mathbf{T} = t_1\mathbf{a}_1 + t_2\mathbf{a}_2 \tag{3}$$

In the tight-binding model in which only π electrons are considered, the electronic structure of a carbon nanotube can be derived directly from that of a graphene sheet via the so-called zone folding of the two-dimensional energy dispersion relation of the graphite. Intuitively, the band spectra of a carbon nanotube are the intersects of the graphene sheet's conduction and valence energy surfaces and a set of parallel planes perpendicular to the Brillouin zone. The hexagon in Fig. 2a is the first Brillouin zone of the graphene sheet, and the parallel dashed lines, which are along the tubule axis $\mathbf{T}$, are the intersects between the

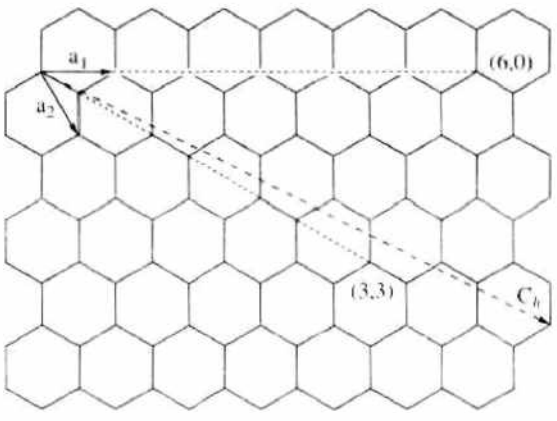


Figure 1. Schematic plot of a two-dimensional graphene sheet showing the lattice vectors $\mathbf{a}_1$ and $\mathbf{a}_2$, and the chiral vector $\mathbf{C}_h = n\mathbf{a}_1 + m\mathbf{a}_2$. An armchair nanotube (3, 3) and a zigzag nanotube (6, 0) are indicated as two examples.

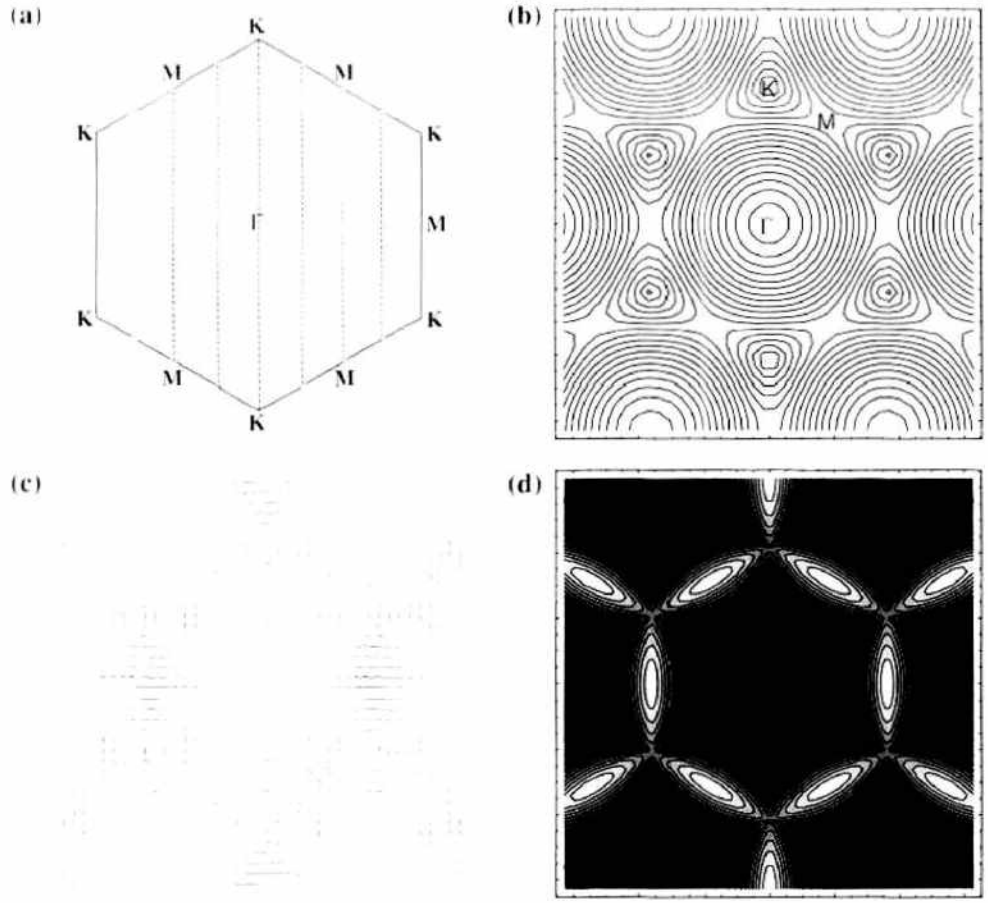


Figure 2. (a) The first Brillouin zone of graphene. The parallel lines (along the tubule axis **T**) are the intersects between the Brillouin zone and the parallel planes that are perpendicular to the Brillouin zone. They represent the allowed states for a (5, 5) SWNT. (b) The contour plot of the conduction band of the graphene sheet. The circle at the center of the first Brillouin zone is the Γ point, and the six surrounding circles are the *K* points at which the conduction and valence bands join. (c) The transition-dipole field lines of graphene in the first Brillouin zone. The orientations of the lines represent the directions of the transition dipoles at that particular **k**, and the lengths of the lines represent the sizes of the transition dipoles. The transition dipole vanishes asymptotically on approaching the Brillouin zone center. The size of transition dipole reaches its maxima on lines connecting neighboring *K* points. A circular pattern is found around *K* points. (d) The contour plot of the oscillator strength $|\mathbf{d}|^2$. The $|\mathbf{d}|^2$ maxima are located at *M* points (light area), and its minima at the Γ point (dark area). Reprinted with permission from [43], Y. Zhao et al., *Chem. Phys. Lett.* 387, 149 (2004). © 2004, Elsevier B. V.

Brillouin zone and the set of parallel planes. The parallel lines shown in Fig. 2a are the allowed states for a (5, 5) SWNT. At *K* points, the conduction (E_c) and valence (E_v) bands join, and the energy difference $E_c - E_v$ vanishes; at Γ point $E_c - E_v$ reaches its maxima, and *M* points are the saddle points at which the energy difference is rather flat nearby. If one of the intersects or parallel lines such as those in Fig. 2a goes through a *K* point in the Brillouin zone, the tube is metallic ($n - m = 3k$). Otherwise, it is a semiconductor ($n - m = 3k \pm 1$). The absorption spectra of SWNTs can be similarly obtained from the optical properties of a graphene sheet via a set of graphic tools, which will be demonstrated in this review.

Neglecting the overlap integral between the adjacent carbon atoms, the conduction band for a graphene sheet is commonly approximated by

$$E_c(\mathbf{k}) = V_{pp\pi}\sqrt{3 + 4\cos\frac{\sqrt{3}k_x}{2}\cos\frac{k_y}{2} + 2\cos k_y} \tag{4}$$

where $V_{pp\pi}$ is the transfer integral between adjacent lattice points, also called the nearest-neighbor $pp\pi$ interaction. A plot of $E_c(\mathbf{k})$ is given in Fig. 2a. The Hamiltonian field of $E_c(\mathbf{k})$ is shown in Fig. 2b. The conduction band describes the π^*-energy antibonding band and the covalence band, which is the mirror image of the conduction with respect to the

Fermi energy, the π-energy bonding band. The two bands are degenerate at K points the Fermi energy crosses. The conduction bands of the SWNTs folded from the planar graphite sheet can be obtained from

$$E_{c,k_2}^{\text{tube}} = E_c(k\mathbf{k}_2|\mathbf{k}_2|^{-1} + \mu\mathbf{k}_1) \tag{5}$$

where $\mu = 1, \ldots, N$, and $-\pi/|\mathbf{T}| < k < \pi/|\mathbf{T}|$. The vertical parallel lines in Fig. 2a, or the intersects, are along the $\mathbf{k}_2$ directions. Therefore, the intersects are also called $\mathbf{k}_2$ cuts.

Kataura et al. plotted energy differences between the ith van Hove singularities in the conduction and valence bands (both numbered from the Fermi energy) $E_{ii}(d_t)$ as a function of the nanotube diameter d_t for all chiralities, and showed that the energy differences $E_{ii}(d_t)$ have a finite width for a given d_t [31]. This width in $E_{ii}(d_t)$ increases with the increasing energy deviation of $E_{ii}(d_t)$ from the Fermi energy. It was later attributed to the trigonal warping effect of the energy bands [44, 45], which is referred to the deviation of the equal-energy contours near the K points in the Brillouin zone away from K-centered circular patterns and toward a triangular shape, with the three neighboring M points as its corners (cf. Fig. 2b).

2.2. Transition Dipole Moment

In 1998 White and Mintmire [46] solved a puzzle posed by Dresselhaus [21] on the density of states (DOS) being independent of translational unit cell sizes and chiral angles of semiconducting carbon nanotubes [47]. White and Mintmire's discovery has since been employed to interpret various optical measurements. Two lowest peaks in the absorption spectra of SWNT bundles were assigned to the semiconducting SWNTs, and the third was attributed to the metallic tubes. The energies of the three peaks are approximately $\frac{2a}{d_t}|V_{pp\pi}|$, $\frac{4a}{d_t}|V_{pp\pi}|$, and $\frac{6a}{d_t}|V_{pp\pi}|$, where a is the C–C bond length. However, such interpretations of absorption spectra of SWNTs in bundles or aqueous micellar suspensions, although likely correct, have not been adequately scrutinized. In addition to the DOS, the absorption lineshape also has a nontrivial dependence on the transition dipole moments between pairs of valence and conduction bands. A finite transition dipole projection along the direction of an external field ensures the appearance of an absorption peak, which links two van Hove singularities in the DOS, whereas a zero or negligible transition dipole moment gives no absorption peaks even if the corresponding DOS diverges. Calculated transition dipole moments in the entire Brillouin zone, however, have remained elusive in the literature. Here we start with the tight-binding model and determine the transition dipole field lines of a graphene sheet. The general absorption spectral features of SWNTs shall become obvious from those field lines and plots of the conduction (valence) band contours shown in the previous section.

Aside from the van Hove singularities of optically connected valence/conduction bands, the transition dipole moment is the single most important quantity for the prediction of the optical absorption lineshapes. Depending on the orientation of the external laser field, the transition dipole moment of the SWNTs can assume quite different characteristics thanks to the selection rules. Optical responses of carbon nanotubes are highly anisotropic [48, 49], and the anisotropy of carbon nanotubes has been investigated experimentally by Walt de Heer and his colleagues [50]. The dielectric function ϵ was found being much larger when the electric field is aligned along the tube axis than when it is aligned perpendicular to the tube axis. This can be attributed to the fact that transition dipole along the tubule axis $\mathbf{T}$ is larger than its component perpendicular to the tubes. We consider here first the scenario in which the external field is parallel to the tubule axis ($\mathbf{E} \parallel \mathbf{T}$), and thus the transition occurs exclusively between the orbitals of the same momenta $\mathbf{k}$ [51, 52]. The scenario in which external field is perpendicular to the tubule axis ($\mathbf{E} \perp \mathbf{T}$) involves transitions between orbitals with momenta differing by $\pm\mathbf{k}_1$. This case will be briefly visited at the closing of the section.

If the external field points along the tubule axis, the transition occurs only vertically between states of the same momenta $\mathbf{k}$. Following, we derive and plot the size and orientation of the vertical transition dipole moment in the entire Brillouin zone. Implications about the absorption spectra will be then discussed. For $\mathbf{E} \parallel \mathbf{T}$, the tubule-axis projection of the

transition dipole can be calculated from

$$\mathbf{d} \cdot \mathbf{E} = \langle \phi_c(\mathbf{k}) | e\mathbf{E} \cdot \mathbf{r} | \phi_v(\mathbf{k}) \rangle \tag{6}$$

where the transition happens between the valence-band wave function $|\phi_v(\mathbf{k})\rangle$ and the conduction-band wave function $|\phi_c(\mathbf{k})\rangle$. Therefore, it is of interest to us to visualize the vector field of the dipole

$$\mathbf{d} = \langle \phi_c(\mathbf{k}) | e\mathbf{r} | \phi_v(\mathbf{k}) \rangle \tag{7}$$

A brief derivation of the transition dipole $\mathbf{d}$ is as follows. The tight-binding wave function for the conduction (valence) band $|\phi_c(\mathbf{k})\rangle$ ($|\phi_v(\mathbf{k})\rangle$) can be written as

$$|\phi_c(\mathbf{k})\rangle = -c|A\rangle + |B\rangle \tag{8}$$

$$|\phi_v(\mathbf{k})\rangle = c|A\rangle + |B\rangle \tag{9}$$

where $|A\rangle$ ($|B\rangle$) is the Bloch function for the A (B) sublattice so that

$$\langle \mathbf{r}|A\rangle = \sum_{\mathbf{R}_A} e^{i\mathbf{k}\cdot\mathbf{R}_A} \psi_A(\mathbf{r} - \mathbf{R}_A) \tag{10}$$

$$\langle \mathbf{r}|B\rangle = \sum_{\mathbf{R}_B} e^{i\mathbf{k}\cdot\mathbf{R}_B} \psi_B(\mathbf{r} - \mathbf{R}_B) \tag{11}$$

and the coefficient c has the form

$$c = \frac{e^{ik_x/4\sqrt{3}} \sqrt{e^{i\sqrt{3}k_x/2} + 2\cos\left(\frac{k_y}{2}\right) + 2e^{i\sqrt{3}k_x/4}\cos\left(\frac{k_y}{2}\right) + 4e^{i\sqrt{3}k_x/2}\cos^2\left(\frac{k_y}{2}\right)}}{1 + 2e^{i\sqrt{3}k_x/2}\cos\left(\frac{k_y}{2}\right)} \tag{12}$$

The fact that $|c|^2 = 1$ ensures the orthogonality of the two wave functions, $|\phi_c(\mathbf{k})\rangle$ and $|\phi_v(\mathbf{k})\rangle$. The transition dipole between the two wave fucntions, $\langle \phi_c | e\mathbf{r} | \phi_v \rangle$, is composed of four terms:

$$\langle \phi_c | e\mathbf{r} | \phi_v \rangle = -\langle A | e\mathbf{r} | A \rangle + \langle B | e\mathbf{r} | B \rangle - c^* \langle A | e\mathbf{r} | B \rangle + c \langle B | e\mathbf{r} | A \rangle \tag{13}$$

Among the four terms on the right-hand side of Eq. (13), only the cross terms survive, and after substituting in the Bloch functions, one obtains

$$\langle \phi_c | e\mathbf{r} | \phi_v \rangle = -\frac{c^*}{iV_{pp\pi}} \vec{\nabla}_{\mathbf{k}} H_{12} \langle a | r | b \rangle + \frac{c}{iV_{pp\pi}} \vec{\nabla}_{\mathbf{k}} H_{21} \langle b | r | a \rangle \tag{14}$$

where $r = |\mathbf{r}|$, the off-diagonal Hamiltonian matrix element

$$H_{12} = V_{pp\pi} \sum_{i=1}^{3} e^{i\mathbf{k}\cdot\mathbf{R}_i} \tag{15}$$

with $\mathbf{R}_i$ ($i = 1, 2, 3$) connecting three nearest-neighbor B-lattice points to an A-lattice point, and $|a\rangle$ and $|b\rangle$ being two neighboring wave functions of the A and B sublattices, respectively

$$\langle \mathbf{r}|a\rangle = \psi_A(\mathbf{r} - \mathbf{R}_A) \tag{16}$$

$$\langle \mathbf{r}|b\rangle = \psi_B(\mathbf{r} - \mathbf{R}_A - \mathbf{R}_i), \qquad i = 1, 2, 3 \tag{17}$$

To arrive at Eq. (14), one has taken advantage of the symmetry of the π orbitals

$$\langle b | \mathbf{r} | a \rangle = |\mathbf{R}_{ab}|^{-1} \mathbf{R}_{ab} \langle b | r | a \rangle \tag{18}$$

where the A sublattice point a is taken as the origin without loss of generality, and $\mathbf{R}_{ab}$ labels the vector pointing from point a to point b.

The sizes and orientations of **d** as a function of the wave vector **k** are shown in Fig. 2c. The transition dipoles form circular patterns around K points and are prominent along lines connecting neighboring K points. The oscillator strength, therefore, is mainly concentrated in the vicinities of the Brillouin zone boundaries; that is, in areas between neighboring K points. One main feature of the **d** orientations is that the transition dipole points tangentially around the K points, forming pseudo-vortices centered at the K points (cf. Fig. 2c). This is significant because the transition dipole consequently has sizable projections along the tubule axis direction around the K points regardless of any nanotube chirality.

Close to the Brillouin zone centers, the size of the transition dipole vanishes asymptotically. This can be understood as follows: At the Brillouin zone center ($\mathbf{k} = 0$), the transition dipole size has to vanish because of symmetry considerations (i.e., to avoid choosing an orientation at the Γ point). Around the zone center, the transition dipole size has the following property

$$|\mathbf{d}|(k_x = 0, k_y \approx 0) = \frac{k_y^2}{9\sqrt{3}} + O|k_y|^4, \qquad |\mathbf{d}|(k_x \approx 0, k_y = 0) = \frac{k_x^2}{9\sqrt{3}} + O|k_x|^4 \tag{19}$$

Therefore, the oscillator strength roughly scales with $|\mathbf{k}|^4$ around the zone center. Along M–Γ–M lines, the transition dipoles are found parallel to the M–Γ–M lines, whereas along K–Γ–K lines, transition dipoles are perpendicular to the K–Γ–K lines.

In the literature, the optical absorption spectra are often calculated from a matrix element containing the momentum operator **P** [52–54]

$$\langle \phi_c(\mathbf{k}) | \mathbf{P} \cdot \mathbf{E} | \phi_v(\mathbf{k}) \rangle \tag{20}$$

This matrix element is proportional to the transition dipole moment we have calculated, as shown by

$$i\hbar\dot{\mathbf{r}} = [\mathbf{r}, \mathbf{H}] \tag{21}$$

where **H** is the nanotube Hamiltonian.

If the external electric field points perpendicular to the tubule axis, the selection rule dictates that the transition dipole take the form [52, 53, 55]

$$\mathbf{d}' \cdot \mathbf{E} = \langle \phi_c(\mathbf{k}) | e\mathbf{E} \cdot \mathbf{r} | \phi_v(\mathbf{k} \pm \mathbf{k}_1) \rangle \tag{22}$$

Similarly, the vector field of the transverse transition dipole $\mathbf{d}'$

$$\mathbf{d}' = \langle \phi_c(\mathbf{k}) | e\mathbf{r} | \phi_v(\mathbf{k} \pm \mathbf{k}_1) \rangle \tag{23}$$

now assumes the importance of the vertical transition dipole **d** field in the previous discussions.

2.3. The Absorption Intensities

The quasi-one-dimensionality of SWNTs gives rise to sharp van Hove peaks in the density of electronic states. Optical properties of SWNTs are thus dominated by transitions between corresponding van Hove singularities on opposite sides of the Fermi level. Contributions to the absorption lineshape from excitations polarized along the tubule axis at the momentum **k** are determined by two factors; namely, the transition dipole **d** projected along the tubule axis $\mathbf{T} \cdot \mathbf{d}$, and the nanotube DOS for the corresponding conduction (valence) band at **k**, which is precisely $(\mathbf{T} \cdot \nabla_{\mathbf{k}} E_c)^{-1}$ $[(\mathbf{T} \cdot \nabla_{\mathbf{k}} E_v)^{-1}]$. In other words, the absorption intensity is proportional to $(\mathbf{T} \cdot \mathbf{d})^2 (\mathbf{T} \cdot \nabla_{\mathbf{k}} E_c)^{-1} (\mathbf{T} \cdot \nabla_{\mathbf{k}} E_v)^{-1}$. Therefore, except where the $\mathbf{k}_2$ cuts are perpendicular to the transition dipole **d** field lines, or where the transition dipole **d** vanishes, the shape of the absorption spectra is determined mostly by the density of states of the nanotube bands [53]. For external excitation fields along the tubule axis, as is the case under discussion here, vertical transitions in the **k** space are expected, and one needs to find out how the DOS contours lines of the graphite conduction (or covalence) band relate to the tubule axis. The contour lines of the conduction band E_c is depicted in Fig. 2b. When the

tubule axis is tangential to a contour line of E_c at a given $\mathbf{k}$, $(\mathbf{T}\cdot\nabla_{\mathbf{k}}E_c)^{-1}$ diverges, and a van Hove singularity appears in the DOS. Because E_c and E_v are mirror images of each other under the approximation that neglects the overlap integral between adjacent carbon atoms, $(\mathbf{T}\cdot\nabla_{\mathbf{k}}E_v)^{-1}$ also diverges at that particular $\mathbf{k}$. If the transition dipole between the valence and conduction orbitals has a nonzero projection along $\mathbf{T}$ at that $\mathbf{k}$, a corresponding peak arises in the absorption spectrum. Of particular interest are those K and M points in the Brillouin zone. An intersect through the K points does not cause a van Hove singularity, but adjacent intersects do, as noted by White and Mintmire. For metallic SWNTs, two intersects with a spacing $\frac{2}{d_t}$ from the K point have van Hove singularities. Because the energy contours are nearly circular around the K point (cf. Fig. 2b), $E_c - E_v$ at both van Hove singularities can be approximated by $\frac{6a}{d_t}|V_{pp\pi}|$. For semiconducting tubes, no intersect goes through K points. The closest intersect to a K point is $\frac{2}{3d_t}$ from the K point, and the next closest is $\frac{4}{3d_t}$ from the K point; $E_c - E_v$ at the two van Hove singularities are $\frac{2a}{d_t}|V_{pp\pi}|$ and $\frac{4a}{d_t}|V_{pp\pi}|$, respectively. Similarly, any intersects at the M points or their vicinities may lead to van Hove singularities. Because the M points are saddle points at which $E_c - E_v$ are relatively flat, transition energies linking these van Hove singularities are approximately $2|V_{pp\pi}|$.

Neither the oscillator strength nor the DOS alone determines the optical absorption lineshapes of the SWNTs. One is therefore led to study the combined effect of the two competing factors. Figure 3 plots the product of the $E_c(\mathbf{k})$ DOS and the oscillator strength, which combines the effects of the transition dipole and DOS on the absorption spectra

$$|\mathbf{d}|^2|\vec{\nabla}_{\mathbf{k}}E_c(\mathbf{k})|^2 = |\langle\phi_c(\mathbf{k})|e\mathbf{r}|\phi_v(\mathbf{k})\rangle|^2|\vec{\nabla}_{\mathbf{k}}E_c(\mathbf{k})|^2 \tag{24}$$

The vanishing oscillator strength annihilates the DOS singularity at the Brillouin zone center. The contour plot that corresponds to the accompanying three-dimensional plot of $|\mathbf{d}|^2|\vec{\nabla}_{\mathbf{k}}E_c(\mathbf{k})|^2$ in Fig. 3 therefore resembles Fig. 2d.

For (5, 5) nanotubes, for example, the external laser field along the nanotube axis is parallel to the energy intersects or the parallel lines in Fig. 2a. Except when an intersect (see Fig. 2a) is orthogonal to the $\mathbf{d}$ field line at momentum $\mathbf{k}$, the transition dipole will have a nonzero component along the external field. Because the $\mathbf{d}$ field lines near the K points are, to the lowest order, circular in pattern, any intersects in its vicinity are guaranteed not to be perpendicular to the $\mathbf{d}$ field lines. It follows that any transitions linking the van Hove singularities near the K points are allowed. Therefore, armed with plots of the conduction (valence) band contours and the transition dipole field lines of the graphene (i.e., Fig. 2b and 2c), we have proven that the lowest-energy peaks in both metallic and semiconducting tubes depend only on the diameter; for tubes of approximately the same diameters, the two lowest-energy peaks from the semiconducting tubes are about 1/3 and 2/3 in energy, respectively, when compared with the lowest-energy peaks from the metallic tubes. The $\mathbf{d}$ field lines near an M point are approximately parallel to each other. If an intersect is perpendicular to the $\mathbf{d}$ field lines near one M point, as a result of the existence of six equivalent M points in

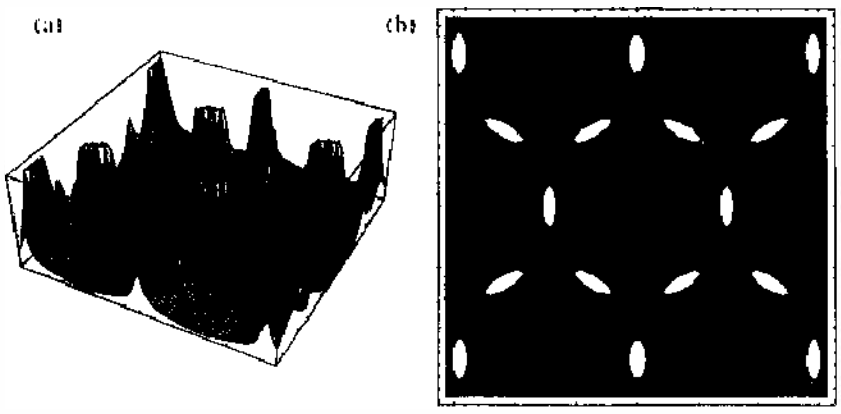


Figure 3. Combined effect of the transition dipole and the SWNT DOS: (a) three-dimensional plot of $|\mathbf{d}|^2|\vec{\nabla}_{\mathbf{k}}E_c(\mathbf{k})|^2$ in the Brillouin zone, and (b) its corresponding contour.

the first Brillouin zone, the same set of intersects parallel to the tubule axis will not be perpendicular to the field lines near the other five M points. We thus expect to observe the absorption peaks attributed to transitions at the M points or in their vicinities, with transition energies at about $2\,|V_{pp\pi}|$. This is helped by the fact that the graphene's oscillator strength, which is proportional to $|\mathbf{d}|^2$, is mainly concentrated near the Brillouin zone boundaries, and especially around the M points, as mentioned earlier from Fig. 2d. Close to the Brillouin zone center, contributions to the absorption lineshapes are weak at best, as the size of the transition dipole vanishes asymptotically at $\mathbf{k} = 0$.

Similar to the orientation plot of transition dipoles in Fig. 2c, the Hamiltonian field of the conduction band also forms vortices around the K points. The area near the M and K points is amplified in Fig. 4a. The Hamiltonian fields along the K–M–K line points almost perpendicular to the K–M–K lines. Therefore, the areas close to the M–K lines are potential candidates for generating absorption peaks as the intersects, or the $\mathbf{k}_2$ cuts, are quite likely tangential to the contour lines along M–K lines. This shall become clearer later. Another stunning feature of the conduction-band Hamiltonian field is that along the M–M lines, the Hamiltonian field points straight toward, or away from, the M points. As a consequence for zigzag $(n, 0)$ tubes, for which all $\mathbf{k}_2$ cuts are parallel to the M–M lines, the $\mathbf{k}_2$ cuts that are close to the M–M lines will contain a finite momentum bracket within which the contour lines are parallel to the tubule axis, and the SWNT DOS diverges. This has implications for the absorption lineshapes calculated by the tight-binding model. For zizag SWNTs, the tight-binding spectra will therefore have large intensities at the energies corresponding to the vicinities of the M–M lines. The spectral contributions from various $\mathbf{k}_2$ cuts for the (9,0) are tabled in Fig. 4b, labeled by μ. The $\mu = 4$ and $\mu = 5$ cuts are close to the M–M lines and are responsible for high-intensity contributions to the absorption spectra over a wide energy range.

For armchair SWNTs, the allowed states, or the $\mathbf{k}_2$ energy cuts, are analogous to the vertical lines in Fig. 2a, which are specifically the $\mathbf{k}_2$ cuts for the $(5, 5)$ tube. To obtain a precise line along which DOS of any armchair SWNT diverges in the vicinities of the K–M–K line, one needs only to solve

$$\mathbf{e}_x \cdot \vec{\nabla}_{\mathbf{k}} E_c(\mathbf{k}) = 0 \tag{25}$$

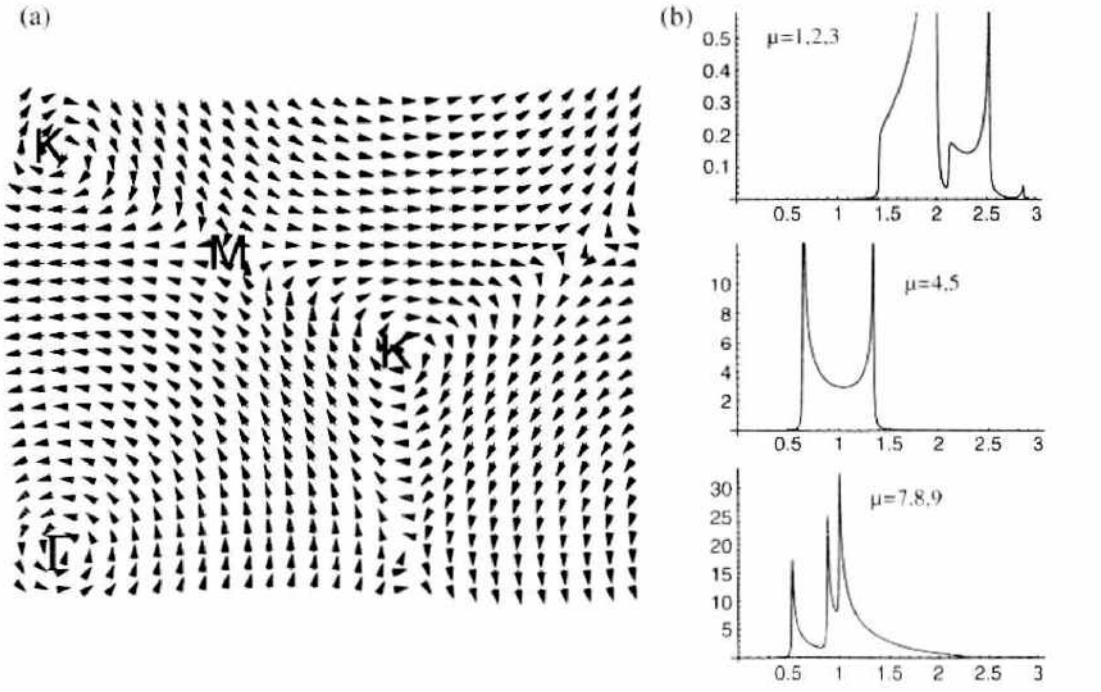


Figure 4. (a) The area near the M and K points is amplified for the Hamiltonian field of the SWNT conduction band; (b) the spectral contributions from various $\mathbf{k}_2$ cuts for the (9, 0) are tabled. Cuts are labelled by μ. The $\mu = 6$ cut goes through the K point, and therefore, it is not listed.

which leads to

$$\cos\frac{\sqrt{3}k_x}{2} + 2\cos\frac{k_y}{2} = 0 \tag{26}$$

The trajectory determined by Eq. (26), as shown by the curvy line in Fig. 5, is very close to the straight line connecting M and K points. For zigzag SWNTs, in contrast, the $\mathbf{k}_2$ cuts are perpendicular to the K–M–K line, which implies that the SWNT DOS diverges on the K–M–K line, and all tight-binding absorption peaks for zigzag SWNTs can be attributed to the points on the K–M–K line. For nanotubes of any other chiralities, the corresponding tight-binding low energy absorption peaks can be attributed to the tiny area in the Brillouin zone that is bordered by the M–K line and the curvy line given by Eq. (26), as shown by the shaded area in Fig. 5.

Recently, Grüneis et al. pointed out an intensity node in optical absorption spectra of a graphene sheet as a function of the electron wave vector $\mathbf{k}$ and the light polarization around the K points in the two-dimensional Brillouin zone [53]. This is in full agreement with the transition-dipole plot, Fig. 2c. For example, in the vicinities of the K points, for a vertical light polarization in Fig. 2c, that is, $\mathbf{E} \parallel \mathbf{k}_y$, there will be zero absorption intensity at wave vectors $\mathbf{k}$, which are on the vertical lines going through the K points. Similarly, for $\mathbf{E} \parallel \mathbf{k}_x$ in Fig. 2c, zero absorption intensity at wave vectors $\mathbf{k}$, which are on the horizontal lines going through the K points. The two nodes on the two sides of the K points, however, are not equivalent, which is also clearly demonstrated in Fig. 2c (cf. Fig. 2 of Ref. [53]).

Optical spectroscopy has been used to reveal detailed composition of bulk SWNT samples providing distributions in both tube diameter and chiral angle. For semiconducting SWNTs, those measurements used small deviations of absorption peaks' dependence on diameter from a linear relation, which can be accounted for qualitatively by the tight-binding model here. For instance, Bachilo et al. observed that the energy ratios between the second and first van Hove singularities, often referred to as E_{22}/E_{11} (E_{22} and E_{11} label the second and first van Hove singularities, respectively), deviate from a central value in the opposite directions for semiconducting nanotubes with $(n - m) \bmod 3 = 1$ and those with $(n - m) \bmod 3 = 2$ [35]. This can be explained by the fact that for wave vectors sufficiently apart from the K points, the energy slope along the K–Γ lines is much steeper than that along the K–M lines (cf. Fig. 2b and Fig. 6).

More features in Fig. 2 of Ref. [35] can be explained intuitively by the graphical tools developed here. In Fig. 6, allowed states for a zigzag semiconducting SWNT can be represented by solid horizontal lines that are perpendicular to the dotted vertical K–Γ line. As the energy slope difference on the two sides of the K point is the largest along the vertical K–Γ line among all lines going through the K point (cf. Fig. 6), zigzag SWNTs exhibit highest deviations of the energy ratios E_{22}/E_{11} from the value two. Furthermore, for zigzag SWNTs, the energy ratio E_{22}/E_{11} for the case of $(n - m) \bmod 3 = 2$ [$(n - m) \bmod 3 = 1$] is below (above) the value two, as clearly demonstrated in Fig. 6a (Fig. 6b). On the other hand, for metallic armchair nanotubes, for which the trigonal warping effect is absent due to

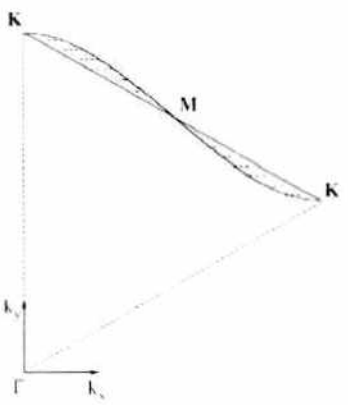


Figure 5. The vicinities of the M–K–M line in the Brillouin zone.

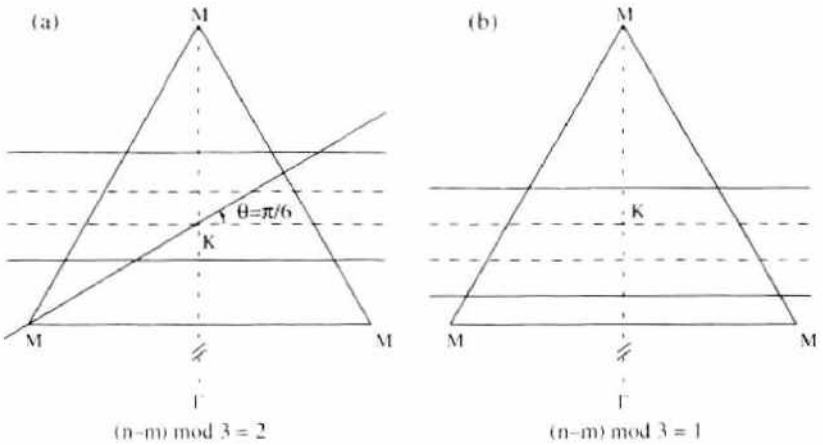


Figure 6. Horizontal solid parallel lines are the allowed states in the Brillouin zone for (a) a zigzag nanotube with $(n - m) \bmod 3 = 2$, and (b) a zigzag nanotube with $(n - m) \bmod 3 = 1$. Dashed parallel lines divide the distance between the solid lines into three equal parts. One of the two dashed lines goes through the K point. The tilted solid line in (a) connecting the M and K points represents a part of allowed states for armchair nanotubes. Dotted lines connect the K and Γ points. Reprinted with permission from [43], Y. Zhao et al., *Chem. Phys. Lett.* 387, 149 (2004). © 2004, Elsevier B. V.

symmetry, the allowed states are represented by lines that are parallel to the tilted solid line connecting M and K points in Fig. 6a (equivalent to the dashed parallel lines in Fig. 2a). For non-armchair metallic SWNTs, the trigonal warping effect brings about a splitting in the first absorption peak in the tight-binding model, as was elaborated by Saito et al. [44] Phonon spectra have recently been used to verify the existence of singularity splitting in the joint density of electronic states in non-armchair metallic SWNTs [56, 57].

Among semiconducting nanotubes, the one with $n - m = 1$ has allowed states that are on lines with the smallest deviation angle from the tilted M–K line in Fig. 6a, and consequently E_{22}/E_{11} for $n - m = 1$ are closest to two. The set of parallel lines representing allowed nanotube states is not unique. Because of the sixfold symmetry of the Brillouin zone, the angle θ between the solid horizontal lines in Fig. 6a and parallel lines representing allowed states for *any* SWNT can be narrowly confined to be within $\pi/6$. This implies that the energies of allowed states for all SWNTs behave in a similar way as those of the zigzag nanotubes. Differing energy slopes of allowed states on the two sides of the K point cause E_{22}/E_{11} to deviate from two, the direction of which depends on whether $(n - m) \bmod 3$ equals 1 or 2. As the angle θ approaches $\pi/6$ (for $n - m = 1$, for example), $|E_{22}/E_{11} - 2|$ minimizes. This also explains why deviation of E_{22}/E_{11} from the value two increases with $n - m$ (and with $|\theta - \pi/6|$) as was observed in Ref. [35].

A rigorous proof has been given here to these interpretations of the experimental results [29, 31, 33, 58] on the low-energy peaks in the absorption spectra of SWNTs. In contrast to those low-frequency features, common absorption peaks attributed to the M points and their vicinities remain to be observed experimentally for SWNTs of various sizes and chiralities. Orbital hybridization and electronic correlations are also more likely to affect higher energy excitations, which makes their experimental confirmations less certain.

3. LOCALIZED-DENSITY-MATRIX METHOD

The prediction of White and Mintmire [46] about the low-lying absorption peaks of SWNTs, and that from the tight-binding model as discussed in length in the previous section, consider only explicitly the π electrons. Rehybridization of σ and π orbitals and electronic correlations are known to affect the band structure, DOS, and transition dipoles of SWNTs and may thus lead to substantial changes of optical absorption spectra. Therefore, more realistic calculations beyond the tight-binding model are needed to examine the implications of White and Mintmire's discovery for optical absorption. In this section we calculate SWNT optical absorption spectra by including explicitly the electron–electron correlations and the σ–π orbital rehybridization in the framework of the LDM method based on the TDHF approximation [48].

In the LDM approach, electronic correlations are taken into account within the framework of TDHF approximation or random phase approximation (RPA) [59]. We have the following equation for the linear response of the density matrix deviation $\delta\rho$ from its ground-state value $\rho^{(0)}$

$$\left(i\hbar\frac{d}{dt}+\gamma\right)\delta\rho(t)=\left[h^{(0)},\delta\rho(t)\right]+\left[\delta h(t),\rho^{(0)}\right]-\mathbf{E}(t)\cdot\left[\mathbf{P},\rho^{(0)}\right] \tag{27}$$

The following approximations are employed to achieve the linear-scaling calculation for the excited state properties [60, 61, 49]:

$$\rho^{(0)}{}_{ab}^{mn}=0 \quad \text{if} \quad r_{ab}>l_0$$
$$\delta\rho_{ab}^{mn}=0 \quad \text{if} \quad r_{ab}>l_1$$

where r_{ab} is the distance between two atoms a and b, and l_0 and l_1 are two cut-off lengths. The DOS operator can be defined as

$$\hat{\sigma}=\sum_m |m\rangle\langle m|\delta(E-E_m) \tag{28}$$

where $|m\rangle$ is the state m, and E_m is its energy. The DOS of a system is then given by [62]

$$\sum_m \delta(E-E_m)=\mathrm{Tr}\hat{\sigma}(E)=\frac{1}{\pi}\mathrm{Im}\sum_m \lim_{\eta\to 0}\frac{1}{E-E_m-i\eta} \tag{29}$$

If the $|m\rangle$ is a Hartree–Fock molecular orbital, Eq. (29) gives the one-electron DOS.

The LDM method has been developed to evaluate the ground- and excited-state properties of very large systems [48, 49, 60, 61, 63–67], which is facilitated by the truncation of the reduced single-electron density matrices. As a result, the computation time of the LDM method can be scaled linearly with the system size. The LDM method has been used to determine the absorption spectra of a few open-ended zigzag SWNTs [49]. The PPP Hamiltonian, which considers only the π electrons, is adopted first in the computation. The fast multipole method (FMM) method is employed to test linear-scaling properties of the calculations [49]. SWNTs are often regarded as one-dimensional nanostructures because their lengths are far greater than their diameters. To illustrate that the LDM method is applicable to three-dimensional systems, we consider the nanotubes with diameters comparable to their lengths. The diameters range from 20.37 to 81.48 Å, and the lengths from 15.63 to 66.78 Å. The number of carbon atoms corresponds to 416, 1664, 3328, 4576, and 6656. The CPU time for propagating the TDHF equation of motion for the single-electron reduced density matrix between a time interval of [−0.5 fs, −0.3 fs] is recorded. The time step is 0.01 fs. The critical lengths l_0 and l_1 are set to be 15 Å. The results are shown in Fig. 7a. The CPU time scales linearly with the system size N. The maximum number of atoms in the smallest box is kept at 26. The corresponding absorption spectrum of a SWNT with 1200 carbon atoms is shown in Fig. 7b. The critical cutoff lengths l_0=28 Å and l_1=43 Å are used. The calculated absorption spectrum, which differs significantly from those of one-dimensional SWNTs, resembles those of graphite [68] if the nanotube radius is large enough.

The curvature of the tube leads to some sp^3 hybridization of π and σ orbitals. Orbital hybridization, which increases with decreasing tube diameters, may alter significantly the zero-order band structure of the SWNTs. Thus, a more realistic model that includes all valence orbitals is needed to describe accurately the electronic structures of the SWNTs, especially those of the capped ones. Semiempirical methods such as CNDO/S [69], INDO [70], MNDO [71], and AM1 [72] consider explicitly all valence electrons including σ electrons in SWNTs. The LDM method using the CNDO/S Hamiltonian was employed for spectral calculations of poly(p-phenylenevynelene) aggregates [67]. Stewart reparametrized MNDO to an MNDO-PM3 (MNDO parametric method 3) or a PM3 [73] Hamiltonian, which substantially reduces errors in calculations of the heats of formation as compared with the MNDO and AM1 Hamiltonians. The PM3 Hamiltonian has been employed to

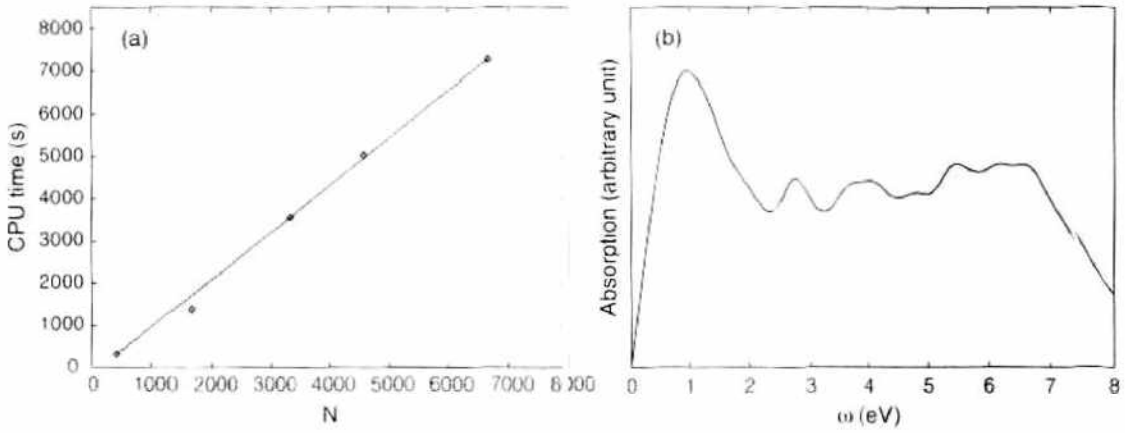


Figure 7. (a) CPU time for the excited-state calculation of three-dimensional CNTs with a large radius. The time interval is [−0.5 fs, −0.3 fs] with the time step 0.01 fs. The critical length l_0 and l_1 are 15 Å. Twenty-six atoms are included in the smallest box. (b) Absorption spectrum of the zigzag (60, 0) nanotube $C_{1200}H_{120}$. $\gamma = 0.4$ eV. Reprinted with permission from [49], W. Z. Liang et al., *J. Am. Chem. Soc.* 104, 2445 (2000). © 2000, American Chemical Society.

calculate the absorption spectra of organic compounds [74–76]. For the remainder of the section, the LDM-PM3 method will be used for calculations of SWNT optical absorption spectra.

A series of SWNTs have been investigated systematically for their optical properties by the LDM method [49, 65, 66]. Two types of dipole-induced excitations have been identified; namely, the end modes (low energy, $\omega < 1.0$ eV) and tube modes (high energy, $\omega > 1.0$ eV). The low-energy excitations are electron–hole pairs confined within the two ends of SWNTs, and the higher energy excitations are located mainly along the tube [65]. Those SWNTs whose optical properties have been calculated are mainly armchair and zigzag SWNTs., and their diameters are much greater than 4 Å. It was found that the absorption spectra of large-diameter carbon nanotubes are determined mainly by their diameters, with a weak dependence on their chiralities.

As early as in 1992, Sawada and Hamada [77] predicted the existence of extremely thin tubules; for instance, 4-Å SWNTs. They calculated the cohesive energies of the SWNTs using the Tersoff's empirical potential for carbon [78] and showed that SWNTs of all diameters, large and small, are energetically more favorable than the graphite sheets of the same width. They thus suggested that the 4-Å SWNTs may exist. Recently, both 4-Å SWNTs and multi-walled CNTs containing 4-Å SWNTs have been synthesized by the mass-selected carbon ion beam deposition (MSIBD) method [79, 80] and the pyrolysis of tripropylamine molecules in channels of porous zeolite $AlPO_4$–5 (AFI) single crystals [32], respectively. The 4-Å SWNTs have three possible structures: the chiral (4, 2) SWNT, the zigzag (5, 0) SWNT, and the armchair (3, 3) SWNT. Their diameters are 4.2, 3.9, and 4.1 Å, respectively. It has been argued that the 4-Å CNTs may be either (3, 3) or (5, 0) because they fit well with the half fullerene C_{20} cap [79]. The corresponding electronic structures and optical properties have been measured experimentally [81]. Three major absorption peaks at 1.35, 2.15, and 3.10 eV are identified in the absorption spectra when the electric field is parallel to the tubes. When the external field is perpendicular to the tube axis, the CNTs are almost transparent, with a very weak absorption peak at an energy slightly lower than 1.35 eV [81].

We first outline the implementation of the LDM method at the PM3 level denoted as the LDM-PM3 method. Next, in Section 3.2 we investigate electronic structures and optical properties of SWNTs of various chiralities, caps, and lengths. The nature of dipole-induced excitations is examined by projecting the corresponding reduced single-electron density matrices onto the Hartree–Fock molecular orbital (HFMO) representation. In Section 3.3, we report calculations of optical absorption spectra of the (4, 2), (3, 3), and (5, 0) SWNTs with diameters around 4 Å [65]. The anisotropy of absorption spectra is investigated by calculating the dynamic polarizabilities for various light polarizations. In Section 3.4, we study the effect of functionalization on metallic SWNTS with the help of the LDM-PM3 method.

3.1. The PM3 Hamiltonian

The PM3 Hamiltonian in the presence of an external field **E** is described as follows.

$$\begin{aligned} H &= H_e + H_{ee} + H_{ext} \\ H_e &= \sum_{ab}\sum_{mn} H_{ab}^{mn} c_{am}^{\dagger} c_{bn} \\ H_{ee} &= \frac{1}{2}\sum_{ab}\sum_{mnij} V_{ab}^{mn,ij} c_{am}^{\dagger} c_{bi}^{\dagger} c_{bj} c_{an} \\ H_{ext} &= -\mathbf{E}(t)\cdot\hat{\mathbf{P}} \end{aligned} \tag{30}$$

where $c_{am}^{\dagger}$ (c_{bn}) is the creation (annihilation) operator for an electron at a localized atomic orbital m (n) on atom a (b). One-electron integral H_{ab}^{mn} may be expressed as

$$H_{ab}^{mn} = \left\langle \chi_a^m \left| -\frac{1}{2}\nabla_{\mathbf{r}}^2 + U(\mathbf{r}) \right| \chi_b^n \right\rangle \tag{31}$$

where χ_a^m (χ_b^n) is the m (n)-th atomic orbital on atom a (b) and $U(\mathbf{r})$ is the one-electron potential. The Hamiltonian H_{ee} is the two-electron part of the Hamiltonian that represents the effective electron–electron Coulomb interaction. The PM3 model uses the neglect of differential overlap for atomic orbitals on different atoms; that is, all the two-electron integrals are set to zero except that when the orbitals m and n belong to the same atom a and i and j belong to atom b. The term $V_{ab}^{mn,ij}$ is expressed as

$$V_{ab}^{mn,ij} = \langle \chi_a^n(1)\chi_b^i(2)|V(r_{12})|\chi_a^m(1)\chi_b^j(2)\rangle \tag{32}$$

The Hamiltonian H_{ext} is the interaction between the valence electrons and an external electric field $\mathbf{E}(t)$, and $\hat{\mathbf{P}}$ is the molecular dipole moment operator. As a consequence, the Fock matrix h may be written as

$$h_{ab}^{mn} = H_{ab}^{mn} + 2\delta_{ab}\sum_{c}\sum_{ij\in c} V_{ac}^{mn,ij}\rho_{cc}^{ij} - \sum_{i\in a}\sum_{j\in b} V_{ab}^{mi,jn}\rho_{ab}^{ij} \tag{33}$$

Similarly, the induced Fock matrix may be expressed as

$$\delta h_{ab}^{mn} = 2\delta_{ab}\sum_{c}\sum_{ij\in c} V_{ac}^{mn,ij}\delta\rho_{cc}^{ij} - \sum_{i\in a}\sum_{j\in b} V_{ab}^{mi,jn}\delta\rho_{ab}^{ij} \tag{34}$$

3.2. Absorption Spectra of Carbon Nanotubes

To test the validity of the PM3 model for determining optical spectroscopy, we calculate the absorption spectrum of a C_{60} molecule. The result is demonstrated in Fig. 8. The geometry optimization for C_{60} is carried out at the Hartree–Fock level, using the PM3 Hamiltonian. The calculated energies of the first three main absorption peaks in C_{60} are 3.8, 4.6, and 5.7 eV, which compare well to the corresponding experimental values of 3.7, 4.6, and 5.7 eV for C_{60} in the n-hexane solution [82]. The calculations are performed using our LDM program with inclusion of all reduced single-electron density matrix elements (i.e., a full TDHF calculation).

A series of CNTs with different chiralities and ends are investigated, and their absorption spectra are determined. The external field is polarized along the tube axis in all calculations. Most CNTs' geometries are optimized at the Hartree–Fock level except those specified otherwise in the text. In our calculations, open-ended CNTs are terminated with hydrogen atoms or other functional groups. Capped CNTs are closed with fullerene-like cages that contain only hexagonal and pentagonal faces. For instance, a C_{60} molecule is bisected at its equator, and the two resulting half spheres may be attached to a open-ended (5, 5) armchair tube or (9, 0) zigzag tube depending on the way that the C_{60} molecule is cut. Six pentagons are needed to form a cap for (5, 0) tube, and six pentagons together with one hexagon for

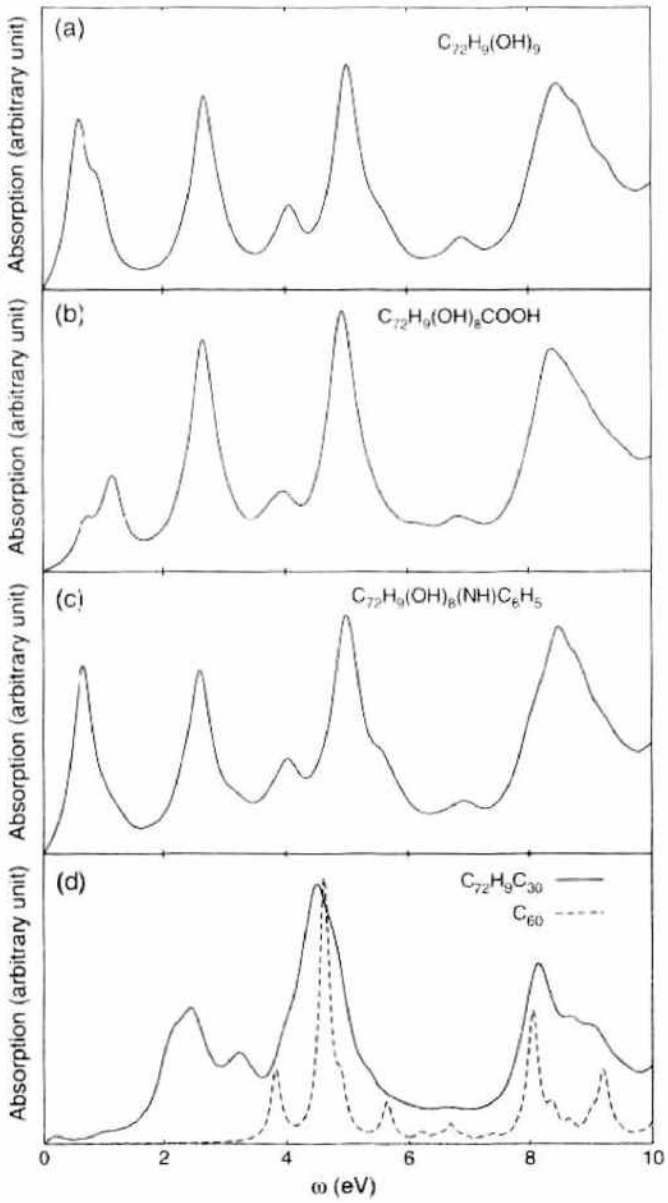


Figure 8. Calculated absorption spectra of C_{60} with dephasing parameter $\gamma = 0.1$ eV and the (9, 0) CNTs with $\gamma = 0.2$ eV. (a), (b), (c), and (d) are the absorption spectra for $C_{72}(H)_9(OH)_9$, $C_{72}(H)_9(OH)_8(COOH)$, $C_{72}(H)_9(OH)_8(NH)(C_6H_5)$, and $C_{72+39}H_9$ by full TDHF, respectively. All the geometries are optimized on the HF level with the PM3 Hamiltonian. Reprinted with permission from [65], W. Z. Liang et al., *J. Am. Chem. Soc.* 122, 11129 (2000). © 2000, American Chemical Society.

(6, 0) tube, as a total of 12 pentagons are needed to form a closed polygon, which is required by the Euler's theorem and sp^2 hybridization.

The influence of two ends on the optical properties of CNTs is examined. The calculated absorption spectra are shown in Fig. 8 for (9, 0) CNTs of same length but different ends. Figure 8a, 8b, 8c, and 8d shows the spectra of $C_{72}H_9(OH)_9$ (one end with —OH and another end with hydrogens), $C_{72}(OH)_8H_9$ (COOH) (one —OH is replaced by —COOH), $C_{72}(OH)_8H_9(NH)C_6H_5$, (one —OH is replaced by $-(NH)C_6H_5$), and $C_{72+30}H_9$ (one end with caps, another end with hydrogens), respectively. Similar line shapes are observed for $\omega >$ 2.0 eV in all spectra. The peaks centered at about 2.7, 4.0, 5.0, 6.9, and 8.0 eV are observed in all systems and are attributed to the excitations along the tubes. They are not much influenced by the different terminating functional groups. Differences appear for $\omega < 2.0$ eV, which should correspond to the excitations at two tube ends. The highest occupied molecular orbits (HOMOs) and lowest occupied molecular orbits (LUMOs) of CNTs have large components at the ends. This is further verified by the density matrices of low-energy excitations.

The physical properties of infinite CNTs are determined by their chiralities and radii [22, 83–86]. For finite size tubes, we find that the tube length also plays a vital role in the electronic structure. Figures 9 and 10 are the calculated absorption spectra of different-size (9, 0) and (5, 5) CNTs. Figure 9 shows absorption spectra of capped (9, 0) CNTs. As the number of carbon atoms increases from 78 to 222 (Fig. 9a–d), the peaks at $\omega < 6.0$ eV red-shift, although these near 8.0 eV change little. As a consequence of the red-shift, the line shapes differ for $\omega < 6.0$ eV until a saturation is reached. The absorption spectra of open-ended (9, 0) tubes differ drastically from those of capped (9, 0) CNTs, especially when the tube lengths are short. Low-energy peaks ($\omega < 2.0$ eV) for $C_{72}H_9(OH)_9$ (Fig. 8a) disappear in the absorption spectrum of C_{72+60} (Fig. 9c). The absorption spectrum of the open-ended (9, 0) tube changes drastically with increasing size. The first two peaks of $C_{72}H_9(OH)_9$ disappear when the number of carbon atoms increases to 162, whereas the third peak shifts below 2.0 eV. Investigation of corresponding density matrices reveals that the first peak for $C_{72}H_9(OH)_9$ corresponds to the electron–hole pairs located at two ends. The two ends play an important role in the optical response of short, open-ended zigzag tubes, and their

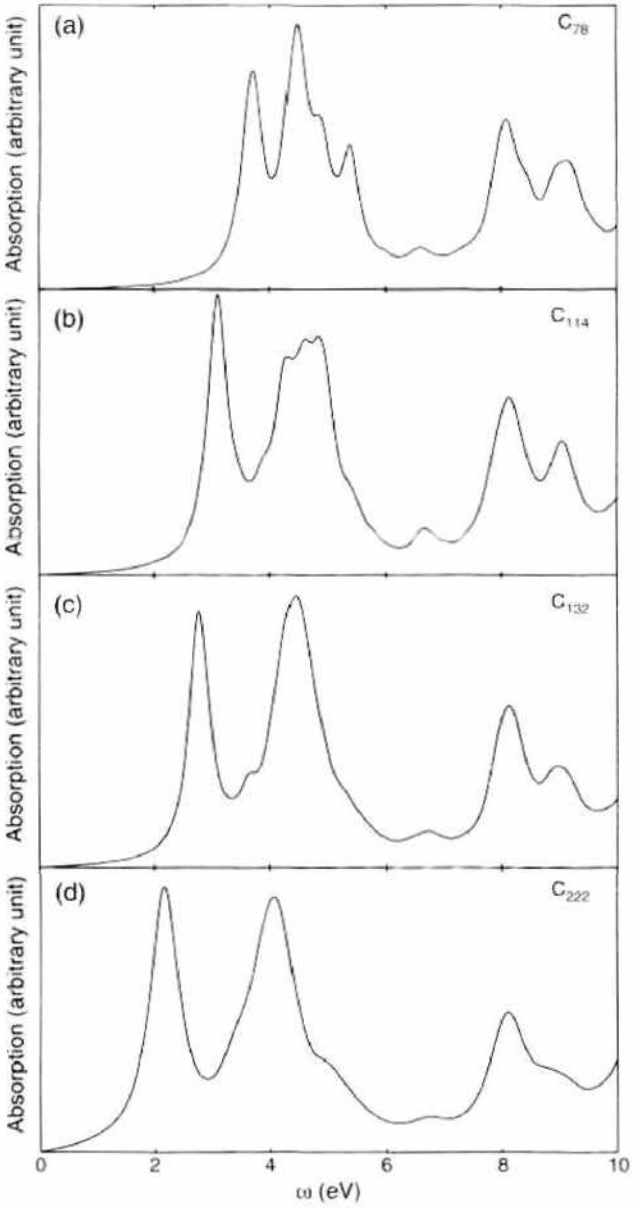


Figure 9. Absorption spectra of capped (9, 0) tubes calculated by full TDHF and LDM method. (a), (b), and (c) are the results calculated by full TDHF with $\gamma = 0.2$ eV, and (d) is the result of LDM with $\gamma = 0.3$ eV. All the geometries are optimized. Reprinted with permission from [65], W. Z. Liang et al., *J. Am. Chem. Soc.* 122, 11129 (2000). © 2000, American Chemical Society.

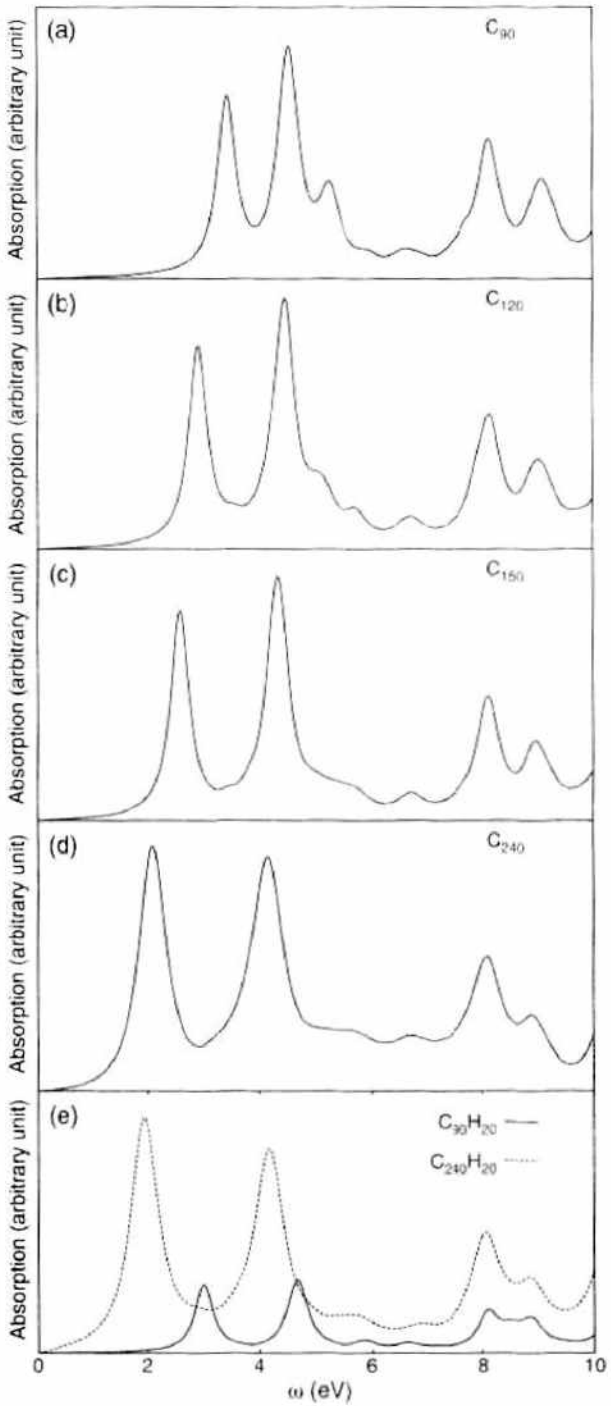


Figure 10. Absorption spectra of open-ended and capped (5, 5) armchair tubes calculated by full TDHF and LDM method. (a), (b), (c), and (d) are the absorption spectra of the capped (5, 5) tube. (e) Absorption spectra of open-ended (5, 5) tubes. The solid line is for $C_{10j}H_{20}$, $j = 9$, and the dashed line is for $j = 24$. (a), (b), (c), and (e) (solid line) are calculated with $\gamma = 0.2$ eV, whereas $\gamma = 0.3$ eV is employed in (d) and (e) (dashed line). All the geometries are optimized. Reprinted with permission from [65], W. Z. Liang et al., *J. Am. Chem. Soc.* 122, 11129 (2000). © 2000, American Chemical Society.

influence diminishes as the size increases. Absorption spectra of several SWNTs with different end groups have been measured, and little variance of the spectra has been observed [87]. This is consistent with our finding that the end groups of long SWNTs have little effect on their absorption spectra. The absorption spectra of (5, 5) CNTs (see Fig. 10) are similar to those of capped (9, 0) CNTs. This is because that both tubes have similar radii [3.5 Å for (9, 0) and 3.4 Å for (5, 5)]. The tight-binding calculations have predicted the same electronic and optical behavior for (5, 5) and (9, 0) tubes [21]. The optical absorption spectra of capped and open-ended (5, 5) look similar. For instance, a mere small red-shift is observed

for the open-ended (5, 5) tube $C_{90}H_{20}$ (solid line in Fig. 10e), as compared to capped C_{90+60} tube (Fig. 10c) of the same cylindrical length.

Computed optical absorption spectra of two SWNTs, (9, 0) (C_{204}) and (5, 5) (C_{280}), are shown in Fig. 11. Both SWNTs are capped by a bisected C_{60} molecule at their equators. Geometries are optimized by PM3, and the LDM-PM3 method is employed for the absorption spectra calculation. Solid lines are for (9, 0) and dashed lines are for (5, 5). In Fig. 11a, the external field **E** is along the tube axis, whereas **E** $\perp$ **T** in Fig. 11b. The diameters of (9, 0) and (5, 5) are 6.9 and 7.1 Å, respectively. In other words, (9, 0) and (5, 5) are of the similar diameters. Clearly, their absorption spectra are quite similar as well. This supports our conclusion that the chirality of a larger-diameter SWNT has little effect on its optical absorption spectrum [65].

The energy of the first major peak (or optical gap) of capped (9, 0) and (5, 5) CNTs versus $1/N$ is plotted in Fig. 12, where N is the number of carbon atoms. The two dashed lines are the linear fits and are almost identical. The optical gaps approach to $E_g = 1.23$ eV as the sizes of both tubes approach infinity. Because the radii of two CNTs are similar, the resulting optical gaps for two infinite long tubes have the same values up the second decimal digits. It may be generalized that the optical gaps for infinite long tubes are finite and their values depend mainly on the tube radii. The optical absorption spectra of finite-size capped (9, 0) and (5, 5) CNTs have been calculated by the tight-binding model with only π orbitals of carbon atoms considered. The tight-binding optical gaps for a (9, 0) tube with $N = 420$ and a (5, 5) tube with $N = 250$ are 1.2 and 1.25 eV, respectively [88, 89]. These are comparable to our extrapolated optical gaps for infinitly long CNTs. The optical signal at ~1.2 eV has been observed in the optical conductivity and absorption measurement of SWNTs [29, 90]. In addition, a gap of 1.2 eV has been found in the calculated DOS spectrum for (9, 0) by the tight-binding method [21].

Absorption spectra of capped (5, 0) tubes with $N = 150$ and 250 and of capped (6, 0) tubes with $N = 156$ and 252 are shown in Fig. 13. The absorption spectra of (5, 0) with $N = 150$ and (6, 0) with $N = 156$ are calculated by the full TDHF method with a dephasing coefficient $\gamma = 0.2$ eV. The absorption spectra of two other tubes are calculated by the LDM method with $\gamma = 0.3$ eV and a cutoff distance $l_0 = l_1 = 29$ Å. As expected, the red-shifts occur when the system sizes increase. The optical gap of infinite (5, 0) tube approaches ~1.0 eV, as shown in the inset of Fig. 13(a). A much weak peak appears at $\omega \sim 1.4$ eV for

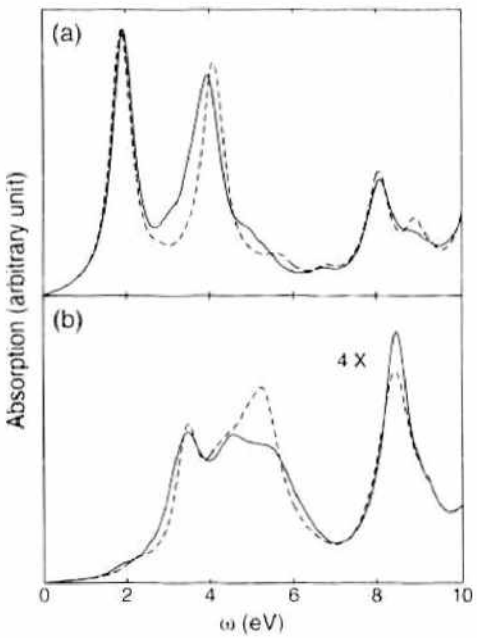


Figure 11. Calculated absorption spectra of capped (9, 0) C_{204} (solid line) and (5, 5) C_{280} (dashed line) SWNTs. The external field **E** is (a) along tube axis and (b) **E** $\perp$ **T**. The absorption spectra in (b) are magnified four times. The dephasing parameter $\gamma = 0.3$ eV. Reprinted with permission from [66], W. Z. Liang et al., *J. Am. Chem. Soc.* 123, 9850 (2001), © 2001, American Chemical Society.

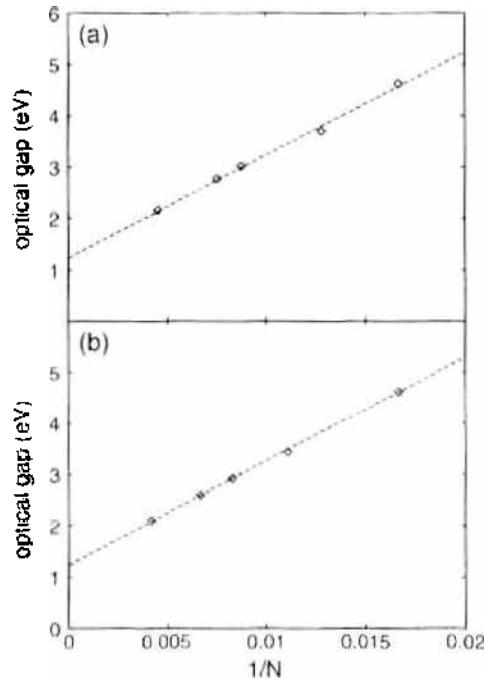


Figure 12. Optical gap via $1/N$ for capped (9, 0) and (5, 5) tubes. (a) (9, 0) tube, (b) (5, 5) tube. Reprinted with permission from [65], W. Z. Liang et al., *J. Am. Chem. Soc.* 122, 11129 (2000). © 2000, American Chemical Society.

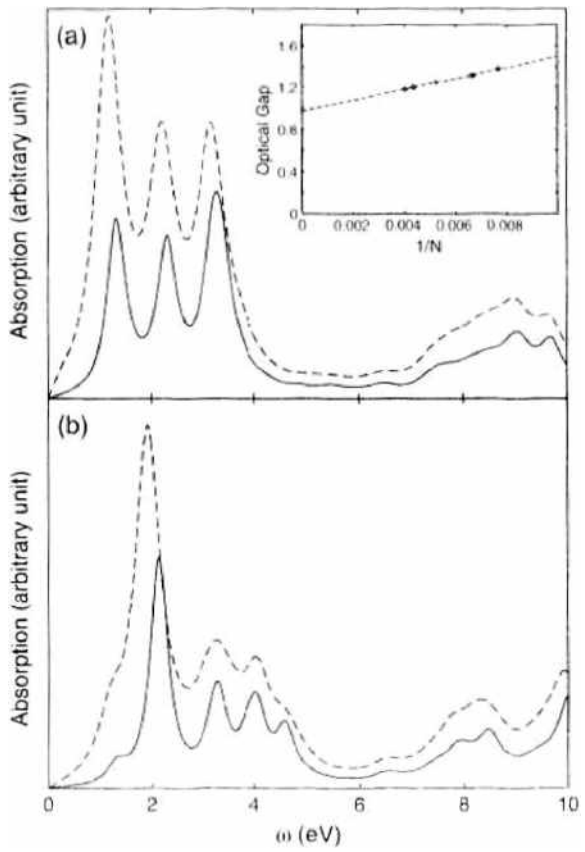


Figure 13. Absorption spectra of capped (5, 0) and (6, 0) tubes. (a) The absorption spectra of (5, 0) C_{40} (the solid line) and C_{50} (dashed line) calculated by full TDHF and LDM with dephasing $\gamma = 0.2$ and $\gamma = 0.3$ eV, respectively. The inset shows the optical gap versus $1/N$ for the (5, 0) tube. (b) The absorption spectra of the capped (6, 0) tube. The solid line for C_{56} is calculated by full TDHF with $\gamma = 0.2$ eV. The dashed line for C_{72} is calculated by LDM with dephasing $\gamma = 0.3$ eV. All the geometries are optimized. Reprinted with permission from [65], W. Z. Liang et al., *J. Am. Chem. Soc.* 122, 11129 (2000). © 2000, American Chemical Society.

he (6, 0) tube. This is because the optimized geometry for the (6, 0) tube has C_{2v} point roup symmetry which is different from the C_{3h} point group of the (9, 0) tube. Tubes like 9, 0) and (5, 5) have the first major absorption peak located at a larger transition energy, nd (5, 0) tubes of the same length possess relatively small optical gaps. This indicates that eometry plays a fundamental role in determining the optical behavior of CNTs.

A broad group of peaks located at 6.0 ~ 7.0 eV is observed in the absorption spectra f capped (5, 0), (6, 0), (9, 0), and (5, 5) tubes. These peaks shift little as the tube lengths ncrease and are identified mainly as $\pi - \pi^*$ transitions by the low-energy EELS experiments [29] and calculations [91]. The relative oscillator strength decreases with the reducing ube radius. This is caused by larger $\pi - \sigma$ hybridization in smaller-radius tubes. The larger he curvature, the more the π and σ hybridization.

Figure 14 shows the optical absorption spectra of an open-ended armchair tube ($C_{4mj}H_{2m}$), vith $j = 4$ and $m = 6$, 8, and 10. The ideal structures, which are rolled up from a single raphite sheet with all bond lengths set to 1.421 Å are adopted here. The absorption spectra ed-shift as the tube radius increases, except that the first peak blue-shifts slightly. The blue-hift may come from the competition between the size effect and the π orbital overlapping. When the external electric field is applied perpendicular to the tube axis, red-shifts of lowest eaks are observed when the radii of tubes increase. This is consistent with the expectation hat the lowest peak red-shifts as the radius increases (see Fig. 15, in which the optical gaps ersus $1/r$ is plotted with r being the radius of the tube). A linear relationship between the ap and $1/r$ is observed, and a gap of 0.7 eV is determined as $r \to \infty$. It was also found hat the absorption threshold is significantly higher for the external field polarized along

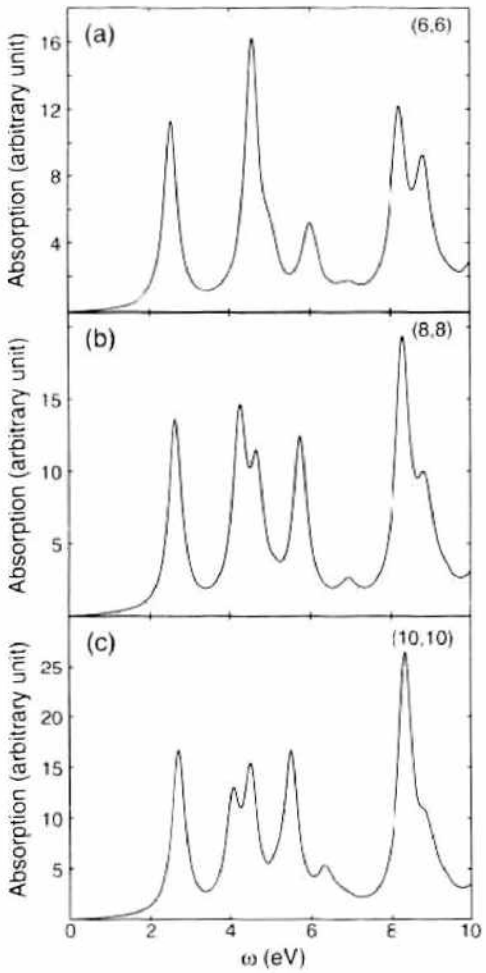


Figure 14. Absorption spectra of (m, m) armchair tube with $m = 6, 8, 10$ and dephasing $\gamma = 0.2$ eV. The ideal structure is employed. (a) (6, 6) $C_{12j}H_{24}$. (b) (8, 8) $C_{16j}H_{32}$. (c) The absorption spectra of (10, 10) $C_{20}H_{40}$ with $j = 8$. Reprinted with permission from [65], W. Z. Liang et al., *J. Am. Chem. Soc.* 122, 11129 (2000). © 2000 American Chemical Society.

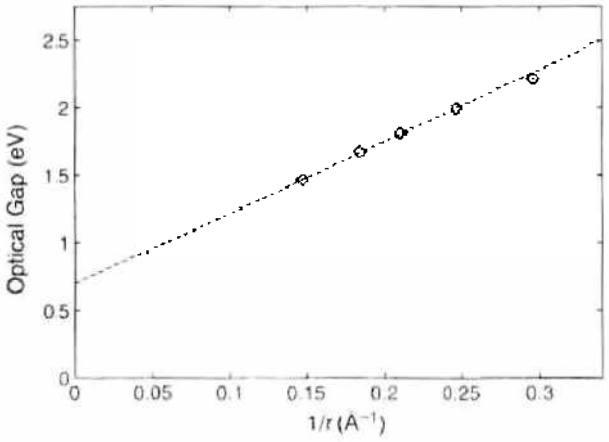


Figure 15. The optical gap via $1/r$ when the external field is applied perpendicular to the tube axis. Reprinted with permission from [65], W. Z. Liang et al., *J. Am. Chem. Soc.* 122, 11129 (2000). © 2000, American Chemical Society.

the tube axis than perpendicular to the tube axis. The absorption spectra of (6, 6), (8, 8), and (10, 10) are very different from that of (5, 5) (solid line in Fig. 10e). This is because different bond lengths are used here. Thus, it is determined that the nature of excitations is sensitive to the bond lengths. The density matrices in molecular orbital (MO) representation reveal that the first peak in armchair tubes come mainly from the HOMO $\rightarrow$ LUMO + 1 and HOMO − 1 $\rightarrow$ LUMO transitions. The HOMO $\rightarrow$ LUMO transition is forbidden. The other peaks red-shift with the increasing tube radius. The relative oscillator strength of these peaks that center at relatively high energies (from 4.3 to 8.2 eV) increase with the increasing radius. These peaks correspond to $\pi - \pi^*$ and small-fraction $\pi - \sigma^*$ transitions.

We proceed to discuss the single-electron density matrices of the dipole-induced excitations. To understand the nature of the electronic excitations, the induced density matrices $\delta\rho^{(1)}(\omega)$ of $C_{72}H_9(OH)_9$ are examined at 0.61, 2.67, and 5.01 eV; C_{72+60} at 2.77, 6.72, and 8.14 eV; and $C_{80}H_{20}$ at 2.89, 4.79, and 8.04 eV. $C_{72}H_9(OH)_9$ and C_{72+60} are open-ended and capped (9, 0) tubes, respectively, and $C_{80}H_{20}$ is an open-ended (5, 5) tube. The results are shown in the Figs. 16, 17, and 18. The atomic orbital (AO) representation is employed. The atomic indices are assigned increasingly from one end of the tube to the other, and the orbital indices are arranged in the order of $2s$, $2p_x$, $2p_y$, and $2p_z$. The absolute values of density matrix elements are shown in the contour plots. Logarithmic scale is employed. The scales employed for ground- and excited-state density matrices are shown in Fig. 16e and Fig. 16f, respectively. From the contour plots in Figs. 16a, 17a, and 18a, it is observed that the ground-state density matrices of the three systems are almost diagonal. However, the electron coherence between the two ends is stronger in the opened-end (9, 0) tube than in the capped (9, 0) and (5, 5) CNTs. The strong electron coherence between the two open ends of (9, 0) has a significant influence on the optical properties. The excited-state density matrices are obtained by the Fourier transformation

$$\delta\rho(\omega) = \int dt\, e^{i\omega t}\delta\rho(t) \tag{35}$$

In the actual TDHF calculation, dephasing γ is added to calculate the time evolution of $\delta\rho(t)$. The term $\delta\rho(\omega)$ contains not only the effect of the mode at ω but also the effect of other modes. When γ is much less than the energy differences of different excitations, the mixing of excitations in $\delta\rho(\omega)$ ($\omega \sim \Omega_v$) is negligible. Therefore, $\mathrm{Im}[\delta\rho(\Omega_v)]$ is a very good approximation of the reduced-density matrix for the excitation at Ω_v. The density matrices shown take the form $\{\mathrm{Im}[\delta\rho(\omega)] + \mathrm{Im}[\delta\rho^t(\omega)]\}/(i\sqrt{2})$, where $\delta\rho^t(\omega)$ is the transpose of $\delta\rho(\omega)$. Three excited-state density matrices of an open-ended (9, 0) tube are shown in Fig. 16b–d. The first excitation at $\omega = 0.61$ eV in $C_{72}H_9(OH)_9$ includes mostly the π electron–hole pairs from the two ends. This peak disappears in the absorption spectra of capped (9, 0) tubes. The contribution from the electron–hole pairs among the π orbitals

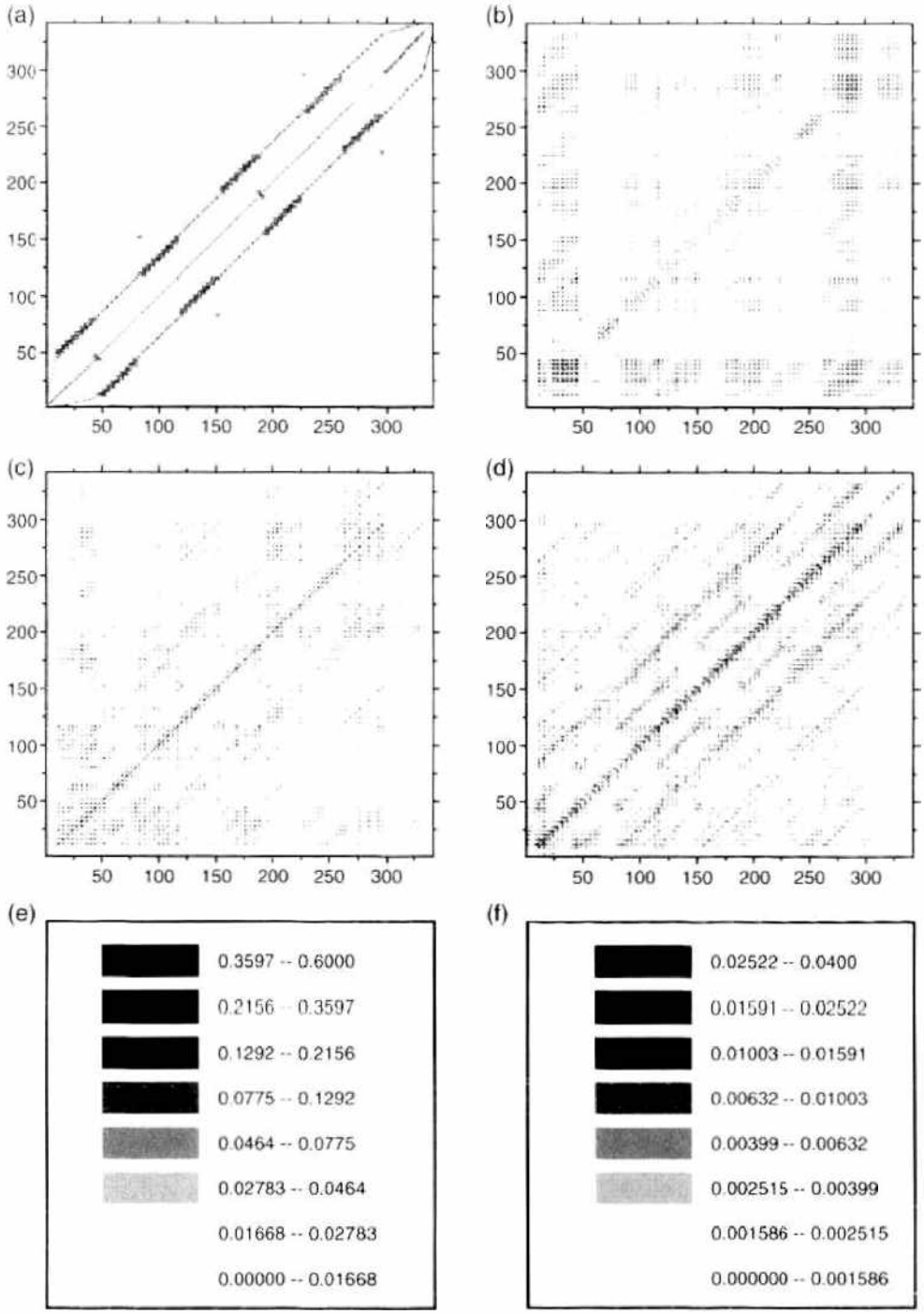


Figure 16. Density matrices of ground states and various excited states for an open ended (9, 0) $C_{72}H_9(OH)_9$ tube in AO representation by full TDHF with $\gamma = 0.1$ eV. Matrix elements are shown on a gray logarithmic scale. (a) The ground state; (b) $\omega = 0.61$ eV; (c) $\omega = 2.67$ eV; (d) $\omega = 5.01$ eV; (e) scale of all the ground-state density matrices; (f) scale of all the excited state density matrices. Reprinted with permission from [65], W. Z. Liang et al., *J. Am. Chem. Soc.* 122, 11129 (2000). © 2000, American Chemical Society.

located in the middle of tube increases for the excitation at $\omega = 2.67$ eV. The excitation at 5.01 eV in $C_{72}H_9(OH)_9$ contains π as well as σ orbital contributions. In general, a contribution that comes from the electron–hole pairs in the middle increases rapidly as energy increases. As expected, the oscillator strengths of these excitations increase and their energies red-shift with increasing sizes [49].

Figure 17b–d depict the excited-state density matrices of capped (9, 0) tube C_{72+60} at energies 2.77, 6.72, and 8.14 eV, whereas Fig. 18b–d show the excited-state density matrices of an open-ended (5, 5) tube at energies 2.89, 4.79, and 8.04 eV, respectively. The contributions from two ends are much weaker than those in $C_{72}H_9(OH)_9$; that is, the electron coherence

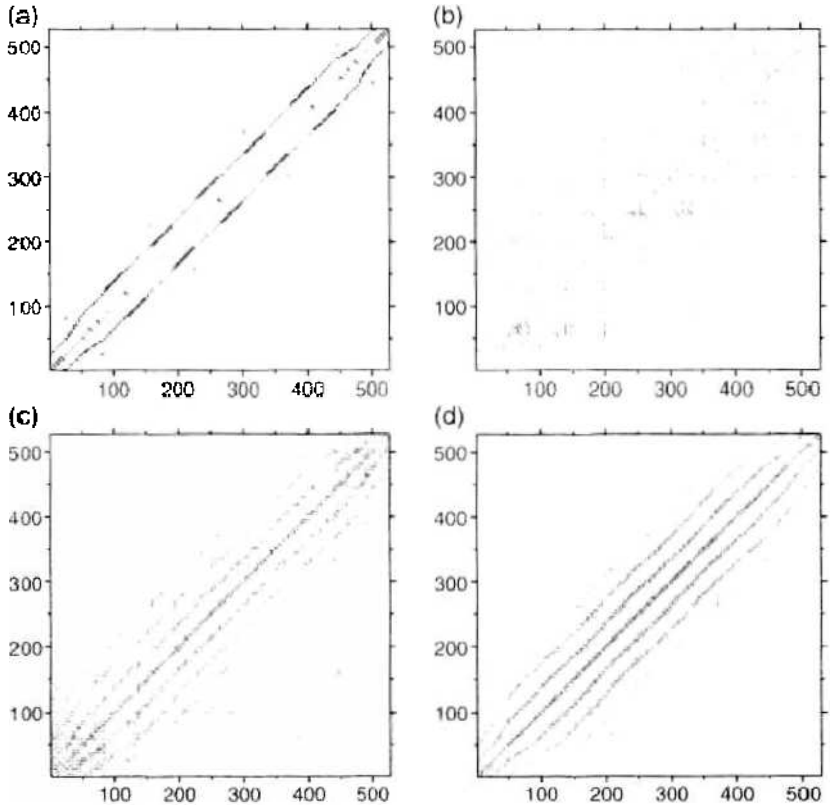


Figure 17. Density matrices of ground states and various excited states for capped (9, 0) tube C_{132} in AO representation by full TDHF with $\gamma = 0.1$ eV. (a) The ground state; (b) $\omega = 2.77$ eV; (c) $\omega = 6.72$ eV; (d) $\omega = 8.14$ eV. Reprinted with permission from [65], W. Z. Liang et al., *J. Am. Chem. Soc.* 122, 11129 (2000). © 2000, American Chemical Society.

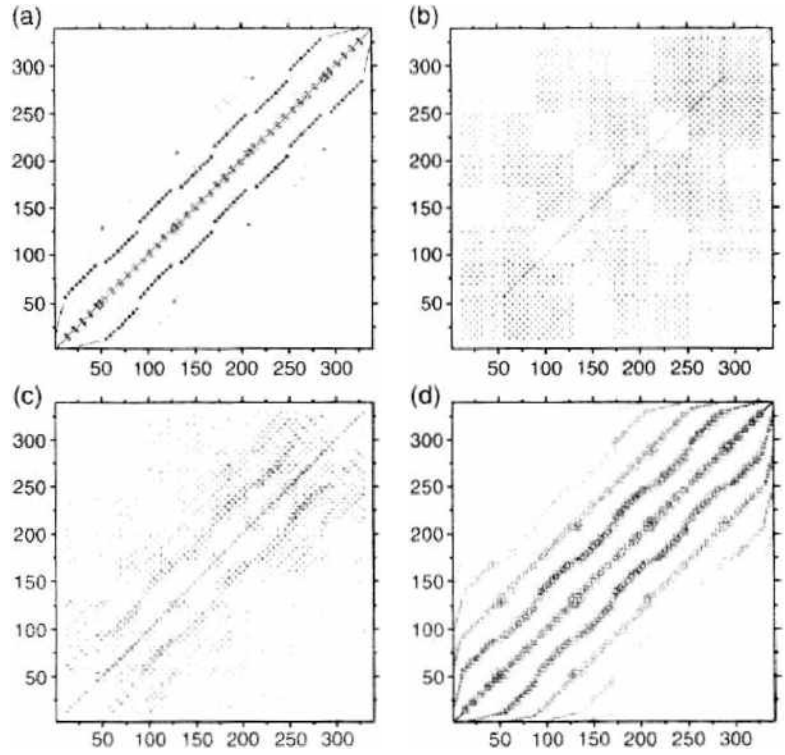


Figure 18. Density matrices of ground states and various excited states for open-ended (5, 5) tube $C_{80}H_{20}$ in AO representation by full TDHF with $\gamma = 0.1$ eV. (a) The ground state; (b) $\omega = 2.89$ eV; (c) $\omega = 4.79$ eV; (d) $\omega = 8.04$ eV. Reprinted with permission from [65], W. Z. Liang et al., *J. Am. Chem. Soc.* 122, 11129 (2000). © 2000, American Chemical Society.

between two ends is weak. The main contributions are from the middle of the tubes. This is the reason why the absorption spectra are rather similar for (5, 5) and capped (9, 0) tubes. The peaks centered at the low-energy of 2.77 eV in Fig. 17b and at 2.89 and 4.79 eV in Fig. 18b and 18c result mostly from π electron–hole pairs. The first peak in both systems comes mostly from HOMO $\rightarrow$ LUMO + 1 and HOMO − 1 $\rightarrow$ LUMO transitions. Transitions HOMO $\rightarrow$ LUMO + 2 and HOMO − 2 $\rightarrow$ LUMO also contribute significantly to the first peak ($\omega = 2.77$ eV) of the capped (9, 0) tube. Other transitions between the HOMO − m and LUMO + n also effect the first peak of the (9, 0) and (5, 5) tubes. The white squares of the "chess board" pattern in Fig. 17b correspond to the electron coherence between two orbitals that belong to two atoms located, respectively, at the nth ring and $n \pm j$th rings, where j is a positive odd number. The antidiagonal part is nearly zero in Figs. 17b and 18b. The white square in the antidiagonal part corresponds to the elements between the two orbitals that belong to the two atoms located, respectively, at two mirror rings, which are symmetric with respect to a center plane equally dividing the nanotube. Their values are almost zero except for the pairs of π orbitals, which belong to two mirror symmetric atoms. The σ electrons have larger contribution in the high- energy range; for instance, the peaks at 6.72 and 8.14 eV (Fig. 17c and 17d) in capped (9, 0) tube and the peak at 8.04 eV (Fig. 18d) in the (5, 5) tube. These peaks originate from the main $\pi - \pi^*$ transition as well as the partial $\pi - \sigma^*$ transition, which is revealed by the corresponding density matrices in MO representation. These excitations have been observed by low-energy EELS experiment [92]. The electron coherence among $\pi - \pi^*$ has larger spatial extents than that among $\pi - \sigma^*$, $\sigma - \pi^*$, and $\sigma - \sigma^*$ (see Fig. 17c, 17d). Examining the excited-state density matrices of these three systems reveals that π orbitals are mainly responsible for electron excitations, with $\omega < 8.0$ eV, while σ electrons are responsible for higher-energy electron excitations. The patterns of the reduced-density matrix contour plots reflect the structure features of CNTs. For instance, the stripes in Figs. 16a, 16d, 17a, 17c, 17d, 18a, 18c, and 18d are the manifestation of the underlying ring structure of CNTs. The "chessboard" pattern in Fig. 17b reflects large electron coherence within the same ring or between adjacent rings (the dark squares), and diminishing electron coherence between adjacent or next-nearest rings.

The STM may probe directly one electron orbital density or DOS. Figure 19 shows the DOS of several CNTs; $C_{300}H_{20}$, $C_{316}H_{20}$, and $C_{320}H_{20}$. The chiralities of above three CNTs are (6, 4), (7, 3), and (5, 5), respectively. The ideal structures are employed. Their radii are 3.415, 3.482, and 3.393 Å, and the lengths are 35.3, 34.91, and 38.14 Å, respectively. The Fermi energy level is between the HOMO and LUMO and is set to zero. According to the tight-binding model, a CNT with a chirality (m, n) is a conductor when $m - n \bmod 3 = 0$. A CNT with other chiralities is a semiconductor. Our calculation shows that the HOMO-LUMO gap for a (5, 5) is 2.8 eV, whereas (7, 3) and (6, 4) have much smaller gaps, although three radii and lengths are of the similar values. The DOS spectra are determined by their chiralities of CNTs. Further, we find that the HOMO-LUMO energy difference decreases

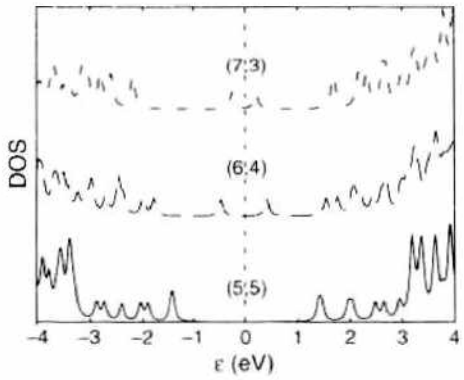


Figure 19. DOS of HF molecular orbitals of (5, 5) $C_{320}H_{20}$, (7, 3) $C_{312}H_{20}$, and (6, 4) $C_{300}H_{20}$ by the PM3 Hamiltonian. The Fermi level is set to $\epsilon = 0.0$ eV. The energy resolution $\eta = 0.05$ eV is employed. Reprinted with permission from [65], W. Z. Liang et al., *J. Am. Chem. Soc.* 122, 11129 (2000). © 2000, American Chemical Society.

overall as the nanotube length increases and the energy gap in the DOS spectrum may be different from the HOMO-LUMO energy difference. This energy gap and DOS spectrum saturate when the length is long enough. Our calculated HOMO-LUMO energy gap of (5, 5)) tubes is larger than that of the tight-binding calculation for CNTs of the same length because the HOMO-LUMO energy gaps predicted by tight-binding calculation saturate more rapidly than those by *ab initio* HF or semiempirical methods [93]. It is worthy pointing out that the DOS of (5, 5) calculated by the tight-binding model has a similar appearance to our result, and a gap of 2.0 eV exists. Our calculated DOS is not symmetric with respect to the Fermi level, which is consistent with the experiment [94] but differs from the tight-binding results [95–97]. The difference stems from $\pi - \sigma$ hybridization [98].

The absorption spectra of a series of CNTs with different sizes, chiralities, ends, and bond lengths are calculated by the LDM method with the PM3 Hamiltonian. The dipole-induced excitations may be categorized into end modes (low energy) and tube modes (high energy). These modes have been characterized by examining their reduced single-electron density matrices. It has been found that their optical properties are very much affected by the tube length, radius, end group, and chirality, as well as the bond length. The density matrices of various excitations exhibit interesting features that relate to the structural features of CNTs. The main results are summarized as follows.

The low-energy dipole-induced excitations appearing in the absorption spectra in Fig. 8 are the end modes; that is, the electron–hole pairs that reside mostly at the two ends of tube. These excitations come mainly from $\pi - \pi^*$ transitions. The corresponding absorption peaks have been observed only for short, open-ended, zigzag CNTs in our calculation. The excitation energy and oscillator strength depend sensitively on the tube length, end groups and chiralities. As the length increases, the low-energy ($\omega \leq 1.0$ eV) absorption peaks red-shift, and their oscillator strengths decrease. When the end groups are altered, the profile of low-energy-absorption spectra changes drastically for short CNTs. Their sensitivities to the end groups may be used to design new SWNT-based materials. It is interesting to note that the capped zigzag CNTs and armchair tubes do not have such low-energy absorption peaks. We emphasize that the absorption spectra of long SWNTs are affected little by the caps or end groups. This has been confirmed experimentally [87].

For $\omega \geq 2.0$ eV in Fig. 8, the dipole-induced excitations are regarded as the tube modes; that is, the electron–hole pairs residing in the middle of the tube. The energies and oscillator strengths of these tube modes may depend on the tube lengths, radii, and bond-length alternations. However, they are not sensitive to the chirality and end groups. It is emphasized that the absorption spectra for capped (9, 0), capped (5, 5), and open-ended (5, 5) tubes of the same length are strikingly similar.

The tube modes may be divided further into two groups of excitations: $2.0 \leq \omega \leq 6.0$ and $\omega \geq 6.0$ eV. For the excitations whose energies ω are between 2.0 and 6.0 eV, the electron–hole pairs arise from $\pi - \pi^*$ transitions. These excitations are sensitive to the tube length. As the tube length increases, the corresponding absorption peaks red-shift (see Figs. 8–13). The absorption spectrum saturates when the tube length reaches a few nanometers. The calculated optical gaps are consistent to the energy gaps measured by STM experiment [29]. When the radii increase, most absorption peaks in the range ($2.0 \leq \omega \leq 6.0$ eV) red-shift except the first absorption peaks of armchair CNTs (see Fig. 14), which blue-shift slightly when the external field is polarized along the tube. This may be explained by the competing effects of the hopping matrix elements of π electrons and the tube radii.

For the dipole-induced excitations with $\omega \geq 6.0$ eV, their energies and oscillator strengths are not sensitive to the length, chiralities, and ends, but their oscillator strengths are sensitive to the tube radii and bond-length alternation. The broad peaks at $6.0 \sim 7.0$ and $8.0 \sim 9.0$ eV are mostly from the $\pi - \pi^*$ transitions, but a small fraction of $\pi - \sigma^*$, $\sigma - \pi^*$, and $\sigma - \sigma^*$ transitions also contribute to these peaks.

For the open-ended tubes containing a few hundreds of atoms, it is observed that the HOMO–LUMO gap for armchair CNTs is quite large [~2.0 eV for (5, 5)]. This confirms results from earlier *ab initio* Hartree–Fock and semiempirical calculations [93]. It was postulated that this larger HOMO–LUMO gap is a result of the finite size effect. As the tube

length increases to infinity, the gap may vanish. Our calculation shows evidence for such a postulation.

Finally, the optical gaps of CNTs are finite and approach nonzero values as the tube lengths become infinity. It is illustrated that the optical gaps depend linearly on $1/N$. The inverse dependence of the optical gap on diameter has been investigated in armchair tube (m, m). It is found that the optical gaps depend inversely on the diameter of the tubes.

In summary, the dipole-induced excitations may be categorized into the end modes and tube modes. The end modes are sensitive to the tube length, chirality, and radius in addition to the bond length. When the tube length becomes long enough or the ends are closed with carbon cages, the corresponding low-energy absorption peaks disappear. The tube modes are of higher energies and are less sensitive to the tube length and chirality. The dipole-induced transitions below 8.0 eV are composed mainly of $\pi - \pi^*$ transitions. Furthermore, it has been found that the optical gap scales linearly with $1/N$ and is finite when the number of carbon atoms N approaches infinity. Although the precise values of excitation energies and oscillator strengths may depend on detailed structures of CNTs, the above qualitative conclusion on the composition and categorization of CNTs' optical excitations remains valid.

3.3. Polarization-Dependent Optical Absorption Spectra of 4-Å Single-Walled Carbon Nanotubes

The narrowest possible nanotube has a diameter of 4 Å, the size at which a SWNT remains energetically stable. Some of the smallest CNTs are first found capped inside an 18-shell nanotube [79]. Lately, 4-Å SWNTs have been fabricated inside inert AFI zeolite channels [32, 33]. Chen et al. have examined the optical properties of 4-Å SWNTs using the LDM method [65].

The two ends of the SWNTs are terminated with hydrogen atoms in the calculations. The ideal structures are employed; that is, the SWNTs are constructed by rolling graphite segments along the tube axis with the C–C bond length set to 1.42 Å, except that the structures of (4, 2) SWNTs are optimized by the PM3 calculations. In a neutral (5, 0) CNT electrons cannot fill the closed-shell structure. The unpaired electrons result in a net dipole moment that prevents the converge of the self-consistent field (SCF) calculation [99]. Four extra electrons are added to the (5, 0) CNT, on which the self-consistent computation converges and a set of MOs are obtained. It is expected that the extra four electrons have little effect on the optical response when the number of carbon atom reaches 200 or more. Figures 20, 21, and 22 show calculated absorption spectra of (4, 2), (3, 3), and (5, 0) SWNTs. Each unit cell of the (4, 2), (3, 3), and (5, 0) CNTs has 56, 12, and 20 carbon atoms, respectively. All solid lines are the absorption spectra corresponding to $\mathbf{E} \parallel \mathbf{T}$, and the dashed lines correspond to $\mathbf{E} \perp \mathbf{T}$. The calculated absorption spectra of (4, 2), (5, 0), and (3, 3) CNTs are quite different despite the fact that their diameters are almost the same. This differs from our previous results from larger-diameter CNTs, which showed that the absorption spectra are insensitive to the chiralities of CNTs when the tube lengths are long enough. In the case in which $\mathbf{E} \parallel \mathbf{T}$, the optical spectra of the three SWNTs depend very much on the tube lengths. As the number of carbon atoms increases, the absorption spectra of the open-ended chiral (4, 2) tubes change drastically, especially when the tube lengths are relatively short. The overall spectra red-shift and the relative amplitude in the low-energy range ($\omega < 1.0$ eV) reduce as the number of carbon atom N increases. When the (4, 2) SWNTs have 200 or more carbon atoms, these peaks disappear when compared to the others (see Fig. 20). The open-ended (3, 3) and (5, 0) SWNTs have a different response to the external field as compared to the (4, 2) SWNTs. In the low-energy range ($\omega < 1$ eV), no peaks are found even for very short tubes (see Figs. 21 and 22). The two ends play an important role in the optical response for some short tubes; for instance, the (4, 2) CNTs. Their influence recedes as the length increases. The optical behaviors of the long tubes are affected little by the two ends, which is consistent with experimental observations [87, 90].

Two absorption peaks are found at 1.60 and 2.9 eV for (4, 2) $C_{332}H_{12}$, and three peaks are found at 1.16, 1.66, and 2.60 eV for (5, 0) $C_{300}H_{10}$. The (3, 3) SWNT has one distinctive

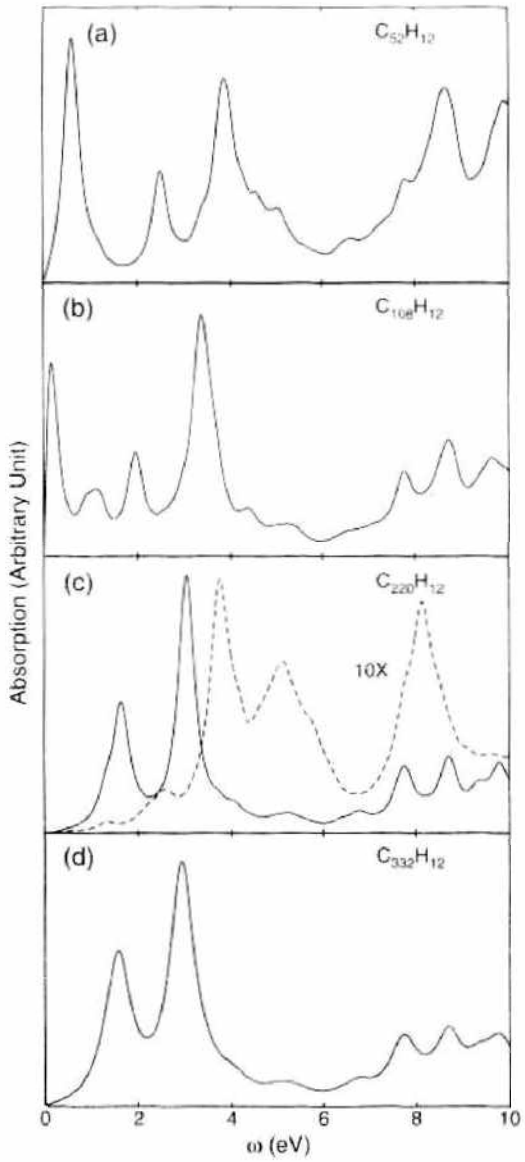


Figure 20. Calculated absorption spectra of (4, 2) SWNTs with dephasing parameter $\gamma = 0.2$ eV (a–c) and $\gamma = 0.3$ eV (d). The external field **E** is along tube axis except the dashed line in part (c) which is the result for $\mathbf{E} \perp \mathbf{T}$. The dashed lines are magnified 10 times. Reprinted with permission from [66], W. Z. Liang et al., *J. Am. Chem. Soc.* 123, 9830 (2001). © 2001, American Chemical Society.

strong absorption peak in the low-energy range for $N > 200$. It red-shifts when the tube length increases and saturates at 0.61 eV as $N \to \infty$ (see Fig. 23b). Similar red-shifts of absorption spectra are found for (4, 2) and (5, 0) SWNTs. These red-shifts are caused by the collective character of the excitation in terms of the single electron–hole excitation picture [59]. Because π electrons delocalize more than σ electrons, the red-shifts are prominent for $\omega < 4$ eV because the corresponding excitations are mainly $\pi - \pi^*$ transition. Note that the spectral profiles do not vary much after N reaches 200 or more. Therefore, $C_{332}H_{12}$, $C_{420}H_{12}$, and $C_{300}H_{10}$ are used to simulate, respectively, the infinite long (4, 2), (3, 3), and (5, 0) SWNTs. Given the fact that the absorption spectra are very different for (4, 2), (3, 3), and (5, 0), it is concluded that for small-diameter SWNTs, the optical response depends on the chirality of the tube, in addition to the diameter and the tube length. Therefore, the observed absorption spectra may be used to determine the structures of 4-Å SWNTs.

To investigate the anisotropy characters of (3, 3), (5, 0), and (4, 2) SWNTs, their absorption spectra are calculated by aligning the external field **E** perpendicular to the tube axis ($\mathbf{E} \perp \mathbf{T}$). The resulting absorption spectra are shown as dashed lines in Figs. 20c, 21a, 21b, 22a, and 22c. Strong anisotropy is observed for all three SWNTs. The absorption intensities

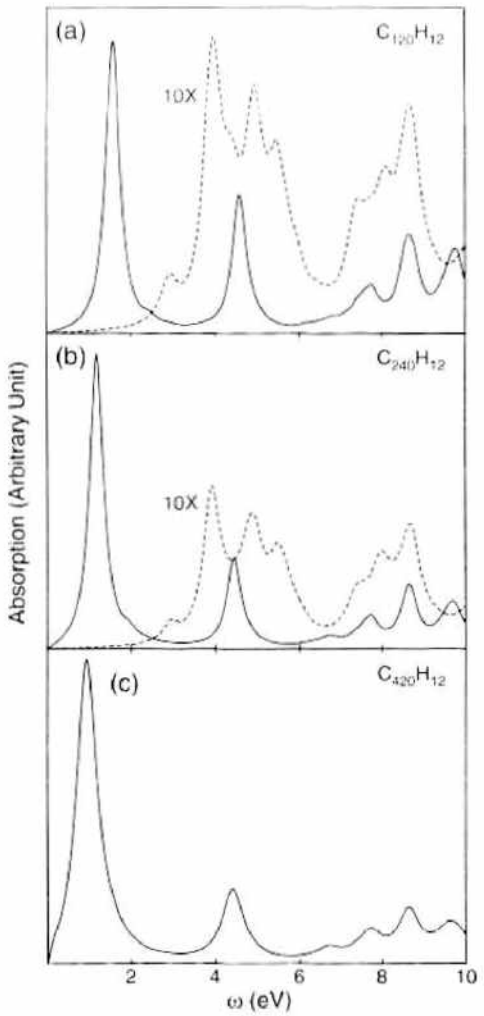


Figure 21. Calculated absorption spectra of (3, 3) SWNTs with dephasing parameter $\gamma = 0.2$ eV (a, b) and $\gamma = 0.3$ eV (c). The solids are the results for $\mathbf{E} \parallel \mathbf{T}$ and the dashed lines for $\mathbf{E} \perp \mathbf{T}$. The dashed lines are magnified 10 times. Reprinted with permission from [66], W. Z. Liang et al., *J. Am. Chem. Soc.* 123, 9830 (2001). © 2001, American Chemical Society.

are much weaker for $\mathbf{E} \perp \mathbf{T}$ as compared to for $\mathbf{E} \parallel \mathbf{T}$. The anisotropy increases as the tube length increases. This is consistent with the experimental observation [81] that the 4-Å SWNTs are opaque for $\mathbf{E} \parallel \mathbf{T}$ but almost transparent for $\mathbf{E} \perp \mathbf{T}$. The lowest-absorption peak red-shifts when the light polarization varies from the parallel direction of the tube to the perpendicular direction [81]. Our calculations show the same phenomenon for (4, 2) and (5, 0) but not for (3, 3). For (4, 2) $C_{220}H_{12}$ and (5, 0) $C_{300}H_{10}$, their respective optical gaps are 1.33 and 0.70 eV for $\mathbf{E} \perp \mathbf{T}$, and 1.64 and 1.16 eV for $\mathbf{E} \parallel \mathbf{T}$. For (3, 3) CNTs, the optical gap is larger for $\mathbf{E} \perp \mathbf{T}$ than for $\mathbf{E} \parallel \mathbf{T}$.

The optical gaps of the open-ended tubes (4, 2), (3, 3), and (5, 0) versus $1/N$ for $\mathbf{E} \parallel \mathbf{T}$ are plotted in Fig. 23a, 23b, and 23c, respectively. The dashed lines are the linear fits. Clearly the gaps depend linearly on $1/N$. The optical gaps of (4, 2), (3, 3), and (5, 0) are respectively 1.5, 0.61, and 0.90 eV as the tubes become infinitely long. In Ref. [65], similar findings have been reported; that is, it was found that the long (6, 0), (9, 0), (5, 5), (8, 8), and (10, 10) SWNTs have finite optical gaps.

The concept of the pyramidalization angle is developed to extend the definition of the π orbital to a nonplanar molecule [100]. To examine the effect of rehybridization of the σ and π electrons on the optical gap, the pyramidalization angle [100] of those carbon nanotubes has been calculated using the π orbital axis vector 1 (POAV1) analysis [100, 101]. The π orbital axis makes equal angles with all three σ bonds at the conjugated carbon atom in question. The angle between π orbital axis vector and the σ bond is 90 degrees for

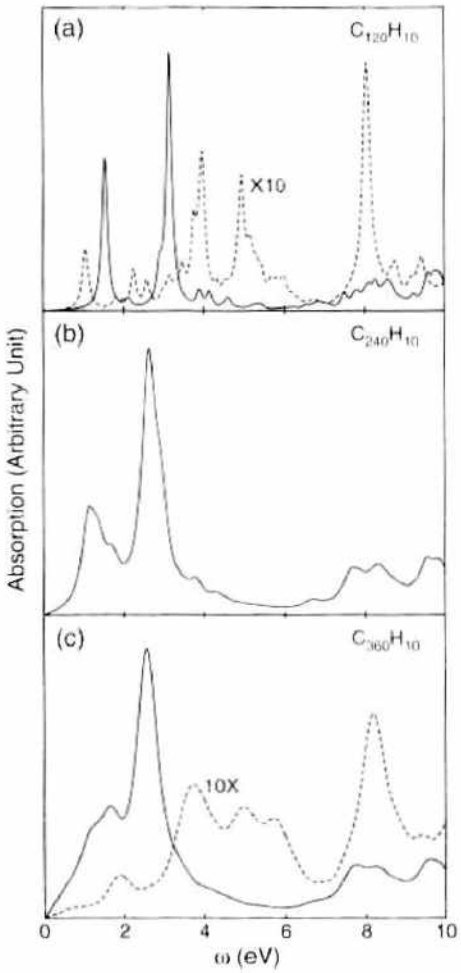


Figure 22. Calculated absorption spectra of zigzag (5, 0) SWNTs with dephasing parameter $\gamma = 0.08$ eV (a) and $\gamma = 0.2$ eV (b and c). The solids are the results for $\mathbf{E} \parallel \mathbf{T}$ and the dashed lines for $\mathbf{E} \perp \mathbf{T}$. The dashed lines are magnified 10 times. Reprinted with permission from [66], W. Z. Liang et al., *J. Am. Chem. Soc.* 123, 9830 (2001). © 2001, American Chemical Society.

planar molecules. The pyramidalization angle θ_P is then defined as the deviation of this angle from 90 degrees. The larger θ_P is, the larger the rehybridization of the σ and π electrons [100]. For (3, 3) and (5, 0), $\theta_P = 9.9$ and 10.0 degrees are obtained, respectively. For (4, 2) $C_{220}H_{12}$, θ_Ps are $9.6 \sim 9.7$ degrees, depending on the atoms. Because the θ_P values have no clear correlation with the optical gaps, it is conjectured that the differences in the optical spectra of (4, 2), (3, 3), and (5, 0) are caused mainly by the relative positions of the carbon atoms.

Armchair and zigzag SWNTs are highly symmetric. Depending on the way we terminate the open CNTs, (3, 3) and (5, 0) CNTs may have D_{3h} or D_{3d} and D_{5h} or D_{5d} symmetry, respectively [99]. When the lengths of the (3, 3) and (5, 0) SWNTs are large enough (>25 Å), the difference in optical response caused by different symmetries can be neglected [93, 99]. Chosen here to be examined are (3, 3) SWNTs with D_{3d} symmetry and (5, 0) with D_{5h}. For the dipole-allowed optical transition, the matrix element $\langle \Phi_i | \mathbf{r} \cdot \mathbf{E} | \Phi_f \rangle$ is nonzero, and $|\Phi_i\rangle$ ($|\Phi_f\rangle$) is the initial (final) electronic state. The variable D_{3d} has six irreducible representations: A_{1g}, A_{2g}, E_g, A_{1u}, A_{2u}, and E_u, and D_{5h} has eight irreducible representations: A_1', A_2', E_1', E_2', A_1'', A_2'', E_1'', and E_2'' [102]. For the electric field $\mathbf{E} \parallel \mathbf{T}$, the following transitions are allowed between the pairs of the molecular orbitals: for (3, 3) SWNTs, $A_{1g} \leftrightarrow A_{2u}$, $A_{2g} \leftrightarrow A_{1u}$, and $E_u \leftrightarrow E_g$; and for (5, 0) SWNTs, $E_1' \leftrightarrow E_1''$, $E_2' \leftrightarrow E_2''$, and $A_1' \leftrightarrow A_2''$. For the electric field $\mathbf{E} \perp \mathbf{T}$, the allowed transitions are, for (3, 3), $A_{1g} \leftrightarrow E_u$, $A_{2g} \leftrightarrow E_u$, $A_{1u} \leftrightarrow E_g$, $A_{2u} \leftrightarrow E_g$, and $E_u \leftrightarrow E_g$; and for (5, 0), $A_1' \leftrightarrow E_1'$, $A_1' \leftrightarrow E_2'$, $A_2' \leftrightarrow E_1'$, $A_2' \leftrightarrow E_2'$, $A_2'' \leftrightarrow E_1''$, $A_2'' \leftrightarrow E_2''$, $A_1'' \leftrightarrow E_1''$, and $A_1'' \leftrightarrow E_2''$. For the open-ended (3, 3) SWNTs, the molecular orbitals with A_{1g}, A_{2g}, A_{1u}, and A_{2u} symmetry are found to be energetically

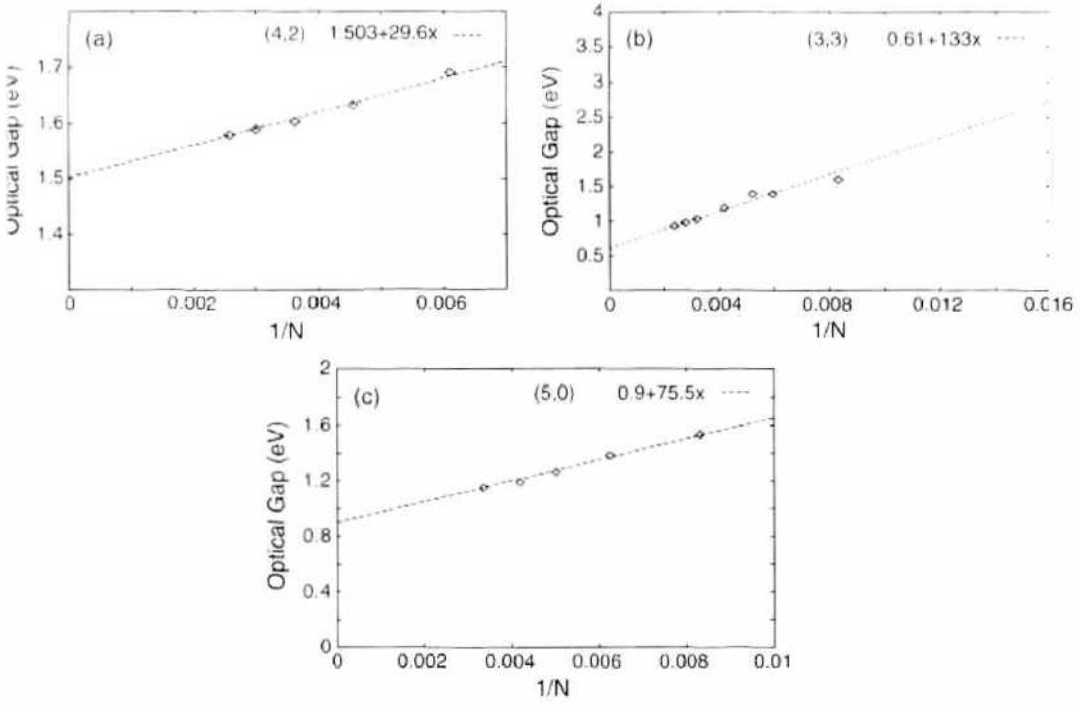


Figure 23. The optical gap versus $1/N$ for open-ended (4, 2), (3, 3), and (5, 0) SWNTs for $\mathbf{E} \parallel \mathbf{T}$, where N is the number of carbon atoms. The dashed lines are linear fits to the calculated results. The optical gaps are 1.50, 0.6 , and 0.90 eV for infinite (4, 2), (3, 3), and (5, 0) SWNTs, respectively. Reprinted with permission from [66]. W. Z. Liang et al., *J. Am. Chem. Soc.* 123, 9830 (2001). © 2001, American Chemical Society.

close to the HOMO and LUMO. The molecular orbitals with E_u and E_g symmetry are quite different in energy from HOMO and LUMO. The excitation at 1.57 eV of (3, 3) $C_{120}H_{12}$ for $\mathbf{E} \parallel \mathbf{T}$ consists mainly of the transitions $A_{1g} \leftrightarrow A_{2u}$ and $A_{2g} \leftrightarrow A_{1u}$, whereas, the first excitation for $\mathbf{E} \perp \mathbf{T}$ consists mainly of $A_{1g} \leftrightarrow E_u$, $A_{1u} \leftrightarrow E_g$, $A_{2g} \leftrightarrow E_u$, and $A_{2u} \leftrightarrow E_g$ transitions. This explains the fact that the optical gap for $\mathbf{E} \perp \mathbf{T}$ is larger than for $\mathbf{E} \parallel \mathbf{T}$ for (3, 3) SWNTs.

After obtaining the induced density matrices $\delta\rho$ of the transitions corresponding to the absorption peaks at 1.57 and 4.58 eV of (3, 3) $C_{120}H_{12}$, 1.53 and 3.14 eV of (5, 0) $C_{120}H_{10}$, and 0.61 and 2.49 eV of (4, 2) $C_{52}H_{12}$, the matrices are projected onto the HFMO representation [103]. In other words, the dipole-induced excitations are decomposed into the transitions between the pairs of molecular orbitals. The results are shown in Tables 1–3. The molecular orbital transitions whose absolute amplitudes are larger than 0.1 are listed. Because $|\delta\rho_{ij}| = |\delta\rho_{ji}|$, only $\delta\rho_{ij}$ $(i < j)$ are shown, where i and j stand for a pair of molecular orbitals. For (3, 3) $C_{120}H_{12}$, the excitation at 1.57 eV is a $\pi^* - \pi$ transition and consists mainly of the HOMO to LUMO + 1 and HOMO − 1 to LUMO transitions, with their respective amplitudes of 0.44 and 0.36 (see Table 1). Other contributions, such as the HOMO − 2 to LUMO + 3 and HOMO − 3 to LUMO + 2, are much smaller. The excitation at 4.58 eV is also a $\pi - \pi^*$ transition. Its main contributions are from the pairs of the molecular orbitals with E_u and E_g symmetry (see Table 1). The transition from HOMO to LUMO is forbidden for (3, 3) CNTs, as they have A_{1g} and A_{2g} symmetry, respectively. For (5, 0) $C_{120}H_{10}$, the excitations at 1.53 and 3.14 eV are from $\pi - \pi^*$ transitions. The molecular orbitals involved in the transitions have E_1', E_1'', E_2', and E_2'' symmetry (see Table 2). The $A_2'' \leftrightarrow A_1'$ transition has made a small contribution to the excitation at 3.14 eV. The HOMO to LUMO transition is forbidden for (5, 0) as well. For (4, 2) $C_{52}H_{12}$, the excitation at 0.61 eV is a $\pi - \pi^*$ transition, and its main contributions are from the HOMO to LUMO, HOMO − 1 to LUMO + 1, and HOMO − 2 to LUMO + 4 transitions. The excitation at 2.49 eV is also a π to π^* transitions, with the HOMO − 1 to LUMO + 1 transition as the main contributor (see Table 3).

The absorption spectra of the 4-Å SWNTs reported in Ref. [32] were measured recently [81]. The polarization of the light was tuned to examine the anisotropy of the optical

Table 1. The compositions of the photo-induced transitions at 1.57 and 4.58 eV of (3, 3) $C_{120}H_{12}$.

1.57 eV	
HOMO − 1(A_{1u}) → LUMO + 0(A_{2g})	0.36
HOMO − 0(A_{1g}) → LUMO + 1(A_{2u})	0.44
HOMO − 3(A_{2g}) → LUMO + 2(A_{1u})	−0.15
HOMO − 2(A_{2u}) → LUMO + 3(A_{1g})	−0.19
4.58 eV	
HOMO − 18(E_u) → LUMO + 9(E_g)	−0.11
HOMO − 10(E_u) → LUMO + 9(E_g)	0.16
HOMO − 9(E_u) → LUMO + 9(E_g)	0.12
HOMO − 17(E_u) → LUMO + 10(E_g)	−0.11
HOMO − 10(E_u) → LUMO + 10(E_g)	0.14
HOMO − 9(E_u) → LUMO + 10(E_g)	0.16
HOMO − 11(E_g) → LUMO + 11(E_u)	−0.23
HOMO − 12(E_g) → LUMO + 12(E_u)	0.23
HOMO − 13(E_g) → LUMO + 13(E_u)	−0.13
HOMO − 14(E_g) → LUMO + 14(E_u)	0.13
HOMO − 15(E_u) → LUMO + 15(E_g)	−0.13
HOMO − 16(E_u) → LUMO + 16(E_g)	0.13

The first column lists the transitions between pairs of molecular orbitals, and the second column lists the corresponding composition coefficients of the specific transitions between the molecular orbital pairs. Reprinted with permission from [66], W. Z. Liang, et al., *J. Am. Chem. Soc.* 123, 9830 (2001). © 2001, American Chemical Society.

Table 2. The compositions of the photo-induced transitions at 1.53 and 3.14 eV of (5, 0) $C_{120}H_{10}$.

1.53 eV	
HOMO − 10(E_1'') → LUMO + 0(E_1')	0.12
HOMO − 11(E_1'') → LUMO + 1(E_1')	0.12
HOMO − 4(E_2') → LUMO + 3(E_2'')	0.19
HOMO − 0(E_2') → LUMO + 3(E_2'')	−0.28
HOMO − 5(E_2') → LUMO + 4(E_2'')	0.19
HOMO − 1(E_2') → LUMO + 4(E_2'')	0.28
HOMO − 6(E_2'') → LUMO + 8(E_2')	0.10
HOMO − 3(E_2'') → LUMO + 8(E_2')	0.13
HOMO − 2(E_2'') → LUMO + 8(E_2')	−0.20
HOMO − 7(E_2'') → LUMO + 9(E_2')	0.10
HOMO − 3(E_2'') → LUMO + 9(E_2')	−0.20
HOMO − 2(E_2'') → LUMO + 9(E_2')	−0.13
HOMO − 1(E_2') → LUMO + 13(E_2'')	0.11
HOMO − 0(E_2') → LUMO + 13(E_2'')	0.11
HOMO − 1(E_2') → LUMO + 14(E_2'')	−0.11
HOMO − 0(E_2') → LUMO + 14(E_2'')	0.11
3.14 eV	
HOMO − 10(E_1'') → LUMO + 0(E_1')	−0.25
HOMO − 11(E_1'') → LUMO + 1(E_1')	−0.25
HOMO − 24(A_2'') → LUMO + 2(A_1')	−0.11
HOMO − 4(E_2') → LUMO + 3(E_2'')	0.19
HOMO − 0(E_2') → LUMO + 3(E_2'')	0.12
HOMO − 5(E_2') → LUMO + 4(E_2'')	0.19
HOMO − 1(E_2') → LUMO + 4(E_2'')	−0.12
HOMO − 14(E_1') → LUMO + 6(E_1'')	0.10
HOMO − 15(E_1') → LUMO + 7(E_1'')	0.10
HOMO − 6(E_2'') → LUMO + 8(E_2')	0.11
HOMO − 7(E_2'') → LUMO + 8(E_2')	0.11

The first column lists the transitions between pairs of molecular orbitals, and the second column lists the corresponding composition coefficients of the specific transitions between the molecular orbital pairs. Reprinted with permission from [66], W. Z. Liang, et al., *J. Am. Chem. Soc.* 123, 9830 (2001). © 2001, American Chemical Society.

Table 3. The compositions of the photo-induced transitions at 0.61 and 2.49 eV of (4, 2) $C_{52}H_{12}$.

1.61 eV	
HOMO − 0 → LOMO + 0	−0.72
HOMO − 1 → LOMO + 1	−0.34
HOMO − 2 → LOMO + 4	−0.25
HOMO − 4 → LOMO + 6	0.12
2.49 eV	
HOMO − 0 → LOMO + 0	−0.15
HOMO − 1 → LOMO + 1	0.53
HOMO − 3 → LOMO + 2	0.19
HOMO − 5 → LOMO + 3	−0.18
HOMO − 2 → LOMO + 4	0.19
HOMO − 4 → LOMO + 6	−0.15
HOMO − 2 → LOMO + 6	0.14

The first column lists the transitions between pairs of molecular orbitals, and the second column lists the corresponding composition coefficients of the specific transitions between the molecular orbital pairs. Reprinted with permission from [66], W. Z. Liang, et al., *J. Am. Chem. Soc.* 123, 9830 (2001). © 2001, American Chemical Society.

response. Several important observations were obtained. First, for $\mathbf{E} \parallel \mathbf{T}$, three absorption peaks were identified at 1.35, 2.15, and 3.10 eV. Second, As **E** deviates from the tube's parallel direction to its normal direction, the absorption intensity weakens significantly, and moreover, the spectra red-shift. Third, the 4-Å SWNTs are transparent for $\mathbf{E} \perp \mathbf{T}$. The absorption spectrum of the (3, 3) SWNT has only one major peak below 4.0 eV. When the electric field is parallel to the tube, there are three distinctive absorption peaks below 4 eV for (5, 0) SWNTs and two peaks below 4.0 eV for (4, 2). Both of these numbers are consistent with the experimental measurement. Because the parameters in the PM3 Hamiltonian were not optimized for the TDHF method, which is employed in the LDM calculation, the calculated absorption spectra red-shift with respect to the experimental spectra. Nevertheless, the calculated and measured absorption spectra agree qualitatively. It is clear that the (3, 3) SWNT alone can not account for the measured absorption spectrum, and it may explain only the first peak at 1.37 eV in the measured spectrum. The calculated absorption spectra of (4, 2) and (5, 0) are consistent with the measured spectra. Because it is energetically unfavorable to fit half C_{20} caps on (4, 2) SWNTs [32], only the *H*-terminated (4, 2) SWNTs may exist. The sharp peak at 1.37 eV in the measured absorption spectrum agrees very well with the first peak in the absorption spectrum of (4, 2). This seems to indicate that (4, 2) SWNT is the most likely candidate. However, the porous AFI crystals were thermally treated at 500°–800°C during the synthesis [32], and it is not clear that *H* atoms inside the pores survive the thermal treatment. This may be resolved by examining the energetics of *H*-terminated (4, 2) or the binding energies of the terminal *H* atoms, and additional calculations are required. If the 4-Å SWNTs synthesized in the porous AFI crystals are of a single chirality, they should be either the *H*-terminated (4, 2) or (5, 0) SWNTs. Of course, a mixture of three SWNTs (4, 2), (3, 3), and (5, 0) cannot be ruled out.

The chirality has a strong influence on the absorption spectra of 4-Å SWNTs. This differs from previous findings [65] of larger-diameter SWNTs, where the chirality has a much abated effect on the absorption spectra. Because of their sensitive dependence on the chirality, the absorption spectra are used to determine the structure of 4-Å SWNT. The optical responses of the 4-Å SWNTs are highly anisotropic. This is because the transition dipole moments parallel to the tube are much larger than the moments perpendicular to the tube. The anisotropy is clearly demonstrated in Figs. 20, 21, and 22. The optical responses of the 4-Å SWNTs are highly anisotropic. This is because the transition dipole moments parallel to the tube are much larger than the moments perpendicular to the tube. The anisotropy is clearly demonstrated in Figs. 20, 21, and 22. Similar to the findings from the larger-diameter SWNTs, the optical gaps of the infinite (3, 3), (4, 2), and (5, 0) SWNTs are all finite, being 0.6, 1.5, and 0.9 eV, respectively. This is despite the fact that the LDA calculation

predicts that all three 4-Å SWNTs are metallic [98]. The existence of the finite optical gap is consistent with the measured absorption spectra [81].

3.4. Effect of Functionalization on Optical Spectra of Metallic Single-Walled Carbon Nanotubes

Modern synthetic chemistry and biological processes primarily take place in the solution phase. However, SWNTs typically exist in forms of raw ropes or bundles about 10–25 nm in diameter and a few micrometers in length. Because SWNT bundles are entangled together in the solid state to form a highly dense, complex network structure, it is difficult to develop a dissolution process for the SWNTs. It is possible to wet the SWNT raw soot in refluxing nitric acid, whereby the end caps of the tubes are oxidized to carboxylic acid and other weakly acidic functionalities [104–107]. In certain instances, covalent chemistry on the side walls of the SWNTs is a viable route to soluble materials [108]. In fact, the ability to carry out controlled covalent chemistry on the side walls of the SWNTs is a very important step, and the achievement of systematic and predictable side-wall chemistry is likely to be a precursor to many of the applications that are currently envisioned for CNTs.

It was previously known that ionic doping with electron acceptors introduces holes into the valence band that can lead to the removal of the interband transitions in semiconducting SWNTs [90]. Recently, it was shown that covalent chemistry can covert the metallic SWNTs to semiconductors, which may lead to efficient nanotube separation [38, 39]. Kamaras et al. showed that reactions with dichlorocarbene rapidly opens a gap near the Fermi level of the metallic CNTs. Strano et al. demonstrated that diazonium agents allow functionalization of SWNTs suspended in aqueous solution with high selectivity and display an autocatalytic effect that functionalizes the entire tube. In particular, metallic species are selected to react to the near exclusion of semiconducting SWNTs under controlled conditions [39]. The optical absorption spectrum is used to monitor the valence-to-conduction electronic transitions.

Here the LDM-PM3 method is employed to study the effect of functionalization on optical absorption spectra of metallic SWNTs. The hydrogenated armchair SWNT (5, 5) ($C_{140}H_{20}$ and $C_{140}H_{20}-C_6H_4Cl$) are selected for absorption spectra calculations. When one functional group is attached to the tube side wall, a double bond is broken, and the functionalized SWNT can receive electrons from neighboring nanotubes or can react with fluoride and diazonium salts. In our model, one ($-C_6H_4Cl$) is attached to the (5, 5) SWNT, and one hydrogen atom is put in to neutralize the free radical. The amount of functionality addition is 7.1 mol per 1000 mol carbon when one ($-C_6H_4Cl$) functionality is attached to the (5, 5) SWNT. An arene-functionalized (5, 5) SWNT is shown in Fig. 24a with $-C_6H_4Cl$ attached to one side of the tube and a hydrogen atom to the other side. Optical absorption spectra are calculated for (5, 5) SWNTs with and without functionalities. In Fig. 24, the effect of functionalization is illustrated by comparing the spectrum of the stand-alone

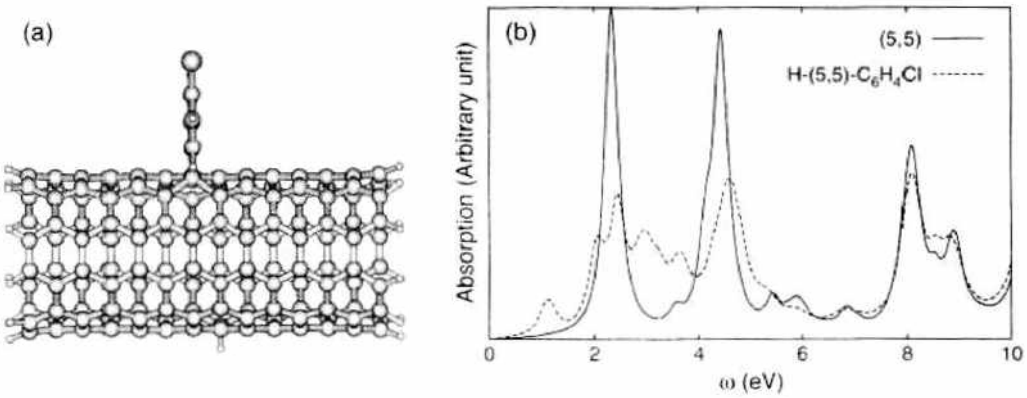


Figure 24. (a) The structure of the functionalized (5, 5) SWNT: $C_{120}H_{20}-C_6H_4Cl-H$. (b) Absorption spectra of the (5, 5) metallic SWNT and the functionalized (5, 5) SWNT.

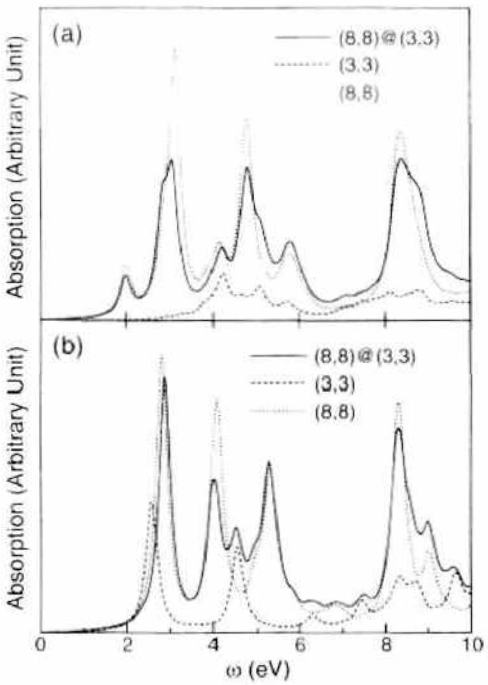


Figure 25. Absorption spectrum of a (3, 3)@(8, 8) DWNT (solid) is shown together with those of a (3, 3) SWNT (dashed) and a (8, 8) SWNT (dotted). The external field is (a) perpendicular or (b) parallel to the nanotube axis. Both tubes in the DWNT are 15.4 Å.

(5, 5) SWNT with that of the arene-functionalized (5, 5) SWNT. The low-energy metallic absorption peaks of the (5, 5) SWNT (below 5 eV) are greatly reduced in intensities on functionalization, in full agreement with measurements [39]. The effect of functionalization on the high-energy features near 8 eV, however, is much less pronounced.

The effect of functionalization on optical absorption spectra of semiconducting SWNTs has also been examined in this study. Similar changes of calculated absorption spectra on

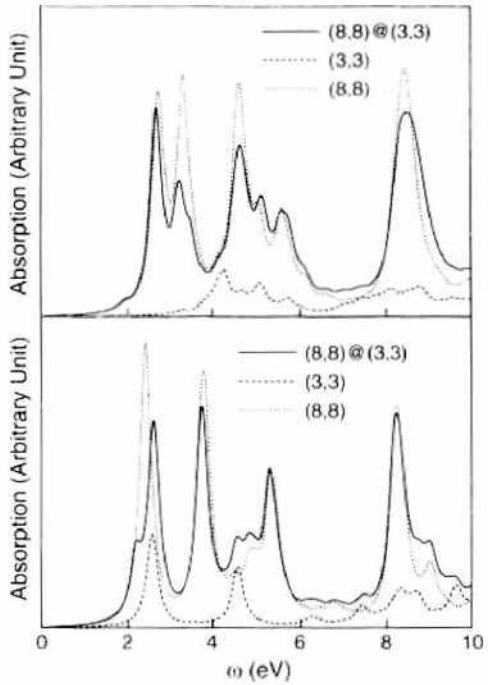


Figure 26. Absorption spectrum of a (3, 3)/(8, 8) DWNT (solid) is shown with those of a (3, 3) SWNT (dashed) and a (8, 8) SWNT (dotted). The DWNT is composed of nanotubes of unequal lengths. The inner tube of the DWNT is 15.4 Å, and the outer tube is 20.4 Å. The external field is (a) perpendicular or (b) parallel to the nanotube axis.

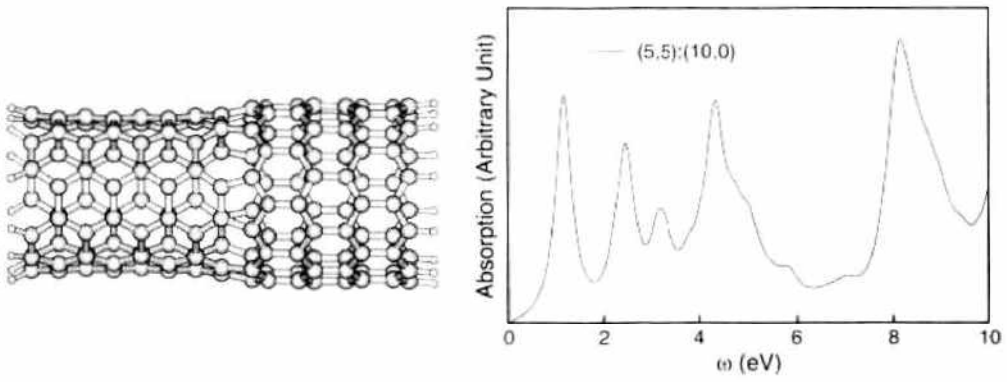


Figure 27. Absorption spectrum for a coaxial CNT junction between a (5, 5) SWNT and a (10, 0) SWNT. The external light field is along the nanotube axis. The lowest absorption peak is attributed to the (10, 0) part of the junction via a density-matrix analysis, whereas the second (~2.4 eV) and fourth (~4.3 eV) peaks have contributions from both the (5, 5) and (10, 0) components of the junction.

functionalization are found. The fact that experimentally semiconducting peaks are not altered on functionalization in the work of Strano et al. [39] confirms the high selectivity of functionalization in the presence of diazonium agents.

3.5. Absorption Spectra of Multiwalled Nanotubes and Carbon Nanotube Junctions

Multiwalled nanotubes (MWNTs) are representative of a unique class of self-organized solids with structural hierarchy. Compared with SWNTs, modifications in the electronic properties are introduced by the internanotube interactions in MWNTs [109–111]. For instance, those interactions are known to cause a small band repulsion and a pseudogap near E_F for metallic CNTs [112, 113]. Thin DWNTs have been synthesized by fusion of fullerenes encapsulated in SWNTs [41, 114] via electron beam irradiation on the carbon peapods, which induces coalescence of fullerenes. Optical absorption spectra of DWNTs will be studied here and compared with those of individual SWNTs that compose the corresponding DWNTs. Also of great interest are optical properties of the nanotube junctions [115] and CNTs with structural defects, which will be examined here with the help of the LDM-PM3 algorithm.

We start out by examining the optical absorption spectra of DWNTs, calculated by the semiempirical LDM-PM3 method. In Fig. 25, the absorption spectrum of a (3, 3)@(8, 8) DWNT is shown together with those of a (3, 3) SWNT and an (8, 8) SWNT of the same lengths as the DWNT. Both nanotubes in the DWNT are 15.4 Å. In the upper panel, Fig. 25a, the lineshapes are obtained after applying an external field perpendicular to the nanotube

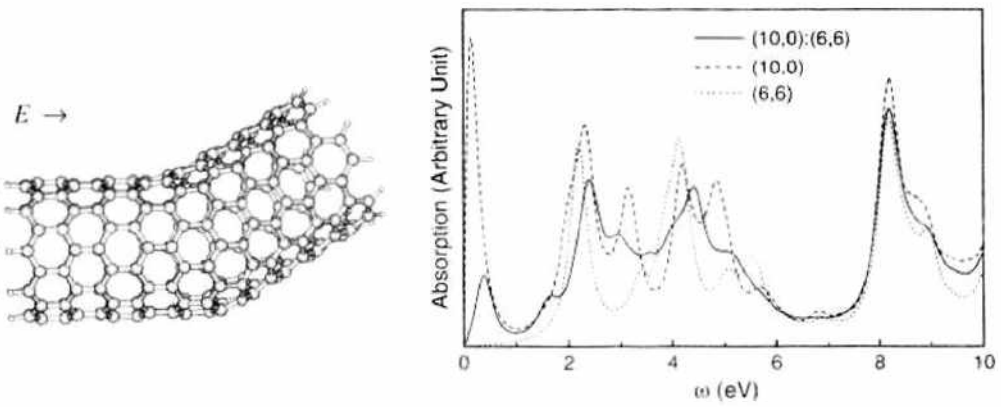


Figure 28. Absorption spectrum for a CNT junction between a (6, 6) SWNT and a (10, 0) SWNT. The (6, 6) and (10, 0) segments are oriented to form an angle of about $\pi/6$. The external light field is along the direction of the (6, 6) SWNT. Absorption spectra of a (6, 6) SWNT and a (10, 0) SWNT are also shown for comparison.

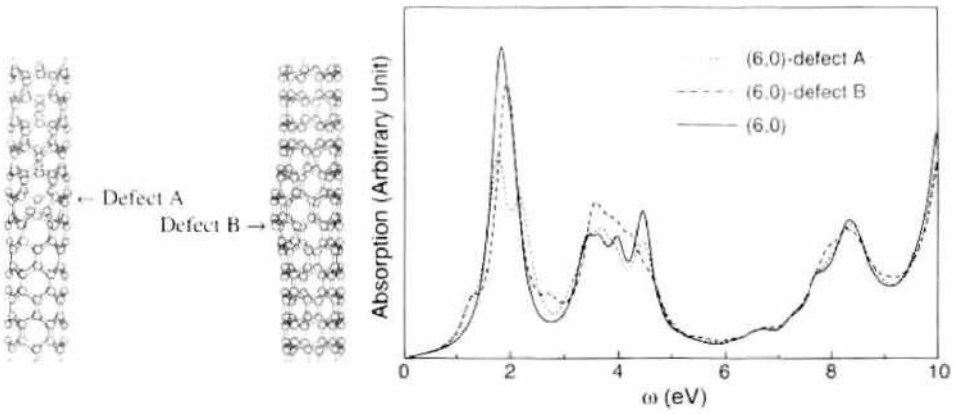


Figure 29. Absorption spectrum for a (6, 0) SWNT is compared with those of two (6, 0) SWNTs with defects.

axis. The contribution from the inner (3, 3) tube seems to be small, and the spectra of the DWNT resemble that of the outer (8, 8) for $\mathbf{E} \perp \mathbf{T}$. In the lower panel, Fig. 25b, the external field is parallel to the nanotube axis ($\mathbf{E} \parallel \mathbf{T}$), and the contribution from the inner (3, 3) tube is much more significant. The first absorption peak of the DWNT has a slightly higher energy than both isolated (8, 8) and (3, 3) SWNTs. High-energy peaks of the SWNT located between 4 and 5.5 eV are a combination of peaks from the isolated SWNTs.

In comparison, absorption spectra of (3, 3) and (8, 8) DWNT tubes of unequal lengths are shown in Fig. 26. The inner tube of the DWNT is 15.4 Å, and the outer tube is 20.4 Å. For $\mathbf{E} \perp \mathbf{T}$, the absorption spectrum of the DWNT in Fig. 26a, although differing from that in Fig. 25a, remains sufficiently close to that of an isolated (8,8) SWNT with a length of 20.4 Å. The contribution from the inner tube for $\mathbf{E} \perp \mathbf{T}$ is minimal. For $\mathbf{E} \parallel \mathbf{T}$, the absorption lineshape of the DWNT is practically a linear combination of those of isolated (3, 3) and (8, 8) SWNTs with similar lengths, as shown in Fig. 26b. Compared with the spectrum in Fig. 25b, the major peak positions in Fig. 26b are located at similar energies, and the change in nanotube lengths merely alters the relative strengths of the peaks.

Shown in Fig. 27 is the optical absorption spectrum of a coaxial CNT junction between a (5, 5) SWNT and a (10, 0) SWNT. The external light field is along their common axis. The lowest absorption peak is attributed to the (10, 0) part of the junction with the help of a density-matrix analysis, whereas the second (~2.4 eV) and the forth (~4.3 eV) peaks receive contributions from both the (5, 5) and (10, 0) components of the junction. The weakest low-energy excitation; that is, the third peak at about 3.2 eV, seems mostly caused by the (10, 0) segment.

A CNT junction can also be constructed in a noncoaxial form, as demonstrated in Fig. 28. The (6, 6) and (10, 0) segments of the junction are oriented to form an angle of about $\pi/6$ between them. The external field is along the direction of the (6, 6) SWNT. Two prominent absorption features at about 2.4 eV and 4.3 eV are found in addition to a structure near 8 eV. Overall, the absorption spectrum of the noncoaxial junction is approximately an average over those of the (6, 6) and (10, 0) segments, as shown in Fig. 28.

Last, we examine the effect of wall defects on the optical properties of SWNTs. A combination of two pairs of the so-called "5+7 member rings" (one pentagon and one heptagon) is the simplest point defect that keeps the carbon nanotube in a straight line with C_{2v} symmetry. In Fig. 29, we modeled two different types of point defects by introducing two pairs of "5 + 7 member-rings" on a (6, 0) SWNT. (In fact, the noncoaxial CNT junction in Fig. 28 is also constructed using two pairs of "5 + 7 member rings.") For example, if one C–C bond at the central part of the SWNT is rotated by $\pi/2$ to arrive at two pairs of "5 + 7 member-rings," the type of point defects formed this way is called Defect A, and the original 66 carbon hexagons are now replaced by 62 hexagons and two pairs of "5 + 7 member-rings." The absorption spectrum for a perfect (6, 0) SWNT is compared with those of two (6, 0) SWNTs with wall defects in Fig. 29. It is found that such localized wall defects have no significant effect on the absorption spectra of the SWNTs.

ST-PRINCIPLES DENSITY-FUNCTION EORY CALCULATIONS

nciples calculations have been carried out on SWNTs [116, 117], including those with llest diameter (4 Å) [118]. Moreover, the detailed plane-wave *ab initio* pseudopo- ɔcal density approximation (LDA) calculations [98] predicted that all small-diameter e conductors regardless their chiralities. Recently, *ab initio* calculations have been ed on the anisotropic dielectric response of small-diameter SWNTs in the framework dependent density-functional theory (DFT) [119] with results in good agreement eriment.

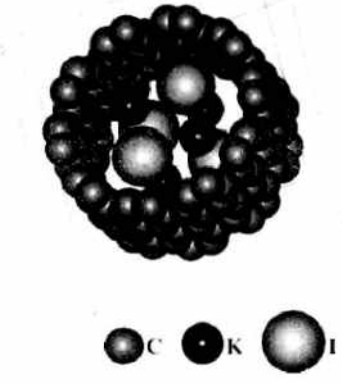
ɔased configurations and electronic structures were optimized for finite open (4, 4) 5) SWNTs [120], and results show that charge distributions and bond lengths/angles ɛnds and those on tube side walls differ significantly. As DFT studies have revealed, ization of σ and π orbitals and interactions among π orbitals caused by the e of the tube lead to significant modifications of the electronic band structures [98]. ıs plausible that orbital rehybridization and electronic interactions that are taken ount in the DFT approach may cause quantitative or even qualitative alternations implified tight-binding absorption spectra. To supplement our semiempirical LDM we have therefore employed a first-principles DFT approach [121, 122] to calculate cal response of isolated SWNTs with periodic boundary conditions. Calculations are ed with the WIEN97 software package [123] based on the full potential linearized ted plane waves (LAPW) method. In the LAPW method, the unit cell is divided ɔ types of regions: the atomic spheres centered at nuclear sites and the interstitial ɔetween the nonoverlapping spheres. Within the atomic spheres, the wave function is ɛd in terms of atomic wave functions, and in the interstitial region the wave function ıded in terms of plane waves. Optical properties of SWNTs are obtained by com- he joint density of states (JDOS) and transition dipole matrix elements. Results are ɛd with those from the tight-binding model and the LDM method in the next section. s been shown [114, 124, 125] that carbon nanotubes can encapsulate fullerenes. [124], this so-called carbon peapod is found to be a metal with multicarries distributed ɔng the nanotube and on the C_{60} chain. These examples show that through chemical the electronic properties of nanotubes can be modified. As mentioned earlier, potas- ɗide has been successfully inserted into a (10, 10) SWNT [41, 42]. Lattice distortions ɛrved that are attributed to the difference in K:I coordination from the bulk crystal interactions between the KI crystal and the tubule wall. Here we employ the first- es DFT approach to calculate the electronic structure of the KI intercalated (10, 10) nanotube. The resulting KI/SWNT composite is a highly anisotropic one-dimensional ɛ whose electronic and optical properties are expected to be considerably modified ɨpect to those of the bulk halide and the encapsulating nanotube.

ıe calculation we adopt the experimental structure of KI@(10, 10) reported in]. The structure is depicted in Fig. 30. The muffin-tin radii are set to 1.30, 2.61, and

16, 117], including those with ıne-wave *ab initio* pseudopo- dicted that all small-diameter *initio* calculations have been ter SWNTs in the framework h results in good agreement

timized for finite open (4, 4) ions and bond lengths/angles s DFT studies have revealed, g π orbitals caused by the ectronic band structures [98]. c interactions that are taken even qualitative alternations ıent our semiempirical LDM proach [121, 122] to calculate y conditions. Calculations are ı the full potential linearized thod, the unit cell is divided ɛlear sites and the interstitial spheres, the wave function is itial region the wave function WNTs are obtained by com- matrix elements. Results are M method in the next section. s can encapsulate fullerenes. ıl with multicarries distributed ɛs show that through chemical . As mentioned earlier, potas- ɨT [41, 42]. Lattice distortions dination from the bulk crystal ıll. Here we employ the first- of the KI intercalated (10, 10) y anisotropic one-dimensional d to be considerably modified nanotube.

of KI@(10, 10) reported in radii are set to 1.30, 2.61, and



). Unit cell of a (10, 10) carbon nanotube having potassium iodide inside.

ınside.

Table 4. Char

Region
C_{120}
K_4I_4
Interstitial

Table 4. Charges distribution of KI@(10, 10), (10, 10) SWNT, and KI.

	Charges		
Region	KI@(10, 10)	(10, 10)	KI
C_{120}	401.73	436.48	—
K_4I_4	281.53	—	282.53
Interstitial	324.74	283.52	5.47

3.50 atomic units for carbon. KI@(10, 10) in our calculati K and I are put into each na capillaries, K and I are spac the spacing is 3.69 Å to ma the nanotubes in the neighb to prevent intertubule intera calculations, we use one k p lation is considered to be ac consecutive iterations does n

Table 4 shows the charges nanotube, (10, 10) nanotube we use the same unit cell as cell contains four potassium amount of charge transfer fi the KI intercalation. On the from the atomic sphere regi electrons per K and I transfe change in the electronic stru

Electronic energy bands f 31b, respectively. The cell us which is three times as man with the unit cell of KI@(10 between (10, 10) and KI@(1 KI@(10, 10) SWNT show the the flat energy bands near −1 in Fig. 31b contain large co

3.50 atomic units for carbon, potassium, and iodine, respectively. The lattice paran KI@(10, 10) in our calculation are a = 18 Å, b = 18 Å, and c = 7.384 Å. Linear c K and I are put into each nanotube, and each unit cell contains $K_4I_4C_{120}$. Across the capillaries, K and I are spaced at intervals of 4.0 Å, whereas along the SWNT ca the spacing is 3.69 Å to maintain the periodicity. The center-to-center distance the nanotubes in the neighboring cells is set to 18 Å, which was found to be large to prevent intertubule interactions. To achieve self-consistency for the electronic s calculations, we use one k point in the irreducible part of the Brillouin zone, and th lation is considered to be achieved when the energy variation for the k point betw consecutive iterations does not exceed 10^{-5} Ry.

Table 4 shows the charges in the atomic spheres and interstitial regions of KI@ nanotube, (10, 10) nanotube, and linear chain KI. In the calculation of a linear c we use the same unit cell as the KI@(10, 10), with the carbon atoms removed. E cell contains four potassium and iodine atoms. The results show that there is a sul amount of charge transfer from the carbon atomic spheres into the interstitial re the KI intercalation. On the average there are 0.3 electrons per carbon atom tra from the atomic sphere region to the interstitial region, although there are only ab electrons per K and I transferred to the interstitial region. Hence, we expect some sig change in the electronic structure of the (10, 10) SWNT on intercalation.

Electronic energy bands for (10, 10) and KI@(10, 10) CNTs are given in Fig. 31b, respectively. The cell used in the calculation for (10, 10) contains 120 carbon which is three times as many as those of the (10, 10) unit cell. This is chosen t with the unit cell of KI@(10, 10), so that a direct comparison of the electronic st between (10, 10) and KI@(10, 10) is more clear (see Fig. 31). Both the (10, 10) SW KI@(10, 10) SWNT show the metallic character. It is found that I's 5p orbitals contr the flat energy bands near −1 eV in Fig. 31b, and the conduction bands between 1 a in Fig. 31b contain large contributions from K's 4s orbitals. We plot the DOS in

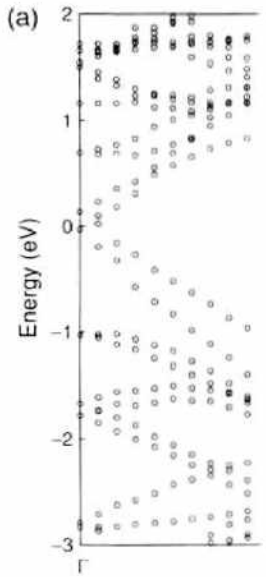


Figure 31. Calculated band structure from the Fermi level energy.

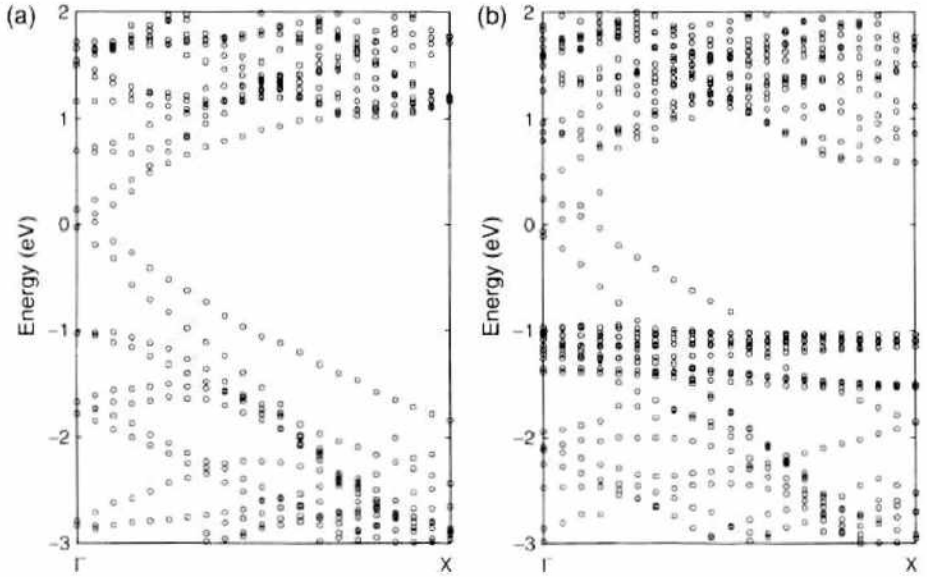


Figure 31. Calculated band structures of (a) (10, 10) carbon nanotube and (b) KI@(10, 10). Energies are from the Fermi level energy.

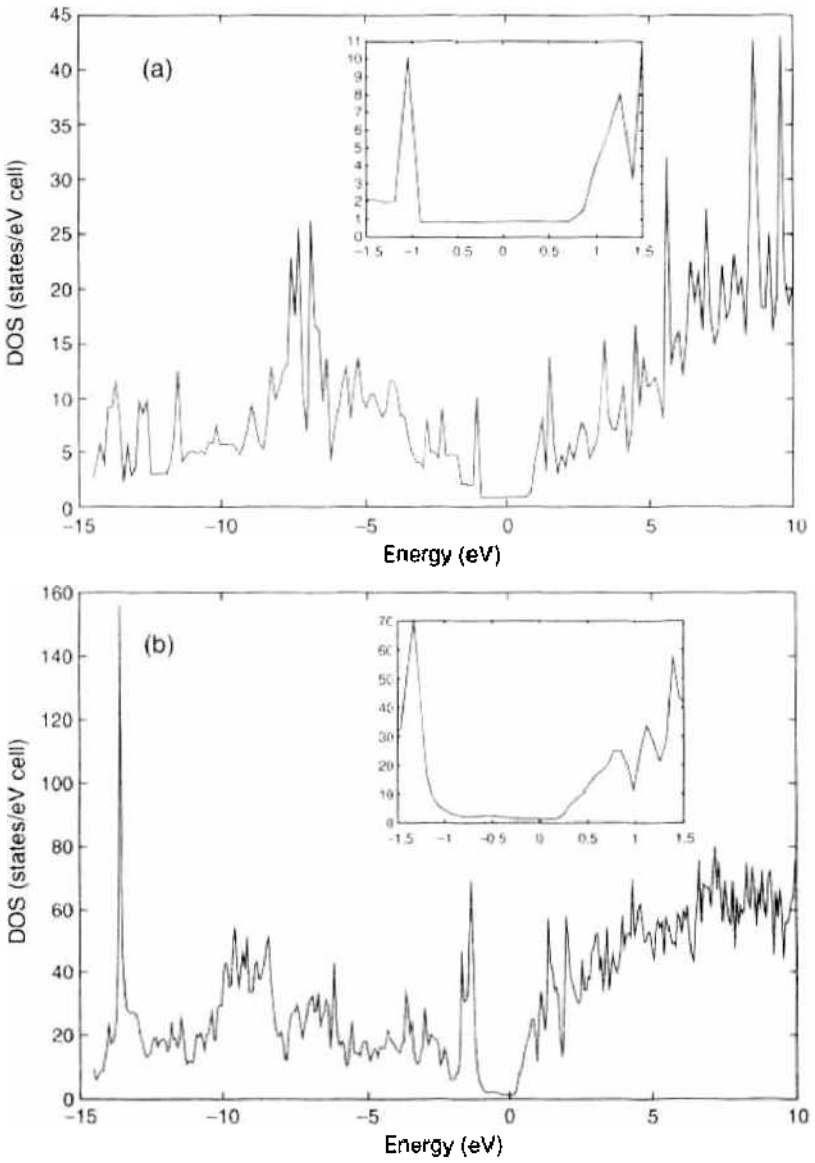


Figure 32. Density of states for (a) (10, 10) carbon nanotube and (b) KI@(10, 10). The Fermi level is at 0 eV.

The huge peak at −15 eV in Fig. 32b belongs to K's 3p orbitals, and the peak at −1 eV corresponds mainly to the I's 5p orbitals. Among other contributions, the broad band of peaks between 1 and 5 eV contains those from K's 4s orbital. The insets in Fig. 32a and 32b show the DOS around the Fermi level.

Figure 33a and 33b show the electron density contour plots for (10, 10) and KI@(10, 10) CNTs, respectively. Figure 33c and 33d show the positive and negative differences between the two, respectively, and positive means the decrease of electron density on the KI intercalation, whereas the negative means the increase of density. Clearly there is electron transfer from K's 4s to I's 5p orbitals. Figure 33c shows the decrease of electron density in the carbon atomic sphere regions of KI@(10, 10) CNTs with respect to those of (10, 10), whereas Fig. 33d shows the increase of electron density in the interstitial region. This shows clearly that electrons move from the atomic sphere region of carbon atoms to the interstitial region of the tube. Thus, the electrons on the tube become more delocalized on intercalation of potassium iodide.

Figure 34 shows the calculated absorption spectra of KI@(10, 10), (10, 10), and KI. The absorption spectrum of the KI@(10, 10) tube is not a simple summation of those of the (10, 10) tube and KI. It is observed that the peak at about 1.8 eV broadens considerably on intercalation. This is consistent with our calculation result that the DOS changes near the Fermi level on the KI intercalation, as indicated by the insets of Fig. 32.

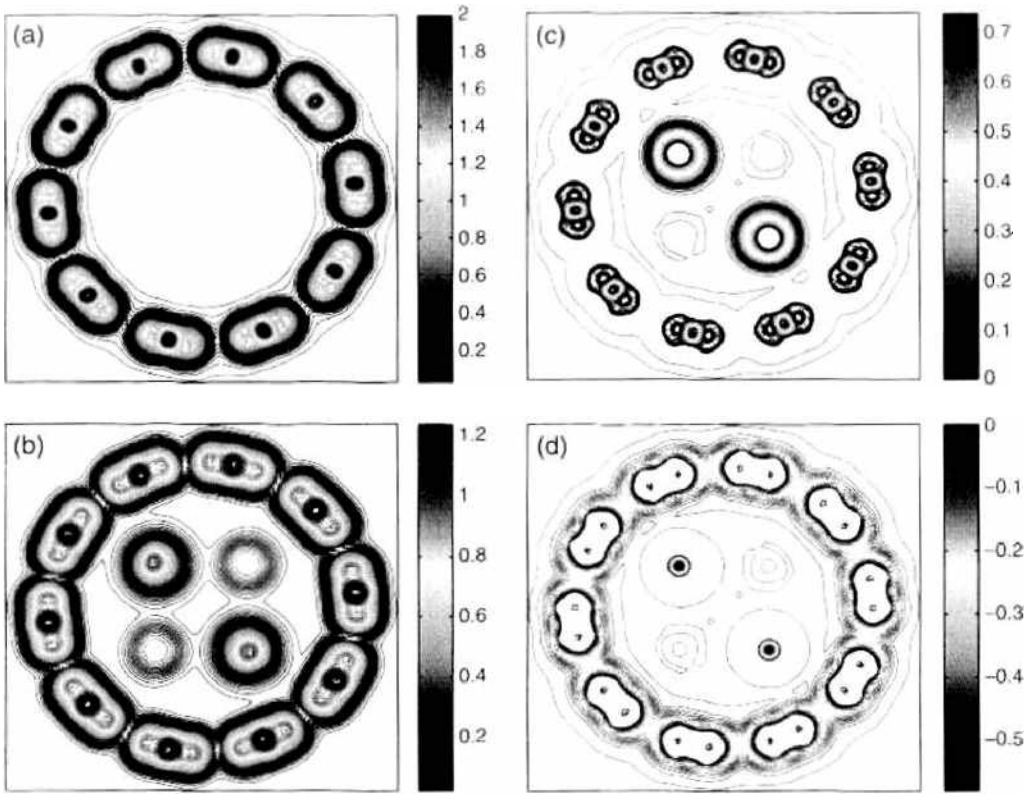


Figure 33. Contour plots of the electron density of (a) (10, 10) carbon nanotube and (b) KI@(10, 10). The contour plots of a more positively charged area and one more negatively charged than a simple sum of two charge densities of the nanotube and KI are shown in (c) and (d) respectively.

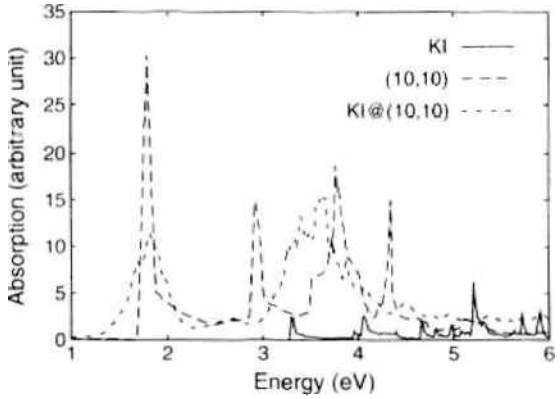


Figure 34. Absorption spectrum of KI@(10, 10), (10, 10) nanotube, and KI.

The electronic structure of potassium iodide intercalated carbon nanotubes has been investigated. On intercalation, the electrons in the SWNT become much more delocalized, indicating strong interactions between KI and the nanotube wall. The interactions alter the electronic structures and the DOS near the Fermi level. The calculated band structures and DOSs reflect these changes. The intercalations thus have the potential to drastically change the physical properties of CNTs.

5. COMPARISON OF THE TIGHT-BINDING, LOCALIZED-DENSITY-MATRIX AND DENSITY-FUNCTION THEORY APPROACHES

Rehybridization of σ and π orbitals and electronic correlations are known to affect the band structure, DOS, and transition dipole, and thus lead to substantial changes of the absorption spectrum. The tight-binding model presented in Section 2 grossly simplifies the electronic

dynamics in the nanotubes. For example, curvature effects are known to introduce small energy gaps in "metallic" zigzag SWNTs that depend inversely on the square of the tube radius [126–128]. Whether the low-energy absorption features as predicted by the tight-binding model survive rehybridization and correlation effects remains to be seen. To compare with the tight-binding low-energy absorption peaks, in this section we explicitly include the electron correlations and the σ–π orbital rehybridization by using the LDM-PM3 method. It is followed by employing the first-principles DFT [123]. It is shown that those low-energy spectral features, as illustrated by the tight-binding theory, persist in the presence of electron–electron Coulomb interactions [129] and σ–π orbital rehybridization, and are thus universal.

The optical responses of carbon nanotubes are highly anisotropic, as has been demonstrated by the tight-binding and LDM calculations in the previous sections [48]. Because of their unique geometric structure, the transition dipoles along the tubule axis **T** greatly exceeds their components perpendicular to **T**. For the purpose here of comparing various theoretical predictions of the SWNT optical absorption lineshapes, we again consider only the scenario in which the external field is parallel to the tubule axis (**E** $\parallel$ **T**), and thus the transition occurs exclusively between the orbitals of the same momenta **k** [51]. Three SWNTs, the (6, 4), (8, 0), and (5, 5) nanotubes, have similar diameters, which are 6.83, 6.26, and 6.78 Å, respectively. According to the tight-binding theory, (8, 0) and (6, 4) are semiconductors, whereas (5, 5) is a metal. In Fig. 35a, we plot the tight-binding absorption spectra of (6, 4), (8, 0), and (5, 5) nanotubes. The first two peaks at $0.42|V_{pp\pi}|$ ($0.48|V_{pp\pi}|$) and $0.80|V_{pp\pi}|$ ($0.83|V_{pp\pi}|$) belong to the semiconducting (6, 4) [(8, 0)] nanotube, and the third peak at $1.18|V_{pp\pi}|$ belongs to the metallic (5, 5) nanotube. All three tubes have large absorption peaks near $2|V_{pp\pi}|$ that are not shown in Fig. 35a.

The first two peaks of the (6, 4) nanotube are 0.36 and 0.68 of the lowest peak of the (5, 5) nanotube in energy, which are close to 1/3 and 2/3, respectively. The slight deviation stems from a minor departure of the $E_c - E_v$ contour from the circular feature near the K points, the so-called trigonal warping effect [44]; and deviation of $E_c - E_v$ from its linear

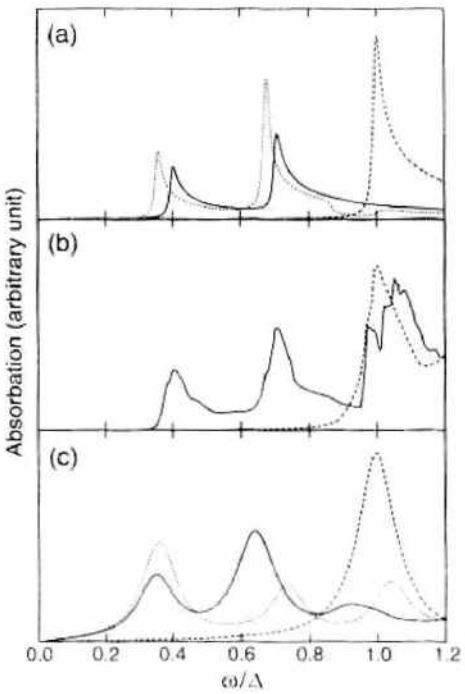


Figure 35. Calculated absorption spectra from (a) the tight-binding model, (b) DFT calculations implemented by the WIEN97 code, and (c) the LDM algorithm are shown for carbon nanotubes of three chiralities: (8, 0) (solid lines), (5, 5) (dashed lines), and (6, 4) (dotted lines); (6, 4) and (8, 0) are semiconducting, and (5, 5) is metallic. The symbol Δ labels the energy of the (5, 5) SWNT's first absorption peak: (a) $\Delta = 1.18|V_{pp\pi}|$, (b) $\Delta = 2.76$ eV, and (c) $\Delta = 2.34$ eV. A dephasing constant of 0.16 eV is used to generate the LDM spectra. Reprinted with permission from [43], Y. Zhao et al., *Chem. Phys. Lett.* 387, 149 (2004). © 2004, Elsevier B. V.

behavior as **k** moves away from the K point. The (8, 0) nanotube has a smaller diameter so that its absorption spectrum is blue-shifted compared to that of (6, 4), and its first and second peaks are 0.41 and 0.70 of the lowest peak of (5, 5), respectively.

Using the full-potential LAPW method implemented in the WIEN97 code [123], we have carried out the spectral calculations for two SWNTs: (5, 5) and (8, 0). Exchange and correlation are included via the local spin-density approximation (LSDA) within the DPT, using the Perdew–Wang parameterization [130]. Optical absorption spectra are calculated via the sum-of-states approach, taking into account pairwise interband transitions [123]. The resulting absorption spectra are shown in Fig. 35b. The lowest peak of the (5, 5) nanotube centers at ~2.76 eV, and the lowest two peaks of (8, 0) are 1.12 eV and 1.95 eV, which are 0.41 and 0.71 of the first peak of (5, 5) (2.76 eV), respectively. The DFT results of the (5, 5) and (8, 0) tubes therefore closely resemble those from the tight-binding model. We conclude that the tight-binding spectral features survive the inclusion of σ–π orbital rehybridization, as shown here by the first-principles DFT calculations.

The LDM method [48], which is size consistent and based on the random phase approximation, considers full single-electron excitation configurations and partial multiple-electron excitation configurations. Because all valence electrons are treated explicitly, the LDM calculation accounts for the σ–π orbital rehybridization as well. Absorption spectra of (5, 5) and (9, 0) SWNTs calculated by the LDM method were found to be similar, which led to our earlier hypothesis that the SWNT absorption lineshapes are mainly determined by their diameters [49]. The calculated LDM absorption spectra of the (8, 0), (5, 5), and (6, 4) SWNTs (C_{128} H_{16}, C_{140} H_{20} and C_{148} H_{20}, respectively) are shown in Fig. 35c. The lengths of the (8, 0), (5, 5), and (6, 4) SWNTs are approximately the same, at 14.92, 14.74, and 15.64 Å, respectively. The PM3 Hamiltonian is employed in the LDM calculations. The first absorption peak of the (5, 5) tube is at 2.34 eV. The lowest two absorption peaks for the (6, 4) tube are at 0.84 and 1.72 eV, which are 0.36 and 0.74 of 2.34 eV, respectively, and for the (8, 0) tube, they are 0.82 and 1.49 eV, which are 0.35 and 0.64 of 2.34 eV, respectively. The LDM results are in approximate agreement with those from the tight-binding model. Discrepancies between the LDM and tight-binding results for the (8, 0) SWNT may be attributed to its small size.

As mentioned earlier, most recent spectrofluorimetric measurements on SWNTs isolated in aqueous surfactant suspensions [35, 131–133] show that for semiconducting nanotubes the tight-binding model can also account for qualitatively small deviations of absorption peaks' dependence on diameter from linear relations, which bear signatures of nanotube chiral angles. Although a consensus on the strength on the electron–hole interaction has yet to emerge [134], that the central value of the energy ratios E_{22}/E_{11} is found to be slightly smaller than 2 [35, 34, 135] is attributed to many-electron correlation effects [35, 136, 137]. First and second van Hove transition wavelengths have been observed directly for 33 semiconducting SWNTs in a surfactant-suspended bulk sample, with each of these species assigned a (n, m) index [35, 131–133], and spectral transitions of six additional SWNT species have been reported in a study of SWNTs of slightly larger diameters [138]. From identified nanotubes ranging from 0.62 to 1.41 nm in diameter and from 3 to 28 degrees in chiral angle, measured transition frequencies have been used to anchor empirical fitting functions that allow extrapolation beyond the measured set [35, 135]. Such extrapolated spectral frequencies have been recently used to constrain theories on SWNT excitonic effects [137].

6. SUMMARY

A large literature of theoretical calculations exists on the electronic structures and optical properties of SWNTs [20–23, 88, 89, 116, 117, 139, 140]. Tight-binding [20–23] and DFT calculations [116, 117] conclude that an (m, n) SWNT is conductor if $m - n \bmod 3 = 0$ and is a semiconductor if $m - n \bmod 3 \neq 0$ [20–23]. Lin et al. studied the absorption spectra of bundles of SWNTs via a tight-binding model, which stressed the spectral differences caused by different chiral angles [51]. Margulis and Gaiduk also proposed a single-electron theory for the absorption spectra of a bundle of diameter-distributed SWNTs that gave good agreement with measurements with regard to both the position and spectral shape of the fundamental

absorption edge [141]. Hagen and Hertel discussed how tight-binding band-structure calculations with a chirality- and diameter-dependent nearest-neighbor-hopping integral may be used to relate well-resolved features in the absorption spectra of individual SWNTs to electronic excitations in specific tube types [142]. Our earlier LDM study [49] on the 4-Å SWNTs shows that the absorption spectra of the (4, 2), (3, 3), and (5, 0) SWNTs are distinct, although their diameters are virtually the same (4.2, 4.1, and 3.9 Å, respectively). When the CNT diameter d_t is comparable to C–C bond length a, adjacent π orbitals overlap. The extent of the orbital overlaps depends on the orientation along which the tube is rolled up, and this leads to the spectral lineshapes that are sensitive to the chiral angles. With support from both the DFT and LDM calculations, it can be concluded that to the lowest order in a/d_t, the simple tight-binding model provides an accurate description of low-energy optical processes in SWNTs [43]. The diameters of the (5, 5), (6, 4), and (8, 0) SWNTs are modestly large compared to the C–C bond length. Greater SWNT diameters are expected to bring the DFT and LDM calculations into closer agreement with the tight-binding results. In addition to its success in describing the nanotubes' band structures, the tight-binding model provides a remarkably accurate overall account of the low-energy absorption lineshapes. Optical properties of a series of finite-size SWNTs, including those with the smallest diameter (4 Å), DWNTs, CNT junctions, and SWNTs with localized defects have been studied systematically. Their absorption spectra are calculated with the LDM method. The Pariser–Parr–Pople (PPP) and MNDO parametric method 3 (PM3) semiempirical Hamiltonians are employed. The finite optical gaps are predicted for the infinite long SWNTs. Strong anisotropy of the dynamic polarizabilities is found for 4-Å SWNTs [143]. Calculated results on 4-Å SWNTs are in good agreement with the recent experimental findings. Furthermore, the compositions of the dipole-induced excitations are examined by projecting the corresponding density matrices onto the Hartree–Fock molecular orbital representation. Via the DFT approach electronic and optical properties of a potassium iodide intercalated (10, 10) nanotube are also calculated.

To summarize, a multiapproached picture is presented for optical absorption spectra of single-walled carbon nanotubes (SWNTs), and theories on SWNT optical properties of incremental sophistication have been reviewed. On the basis of the simplified treatment of the tight-binding model, a visual, intuitive connection is given between optical absorption lineshapes and the underlying carbon nanotube structures. Within the tight-binding model, the absorption spectra of SWNTs can be linked directly to plots of energy contours and transition dipoles of a graphene sheet. Calculations based on two distinctively different methods, the LDM and DFT approaches, result in common low-energy absorption spectral features for SWNTs with diameters much larger than the C–C bond length. These spectral features survive σ–π orbital rehybridization and electronic correlations and are further supported by measured absorption spectra and EELS of SWNT bundles. Among many useful results presented here is the vertical transition dipole $\mathbf{d}$ between the valence-band wave function $|\phi_v(\mathbf{k})\rangle$ and the corresponding conduction-band wave function $|\phi_c(\mathbf{k})\rangle$ defined as $\mathbf{d} \equiv \langle \phi_c(\mathbf{k})|e\mathbf{r}|\phi_c(\mathbf{k})\rangle$ (cf. Fig. 2c).

Emerging fields of nanotechnology hold the promise of overcoming limitations of existing technologies on nanoscale manipulation. In particular, these new developments provide approaches for the creation of chemical–biological hybrid nanocomposites that can have a variety of applications such as being introduced into cells, and they can subsequently be used to initiate intracellular processes or biochemical reactions. Such nanocomposites may advance medical biotechnology as much as introduce new possibilities in chemistry and material sciences. SWNTs and their derivatives are among the simplest nanocomposites that are currently undergoing intense experimental and theoretical investigations. Various manufacturing methods produce SWNTs in form of bundles or ropes intertwined with the van der Waals interactions, and identification of spectroscopic signatures and correlation with nanotube structures helps purify and segregate SWNTs. Optical properties of SWNTS therefore become indispensable to structure-based nanotube characterization and separation.

Recently discovered near-infrared band-gap photoluminescence of dispersed, micelle-isolated semiconducting SWNTs allow optical-spectra-based characterization of SWNT structures. Optical measurements of individual SWNTs in aqueous surfactant suspensions provide

information that is sensitive to both SWNT chiralities and diameters [35, 131–133]. Time-resolved carrier dynamics in SWNTs have been investigated by means of two-color pump-probe techniques [144]. Using various pump-probe wavelengths and intensities, femtosecond dynamics of photoexcitations in films containing semiconducting and metallic SWNTs have also been studied to probe confined excitons and charge carriers [145]. More stringent tests on theoretical models of SWNT optical properties will soon be available from Raman, luminescence, and ultrafast spectroscopic measurements that are being carried out for isolated SWNTs as well as for SWNT bundles.

ACKNOWLEDGMENTS

This work is supported by the Hong Kong Research Grant Council (RGC) and the Committee for Research and Conference Grants (CRCG) of the University of Hong Kong. The authors would like to thank W. Z. Liang, C. Y. Yam, and S. Yokojima for their help in preparing the manuscript.

REFERENCES

1. S. Iijima, *Nature* 354, 56 (1991).
2. T. W. Ebbesen and P. M. Ajayan, *Nature* 358, 220 (1992); T. W. Ebbesen, H. Hiura, J. Fujita, Y. Ochiai, S. Matsui, and K. Tanigaki, *Chem. Phys. Lett.* 209, 83 (1993).
3. S. Iijima and T. Ichihashi, *Nature* 363, 603 (1993).
4. D. S. Bethune, C. H. Kiang, M. S. Devries, G. Gorman, R. Savoy, J. Vazquez, and R. Beyers, *Nature* 363, 605 (1993).
5. S. Iijima, *Mater. Sci. Eng. B* 19, 172 (1993).
6. K. Sattler, *Carbon*, 33, 915 (1995).
7. M. Bockrath, D. H. Cobden, P. L. McEuen, N. G. Chopra, A. Zettl, A. Thess, and R. E. Smalley, *Science* 275, 1922 (1997).
8. K. Tsukagoshi, B. W. Alphenaar, and H. Ago, *Nature* 401, 572 (1999).
9. R. D. Antonov and A. T. Johnsons, *Phys. Rev. Lett.* 83, 3274 (1999).
10. Z. Yao, H. W. Ch. Postma, L. Balents, and C. Dekker, *Nature* 402, 273 (1999).
11. R. E. Tuzun, D. W. Noid, and B. G. Sumpter, *Nanotechnology* 6, 64 (1995).
12. D. Srivastava, *Nanotechnology* 8, 186 (1997).
13. J. Cumings and A. Zettl, *Science* 289, 602 (2000).
14. S. Berber, Y. K. Kwon, and D. Tománek, *Phys. Rev. Lett.* 91, 165503 (2003).
15. Q. S. Zheng and Q. Jiang, *Phys. Rev. Lett.* 88, 045503 (2002); Q. S. Zheng, J. Z. Liu, and Q. Jiang, *Phys. Rev. B* 65, 245209 (2002).
16. Y. Zhao, C. C. Ma, G. H. Chen, and Q. Jiang, *Phys. Rev. Lett.* 91, 175504 (2003).
17. S. B. Legoas, V. R. Coluci, S. F. Braga, P. Z. Coura, S. O. Dantas, and D. S. Galvão, *Phys. Rev. Lett.* 90, 055504 (2003)
18. J. L. Rivera, C. McCabe, and P. T. Cummings, *Nano. Lett.* 3, 1001 (2003).
19. W. L. Guo, Y. F. Guo, H. J. Gao, Q. S. Zheng, and W. Y. Zhong, *Phys. Rev. Lett.* 91, 125501 (2003).
20. N. Hamada, S. Sawada, and A. Oshiyama, *Phys. Rev. Lett.* 68, 1579 (1992).
21. M. S. Dresselhaus, G. Dresselhaus, and P. C. Eklund, "Science of Fullerenes and Carbon Nanotubes," Academic Press, San Diego, CA, 1996.
22. R. Saito, M. Fujita, G. Dresselhaus, and M. S. Dresselhaus, *Appl. Phys. Lett.* 60, 2204 (1992); R. Saito, M. Fujita, G. Dresselhaus, and M. S. Dresselhaus, *Phys. Rev. B* 46, 1804 (1992).
23. R. Saito, G. Dresselhaus, and M. S. Dresselhaus "Physical Properties of Carbon Nanotubes," Imperial College Press, London, 1998.
24. C. T. White, D. H. Robertson, and J. W. Mintmire, *Phys. Rev. B* 47, 5485 (1993).
25. K. Keren, R. S. Berman, E. Buchstab, U. Sivan, and E. Braun, *Science* 302, 1380 (2003).
26. M. Freitag, Y. Martin, J. A. Misewich, R. Martel, and P. Avouris, *Nano. Lett.* 3, 1067 (2003).
27. J. A. Misewich, R. Martel, Ph. Avouris, J. C. Tsang, S. Heinze, and J. Tersoff, *Science* 300, 783 (2003).
28. M. Zheng, A. Jagota, M. S. Strano, A. P. Santos, P. Barone, S. G. Chou, B. A. Diner, M. S. Dresselhaus, R. S. Mclean, G. B. Onoa, G. G. Samsonidze, E. D. Semke, M. Usrey, and D. J. Walls, *Science* 302, 1545 (2003).
29. T. Pichler, M. Knupfer, M. S. Golden, J. Fink, A. Rinzler, and R. E. Smalley, *Phys. Rev. Lett.* 80, 4729 (1998).
30. O. Jost, A. A. Gorbunov, W. Pompe, T. Pichler, R. Friedlein, M. Knupfer, M. Reibold, H.-D. Bauer, L. Dunsch, M. S. Golden, and J. Fink, *Appl. Phys. Lett.* 75, 2217 (1999).
31. H. Kataura, Y. Kumazawa, Y. Maniwa, I. Umezu, S. Suzuki, Y. Ohtsuka, Y. Achiba, *Synth. Met.* 103, 2555 (1999).
32. N. Wang, Z. K. Tang, G. D. Li, and J. S. Chen, *Nature* 408, 50 (2000).

33. Z. M. Li, Z. K. Tang, H. J. Liu, N. Wang, C. T. Chan, R. Saito, S. Okada, G. D. Li, J. S. Chen, N. Nagasawa, and S. Tsuda, *Phys. Rev. Lett.* 87, 127401 (2001).
34. M. J. O'Connell, S. M. Bachilo, C. B. Huffman, V. C. Moore, M. S. Strano, E. H. Haroz, K. L. Rialon, P. J. Boul, W. H. Noon, C. Kittrell, J. P. Ma, R. H. Hauge, R. Bruce Weisman, and R. E. Smalley, *Science* 297, 593 (2002).
35. S. M. Bachilo, M. S. Strano, C. Kittrell, R. H. Hauge, R. E. Smalley, and R. Bruce Weisman, *Science* 298, 2361 (2002).
36. M. S. Strano, S. K. Doorn, E. H. Haroz, C. Kittrell, R. H. Hauge, and R. E. Smalley, *Nano. Lett.* 3, 1091 (2003).
37. M. Zheng, A. Jagota, E. D. Semke, B. A. Diner, R. S. Mclean, S. R. Lustig, R. E. Richardson, N. G. Tassi, *Nat. Mater.* 2, 338 (2003).
38. K. Kamaras, M. E. Itkis, H. Hu, B. Zhao, and R. C. Haddon, *Science* 301, 1501 (2003).
39. M. S. Strano, C. A. Dyke, M. L. Usrey, P. W. Barone, M. J. Allen, H. W. Shan, C. Kittrell, R. H. Hauge, J. M. Tour, and R. E. Smalley, *Science* 301, 1519 (2003).
40. R. S. Lee, H. J. Kim, J. E. Fischer, A. Thess, and R. E. Smalley, *Nature* 388, 255 (1997).
41. J. Sloan, M. C. Novotny, S. R. Bailey, G. Brown, C. Xu, V. C. Williams, S. Friedrichs, E. Flahaut, R. L. Callender, A. P. E. York, K. S. Coleman, M. L. H. Green, R. E. Dunin-Borkowski, and J. L. Hutchison, *Chem. Phys. Lett.* 329, 61 (2000).
42. R. R. Meyer, J. Sloan, R. E. Dunin-Borkowski, A. I. Kirkland, M. C. Novotny, S. R. Bailey, J. L. Hutchison, and M. L. H. Green, *Science* 289, 1324 (2000).
43. Y. Zhao, X. J. Wang, C. C. Ma, and G. H. Chen, *Chem. Phys. Lett.* 387, 149 (2004).
44. R. Saito, G. Dresselhaus, and M. S. Dresselhaus, *Phys. Rev. B* 61, 2981 (2000).
45. S. Reich and C. Thomsen, *Phys. Rev. B* 62, 4273 (2000).
46. C. T. White and J. W. Mintmire, *Nature* 394, 29 (1998).
47. J. W. G. Wildöer, L. C. Venema, A. G. Rinzler, R. E. Smalley, and C. Dekker, *Nature* 391, 59 (1998).
48. S. Yokojima and G. H. Chen, *Chem. Phys. Lett.* 300, 540 (1999).
49. W. Z. Liang, S. Yokojima, D. H. Zhou, and G. H. Chen, *J. Phys. Chem. A* 104, 2445 (2000).
50. W. A. de Heer, W. S. Baesa, A. Châtelain, T. Gerfin, R. Humphrey-Baker, L. Forro, and D. Ugarte, *Science* 268, 845 (1995).
51. M. F. Lin and W. K. Shung, *J. Phys. Soc. Jpn.* 66, 3294 (1997).
52. M. F. Lin, *Phys. Rev. B* 62, 13153 (2000).
53. A. Grüneis, R. Saito, Ge. G. Samsonidze, T. Kimura, M. A. Pimenta, A. Jorio, A. G. Souza Filho, G. Dresselhaus, and M. S. Dresselhaus, *Phys. Rev. B* 67, 165402 (2003).
54. F. L. Shyu, C. P. Chang, R. B. Chen, C. W. Chiu, and M. F. Lin, *Phys. Rev. B* 67, 045405 (2003).
55. H. Ajiki and T. Ando, *Physica B* 201, 349 (1994).
56. A. G. Souza Filho, A. Jorio, Ge. G. Samsonidze, G. Dresselhaus, M. S. Dresselhaus, A. K. Swan, M. S. Ünlüe, B. B. Goldberg, R. Saito, J. H. Hafner, C. M. Lieber, and M. A. Pimenta, *Chem. Phys. Lett.* 354, 62 (2002).
57. Ge. G. Samsonidze, R. Saito, A. Jorio, A. G. Souza Filho, A. Grüneis, M. A. Pimenta, G. Dresselhaus, and M. S. Dresselhaus, *Phys. Rev. Lett.* 90, 027403 (2003).
58. X. Liu, T. Pichler, M. Knupfer, M. S. Golden, J. Fink, H. Kataura, and Y. Achiba, *Phys. Rev. B* 66, 045411 (2002).
59. P. Ring and P. Schuck, "The Nuclear Many-Body Problem," Springer, New York, 1980.
60. S. Yokojima and G. H. Chen, *Chem. Phys. Lett.* 292, 379 (1998).
61. S. Yokojima and G. H. Chen, *Phys. Rev. B* 59, 7259 (1999).
62. M. F. Li, "Modern Semiconductor Quantum Physics," World Scientific, Singapore, 1995.
63. S. Yokojima, D. H. Zhou, and G. H. Chen, *Chem. Phys. Lett.* 302, 495 (1999).
64. W. Z. Liang, S. Yokojima, and G. H. Chen, *J. Chem. Phys.* 110, 1844 (1999).
65. W. Z. Liang, X. J. Wang, S. Yokojima, and G. H. Chen, *J. Am. Chem. Soc.* 122, 11129 (2000).
66. W. Z. Liang, S. Yokojima, M. F. Ng, G. H. Chen, and G. Z. He, *J. Am. Chem. Soc.* 123, 9830 (2001).
67. S. Yokojima, X. J. Wang, D. H. Zhou, and G. H. Chen, *J. Chem. Phys.* 111, 10444 (1999).
68. R. F. Willis, B. Feuerbacher, and B. Fitton, *Phys. Rev. B* 4, 2441 (1971).
69. J. Del Bene and H. H. Jaffe, *J. Chem. Phys.* 48, 1807 (1968); *J. Chem. Phys.* 48, 4050 (1968).
70. J. A. Pople, D. L. Beveridge, and P. A. Dobosh, *J. Chem. Phys.* 47, 2026 (1967).
71. M. J. S. Dewar and W. Thiel, *J. Am. Chem. Soc.* 99, 4899 (1977).
72. M. J. S. Dewar, E. G. Zoebisch, E. F. Healy, and J. J. P. Stewart, *J. Am. Chem. Soc.* 107, 3902 (1985).
73. J. J. P. Stewart, *J. Comput. Chem.* 10, 209 (1989).
74. X.-M. Duan, H. Konami, S. Okaka, H. Oikawa, H. Matsuda, and H. Nakanishi, *J. Phys. Chem.* 100, 17780 (1996).
75. U. S. Choi, T. W. Kim, S. W. Jung, and C. J. Kim, B. Kor, *Chem. Soc.* 19, 299 (1998).
76. N. Matsuzawa and D. A. Dixon, *J. Phys. Chem.* 96, 6232 (1992).
77. S. Sawada and N. Hamada, *Solid State Commun.* 83, 917 (1992).
78. J. Tersoff, *Phys. Rev. Lett.* 61, 2879 (1988).
79. L.-C. Qin, X. Zhao, K. Hirahara, Y. Miyamoto, Y. Ando, and S. Iijima, *Nature* 408, 50 (2000).
80. H. Y. Peng, N. Wang, Y. F. Zheng, Y. Lifshitz, J. Kulik, R. Q. Zhang, C. S. Lee, and S. T. Lee, *Appl. Phys. Lett.* 77, 2831 (2000).
81. Z. K. Tang, private communication.

92. S. Leach, M. Vervloet, A. Despres, E. Breheret, J. P. Hare, T. J. Dennis, H. W. Kroto, R. Taylor, and D. R. M. Walton, *Chem. Phys.* 160, 451 (1992).
93. R. Saito, T. Takeya, T. Kimura, G. Dresselhaus, and M. S. Dresselhaus, *Phys. Rev. B* 59, 2388 (1999).
94. Z. Zhang and C. M. Lieber, *Appl. Phys. Lett.* 62, 2792 (1993).
95. T. W. Odom, J.-L. Huang, P. Kim, and C. M. Lieber, *Nature* 391, 62 (1998).
96. E. W. Wong, P. E. Sheehan, and C. M. Lieber, *Science* 277, 1971 (1997).
97. 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).
98. J. Ma and R.-K. Yuan, *Phys. Rev. B* 57, 9343 (1998).
99. J. Jiang, J. Dong, and D. Y. Xing, *Phys. Rev. B* 59, 9838 (1999); J. Jiang, J. M. Dong, X. G. Wan, and D. Y. Xing, *J. Phys. B* 31, 3079 (1998).
100. J. Chen, M. A. Hamon, H. Hu, Y. Chen, A. M. Rao, P. C. Eklund, and R. C. Haddon, *Science* 282, 95 (1998).
101. M.-F. Lin and K. W.-K. Shung, *Phys. Rev. B* 50, 17744 (1994); M.-F. Lin, K. W.-K. Shung, D. S. Chuu, C. S. Huang, and Y. K. Lin, *Phys. Rev. B* 53, 15493 (1996).
102. P. M. Ajayan, S. Iijima, and T. Ichihashi, *Phys. Rev. B* 47, 6859 (1993).
103. A. Rochefort, D. R. Salahub, and P. Avouris, *J. Phys. Chem. B* 103, 641 (1999).
104. P. Kim, T. W. Odom, J. L. Huang, and C. M. Lieber, *Phys. Rev. Lett.* 82, 1225 (1999).
105. J. C. Charlier, *Phys. Rev. B* 57, R15037 (1998).
106. A. Rochefort, D. S. Salahub, and P. Avouris, *Chem. Phys. Lett.* 297, 45 (1998).
107. A. M. Rao, E. Richter, S. Bandow, B. Chase, P. C. Elund, K. A. Williams, S. Fang, K. R. Subbaswamy, M. Menon, A. Thess, R. E. Smalley, G. Dresselhaus, and M. S. Dresselhaus, *Science* 275, 187 (1997).
108. X. Blase, L. X. Benedict, E. L. Shirley, and S. G. Louie, *Phys. Rev. Lett.* 72, 1878 (1994).
109. L. G. Bulusheva, A. V. Okotrub, D. A. Romanov, and D. Tomanek, *J. Phys. Chem. A* 102, 975 (1998).
110. R. C. Haddon, *Acc. Chem. Res.* 21, 243 (1988).
111. R. C. Haddon, *J. Phys. Chem. A* 105, 4164 (2001).
112. W. Ludwig and C. Falter, "Symmetries in Physics: Group Theory Applied to Physical Problems," Springer, New York, 1996.
113. G. H. Chen and S. Mukamel, *J. Am. Chem. Soc.* 117, 4945 (1995).
114. B. Zhou, Y. Lin, H. Li, W. Huang, J. W. Connell, L. F. Allard, and Y. P. Sun, *J. Phys. Chem. B* 107, 13588 (2003).
115. S. Banerjee and S. S. Wong, *J. Am. Chem. Soc.* 124, 8940 (2002).
116. F. Pompeo and D. E. Resasco, *Nano. Lett.* 2, 369 (2002).
117. Z. Jin, X. Sun, G. Xu, S. H. Goh, and W. Ji, *Chem. Phys. Lett.* 318, 505 (2000).
118. J. L. Bahr and J. M. Tour, *J. Mat. Chem.* 12, 1952 (2002).
119. S. Okada and A. Oshiyama, *Phys. Rev. Lett.* 91, 216801 (2003).
120. Y. Miyamoto, S. Saito, and D. Tománek, *Phys. Rev. B* 65, 041402 (2001).
121. H. Stahl, J. Appenzeller, R. Martel, Ph. Avouris, and B. Lengeler, *Phys. Rev. Lett.* 85, 5186 (2000).
122. P. Delaney, H. J. Choi, J. Ihm, S. G. Louie, and M. L. Cohen, *Nature* 391, 466 (1998).
123. Y.-K. Kwon, S. Saito, and D. Tománek, *Phys. Rev. B* 58, R13314 (1998).
124. B. W. Smith, M. Monthioux, and D. E. Luzzi, *Nature* 396, 323 (1998).
125. L. Chico, V. H. Crespi, L. X. Benedict, S. G. Louie, and M. L. Cohen, *Phys. Rev. Lett.* 76, 971 (1996).
126. A. Rubio, *Appl. Phys. A* 68, 275 (1999).
127. A. Rubio, D. Sanchez-Portal, E. Artacho, P. Ordejón, and J. M. Soler, *Phys. Rev. Lett.* 82, 3520 (1999).
128. H. J. Liu and C. T. Chan, *Phys. Rev. B* 66, 115416 (2002).
129. A. G. Marinopoulos, L. Reining, A. Rubio, and N. Vast, *Phys. Rev. Lett.* 91, 046402 (2003).
130. D.-S. Wu, W.-D. Cheng, H. Zhang, X.-D. Li, Y.-Z. Lan, D.-G. Chen, Y.-J. Gong, and Y.-C. Zhang, *Phys. Rev. B* 68, 125402 (2003).
131. P. Hohenberg and W. Kohn, *Phys. Rev.* 136, B864 (1964).
132. W. Kohn and L. J. Sham, *Phys. Rev.* 140, A1133 (1965).
133. P. Blaha, K. Schwarz, and J. Luitz, "WIEN97, A Full Potential Linearized Augmented Plane Wave Package for Calculating Crystal Properties," Karlheinz Schwarz, Techn. Universitat, Wien, 1999.
134. S. Okada, S. Saito, and A. Oshiyama, *Phys. Rev. Lett.* 86, 3835 (2001).
135. T. Pichler, X. Liu, M. Knupfer, and J. Fink, *New J. Phys.* 5, 156 (2003).
136. M. Ouyang, J. L. Huang, C. L. Cheung, and C. M. Lieber, *Science* 292, 702 (2001).
137. J. W. Ding, X. H. Yan, and J. X. Cao, *Phys. Rev. B* 66, 073401 (2002).
138. M. E. Itkis, S. Niyogi, M. E. Meng, M. A. Hamon, H. Hu, and R. C. Haddon, *Nano. Lett.* 2, 155 (2002).
139. M. Ichida, S. Mizuno, Y. Saito, H. Kataura, Y. Achiba, and A. Nakamura, *Phys. Rev. B* 65, 241407 (2002).
140. J. P. Perdew and Y. Wang, *Phys. Rev. B* 45, 13244 (1992).
141. S. M. Bachilo, L. Balzano, J. E. Herrera, F. Pompeo, D. E. Resasco, and R. B. Weisman, *J. Am. Chem. Soc.* 125, 11186 (2003).
142. S. Maruyama, Y. Miyauchi, Y. Murakami, and S. Chiashi, *New J. Phys.* 5, 149 (2003).
143. S. Lebedkin, K. Arnold, F. Hennrich, R. Krupke, B. Renker, and M. M. Kappes, *New J. Phys.* 5, 140 (2003).
144. A. Hartschuh, H. N. Pedrosa, L. Novotny, and T. D. Krauss, *Science* 301, 1354 (2003).
145. R. B. Weisman and S. M. Bachilo, *Nano. Lett.* 3, 1235 (2003).
146. C. L. Kane and E. J. Mele, *Phys. Rev. Lett.* 90, 207401 (2003).

137. C. D. Spataru, S. Ismail-Beigi, L. X. Benedict, and S. G. Louie, *Phys. Rev. Lett.* 92, 07742 (2004).
138. S. Lebedkin, F. Hennrich, T. Skipa, and M. M. Kappes, *J. Phys. Chem. B* 107, 1949 (200).
139. X. Wan, J. Dong, and D. Y. Xing, *Phys. Rev. B* 58, 6756 (1998).
140. T. Ando, *J. Phys. Soc. Jpn.* 9h, 1066 (1997).
141. V. A. Margulis and E. A. Gaiduk, *Phys. Lett. A* 281, 52 (2001).
142. A. Hagen and T. Hertel, *Nano. Lett.* 3, 383 (2003).
143 W. Z. Liang, G. H. Chen, Z. M. Li, and Z. K. Tang, *Appl. Phys. Lett.* 80, 3415 (2002).
144. J.-S. Lauret, C. Voisin, G. Cassabois, C. Delalande, P. Roussignol, O. Jost, and L. Capo, ***Phys. Rev. Lett.*** 90, 057404 (2003).
145. O. J. Korovyanko, C. X. Sheng, Z. V. Vardeny, A. B. Dalton, and R. H. Baughman, ***PhysRev. Lett.*** 92, 017403 (2004).

CHAPTER 3

Nonlinear Optical Properties of Carbon Nanostructures

Rui-Hua Xie
Department of Chemistry, Queen's University, Kingston, Ontario, Canada

Tapas Kar
Department of Chemistry and Biochemistry, Utah State University, Logon, Utah, USA

Zhigang Li
State Key Laboratory of Applied Optics, Changchun Institute of Optics and Fine Mechanics Chinese Academy of Sciences, Changchun, People's Republic of China

CONTENTS

ISBN 1-58883-149-7

Handbook of Theoretical and Computational Nanotechnology
Edited by Michael Rieth and Wolfram Schommers
Volume 7: Pages (127–207)

1. INTRODUCTION

1.1. Nanoworld

The little word *nano* has been rapidly insinuating itself into our consciousness because of its big potential [1]. In the media, *nano* has captured headlines on television news channels and almost every online technical and scientific journal. Physicists, chemists, biologists, engineers, computer scientists, physicians, and other researchers are intimately involved in *nano* development and are able to control and study materials in the nanoworld, leading to great achievements in nano research [1]. For example, the Nobel Prize has been awarded several times for nano research, the Feynman Prize was created to recognize the accomplishments of nanoscientists, and *Science* magazine named a nano development as Breakthrough of the Year in 2001. Innovations in nano-related fields have already sparked a flurry of commercial inventions, for example, faster-burning rocket fuel additives, nano skin creams and suntan lotions, and nano-enhanced tennis balls [1]. Then, what is the nano? The *nano* means one billionth. One nanometer (abbreviated as 1 nm) is 10^{-9} of a meter. To get a sense of the nano scale, a human hair measures 50,000 nm wide, and 10 hydrogen atoms in a line make up about 1 nm.

What are *nanoscience* and *nanotechnology*? *Nanoscience* is the study of the fundamental principles of molecules and structures with at least one dimension roughly between 1 and 100 nm [1]. These structures are the so-called nanostructures. As for *nanotechnology*, it talks about the technological applications of these nanostructures into useful *nanoscale* devices [1]. In recent years, nanoscience and nanotechnology have become one of the most important and exciting fields in physics, chemistry, biology, medicine, material science, and engineering, promising us many breakthroughs and showing us new technological directions for a wide range of applications. All these are mainly based on two fundamental facts: (1) particles in the nanoworld are smaller than the characteristic lengths associated with particular phenomena, and (2) these nanoparticles will lead to new behaviors that strongly depend on their sizes.

Actually, functional nanodevices and nanostructures have been on Earth since the beginning of life itself and are of great interest to many researchers who try to understand the unusual behaviors of particles or matters displayed in the nanoworld. For example, the strong shells of abalone represent a natural demonstration that a structure fabricated from nanoparticles can be much stronger [2]. Roman glassmakers in the fourth century A.D. fabricated glasses containing nanosized metals [2]. The Lycurgus cup, which depicts the death of King Lycurgus and resides in the British Museum in London, is made from a kind of glass that contains gold and silver nanoparticles [2]. As a light source is placed inside the cup, the color of the Lycurgus cup varies from a deep red to green. In 1857, Michael Faraday attempted to understand how metal particles affect the color of church windows [2]. In 1908, Gustav Mie was the first to explain why the color of the glasses depends on the metal size [2]. Photography depends on production of silver nanoparticles that are sensitive to light [2].

In the 1980s interest in nanoscience and nanotechnology increased because of the rapid development of appropriate fabrication techniques for nanostructures (see details in

Ref. [2]). So far, the electronic structure, conductivity, reactivity, melting temperature, and magnetic and mechanical properties of particles or matters have been observed to change greatly as these particles or matters become smaller than critical sizes. And these size-dependent properties have led to tailorable or engineered techniques that expose to us a wide and diverse range of technological areas, for example, making stronger and lighter materials, increasing the storage capacity of magnetic tapes, providing faster sensors and switches for computers, and shortening the delivery time of nanostructured pharmaceuticals to the body's circulatory systems. Obviously, these extensive studies in nanoscience and nanotechnology have led to a range of commercial applications in electronics, photonics, spintronics, sensors, military uses, medicine, solar cells, fuel cells, and so on [3]. As mentioned, the research area of nanoscience and nanotechnology is interdisciplinary, covering a wide variety of subjects in physics, chemistry, biology, medicine, material science, and engineering. Certainly, the best and most exciting days in nanoscience and nanotechnology are still ahead. This chapter is concerned with only the nonlinear optical (NLO) properties of carbon nanostructures.

1.2. Carbon Nanostructures

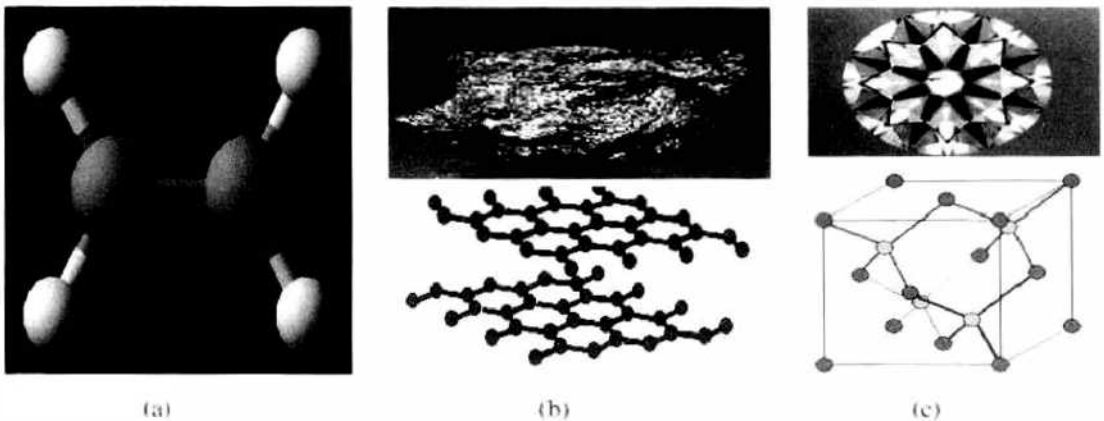
Carbon was known in ancient times, and its name came from the Latin *carbo*, "coal." It is the key element not only of terrestrial life but also of minerals (carbonates) and fossil fuels (oil, gas, and coal), as well as a minor but essential component of our atmosphere.

1.2.1. Nature of Carbon Bonds

Because carbon bonding plays an important role in building organic molecules of life and carbon nanostructures, here we briefly introduce the nature of the carbon bond. It is known that carbon atom contains six electrons distributed over the lowest energy levels of the carbon atom. The lowest energy level is 1s, occupied by two electrons with oppositely paired electron spins. Both 1s electrons do not participate in the chemical bonding. The next four electrons are in the 2s, $2p_x$, $2p_y$, and $2p_z$ orbitals. The 2s and the three 2p orbitals are hybridized to form three kinds of chemical bonds of carbon with other atoms, namely, sp^n hybridization: (1) diagonal sp^1 with the resulting bond angle of 180°, (2) trigonal sp^2 with the bond angle of 120°, and (3) tetrahedral sp^3 with the bond angle of 109°28′. It is the diverse nature of the carbon bond that allows carbon to form many interesting nanostructures, particularly carbon nanotubes. See examples shown in Fig. 1.

1.2.2. Graphite and Diamond

Elemental carbon is found in nature as its allotropes graphite and diamond, which are of vastly differing abundance and thus also of differing value. Graphite is known as one stable and abundant solid form of pure carbon [4, 5]. As shown in Fig. 1b, this form has three strong



(a) (b) (c)

Figure 1. Examples of sp^n hybridization ($n = 1, 2, 3$): (a) acetylene with a diagonal sp^1; (b) graphine with a trigonal sp^2; and (c) diamond with a tetrahedral sp^3.

sp^2 trigonal bonds, with an equal distance of 0.142 nm and weak π bonds perpendicular to successive sheets with an interplane distance of 0.34 nm [6]. Diamond is the hardest of all natural materials and slightly less stable and less abundant crystallographic form of pure carbon [4]. As shown in Fig. 1e, each carbon atom in diamond is covalently bonded to four neighbors via sp^3 hybridization at the apexes of a regular tetrahedron.

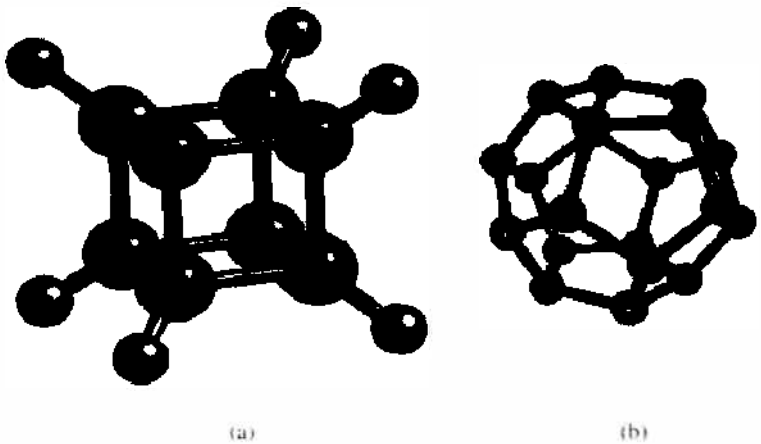
1.2.3. Cubane and Carbon Dodecahedron

Before 1964, it was generally believed that there exist only three kinds of carbon bond angles, as mentioned previously. In that year, Phil Eaton of the University of Chicago synthesized a square carbon molecule, C_8H_8, named *cubane* [2], as shown in Fig. 2a. Twenty years later, L. Paquette of the Ohio State University synthesized a $C_{20}H_{20}$ molecule that has a dodecahedron shape [2] as shown in Fig. 2b. Both hydrocarbon molecules have carbon bond angles ranging from 108° to 110°, different from the standard hybridization values mentioned before. This implies the formation of new carbon nanostructures, which may require different bonding angles.

1.2.4. Spherically Shaped Carbon Nanostructures

This kind of carbon nanostructure includes fullerenes, carbon blacks, carbon nanoparticles, carbon onions, and porous carbons [4].

1.2.4.1. Fullerenes Laser evaporation of a carbon substrate such as graphite can be used to make carbon clusters with the apparatus shown in Fig. 3. The neutral cluster beam is photoionized by a UV laser and analyzed by a mass spectrometer. In 1984, Rohlfing, Cox, and Kaldor [7] reported a typical mass spectrum of carbon clusters. For the carbon number N less than 30, there are clusters for every N. These small clusters have linear structures if N is odd and closed nonplanar monocyclic geometries if N is even. The carbon structures with $N = 3$, 11, 15, 19, and 23 have the sp hybridization and the standard bond angle (180°) and are more prominent and more stable [8]. However, the closed structures have angles between the carbon bonds, which are different from the standard hybridization process previously mentioned. More details can be found in Ref. [8]. For the carbon number $N \geq 40$, there are carbon clusters for even N. Especially, there is a very large mass peak at 60. In 1985, to explain and understand this peak, Kroto et al. [9] proposed a C_{60} structure (see Fig. 4a), which Osawa [10] theoretically predicted to be chemically stable in 1972. This structure has 20 hexagonal and 12 pentagonal faces and 60 vertices (each vertice is at the intersection of two hexagonal and one pentagonal faces). This was also a somewhat fortuitous result of research on the nature of matter in outer space that involved studies of light transmission through interstellar dust, the small particles of matter that fill the regions of outer space between stars and galaxies. This discovery brought us a new form of pure carbon, named *fullerenes* [11–16] (see Fig. 4 for detailed structures of several fullerenes that are discussed in this chapter), which are made of an even number of three-coordinated



(a) (b)

Figure 2. Carbon structures: (a) a cubane (C_8H_8); (b) carbon dodecahedron structure $C_{20}H_{20}$ (H atoms not shown).

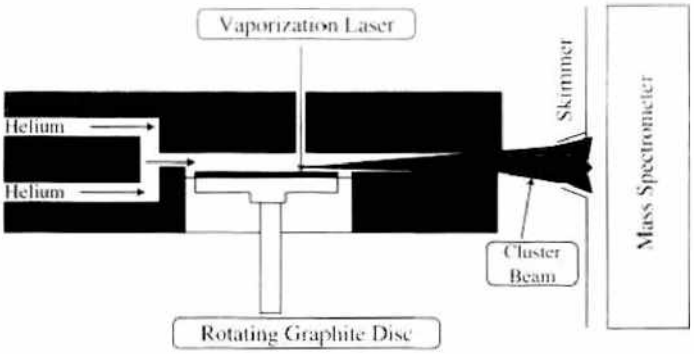


Figure 3. Apparatus to make carbon nanoparticles by laser-induced evaporation of carbon atoms from the surface of a graphite. Helium is used to cool the hot jet of carbon and form carbon clusters, which are skimmed to a mass spectrometer.

sp^2 carbon atoms that arrange themselves into 12 pentagonal faces and any number (>1) of hexagonal faces. This kind of carbon nanostructure can crystallize in a variety of three-dimensional structures [17, 18]. The macroscopic synthesis of a soot [17], which contains C_{60} sand other fullerenes in large compounds, and the straightforward purification techniques of the soot, which make the pure fullerene materials available, have led to extensive studies on the structural, mechanical, electronic, magnetic, and optical properties of fullerenes [19–29]. Meanwhile, fullerenes can be doped in several different ways—for example, endohedral, exohedral, and substitutional doping—because of their unique electronic properties [22] (see Fig. 5). These doped fullerenes can be semiconducting, metallic, or even superconducting [22]. Fullerenes are entirely insoluable in water. However, suitable functionalization makes the fullerenes soluable, and studies on water-soluable fullerene derivatives led to the discovery of the interaction organofullerenes with DNA, proteins, and living cells. See details reviewed by Nakamura and Isobe [30].

1.2.4.2. Carbon Blacks Classical carbon blacks, which are finely divided carbon particles, are produced by hydrocarbon dehydrogenation [31]. They are widely used in industry as a filler to modify the mechanical, electrical, and optical properties of the materials in which they are dispersed [31]. The microcrystalline structures of several types of carbon blacks in sizes of 100 nm and higher were established in as-synthesized or heat-treated (up to 3000°C) samples. The as-synthesized carbon blacks are composed of small graphite like layers, where the dimensions of the layers are described by two characteristic lengths, L_a (the crystalline size in the plane of the graphene layers) and L_c (the size along the c-axis perpendicular to the planes), and a concentric organization of the graphite layers in each individual particle.

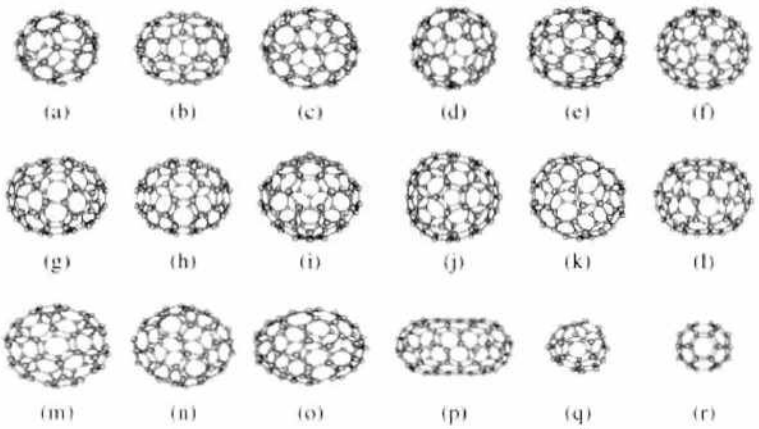


Figure 4. Some fullerene structures discussed in this chapter: (a) C_{60} (I_h); (b) C_{70} (D_{5h}); (c) C_{76} (D_2); (d) C_{76} (Td); (e) C_{78} (C_{2v}:1); (f) C_{78} (C_{2v}:2); (g) C_{78} (D_3); (h) C_{78} (D_{3h}:1); (i) C_{78} (D_{3h}:2); (j) C_{80} (C_{2v}); (k) C_{80} (D_2); (l) C_{80} (D_{5d}); (m) C_{84} (D_{6h}); (n) C_{84} (D_2); (o) C_{84} (D_{2d}); (p) C_{540} (D_{6h}); (q) C_{70} (D_{5h}); and (r) C_{20} (D_{3d}). The specific symmetry of each carbon structure is given.

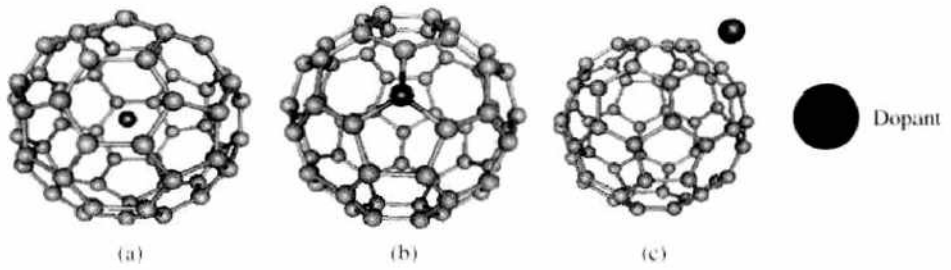


Figure 5. Three doping methods of permanently altering the charge distribution and electronic structures of the parent fullerenes: (a) the inclusion of atoms or ions inside the fullerene cage, (b) the substitution of one or more of the carbon atoms with dopants, and (c) the introduction of dopant species surrounding the fullerene.

Subsequent heat treatment of as-synthesized carbon blacks produced polygonized particles with an empty core and a well-graphitized carbon shell centered around the growth starting point [31].

1.2.4.3. Carbon Nanoparticles The carbon nanoparticles are often found along the synthesis of carbon nanotubes (as described later). They are hollow nanoparticles or filled nanocapsules. The filled nanocapsules are synthesized as the carbon electrode or carbon target is fabricated from a mixture of carbon with a small amount of an appropriate transition metal or rare earth metal [4]. The carbon-coated, for example, YC_2 nanoparticle (30 to 70 nm in diameter) [32] consists of a few nanometers of disordered carbon or of polygonized graphene shells.

1.2.4.4. Carbon Onions Upon intense electron beam irradiation of carbon nanoparticles, hollow concentric carbon spheres (spherical multishells) can be formed with faceted shapes [33]. An innermost sphere with an inner diameter of 0.71 nm corresponds to C_{60}. This kind of structure can be synthesized up to 10 nm. The inner spheres contain no dangling bonds and are stable under further electron bombardment [4]. Carbon onions can serve as nanoscopic pressure cells for diamond formation when they are heated to 700°C and irradiated with electrons [4].

1.2.4.5. Porous Carbons Porous carbons can be macropores (larger than 100 nm in diameter), mesopores (2 to 100 nm in diameter), or micropore (less than 2 nm in diameter) [34]. They have very high surface areas and pores of nanometer size, similar to those of fullerenes, and may be in the form of cages or tunnels. Although the surfaces of the carbon nanopores contain a high density of dangling bonds and surface states, the surfaces of fullerenes have no dangling bonds [4].

1.2.5. Rod-Shaped Carbon Nanostructures

The study of rod-shaped carbon nanostructures was greatly stimulated by the discovery of the existence of carbon tubules or nanotubes [35] and the subsequent report of the synthesis of large quantities of carbon nanotubes [22, 36–56]. Generally speaking, single-walled carbon nanotubes (SWNT) can be made if only one graphene sheet is rolled, and multiwalled carbon nanotubes (MWNT) can be obtained if a few stacked graphitic shells are built from perfectly concentric cylinders or a single graphene sheet is rolled as a scroll. Theoretically, there are three major classifications of C_{60}-based tubules or nanotubes [22, 57–61]: armchair carbon nanotube (see Fig. 6a), zigzag carbon nanotube (see Fig. 6b), and chiral carbon nanotubes (see Fig. 6c), depending on whether they are related to a fivefold, threefold, or twofold axis relative to the C_{60} molecule. First we consider the tube formed along a fivefold axis, which is the easiest to visualize and can be represented by the formula $C_{60+i\times10}$, where i is a positive integer. We can think of this tubule as follows: by cutting a C_{60} molecule into two parts along its equatorial line and then inserting one row of five armchair hexagons, one obtains C_{70}. More generally, adding i rows of armchair hexagons, one then obtains a $C_{60+i\times10}$ molecule (armchair tubule [22, 57–61]). Figure 6a shows an example of an armchair nanotube. Closely related to the armchair tube based on a fivefold axis is $C_{60+i\times18}$, which is based on a threefold axis. In detail, it is formed as follows: by cutting C_{60} into two parts

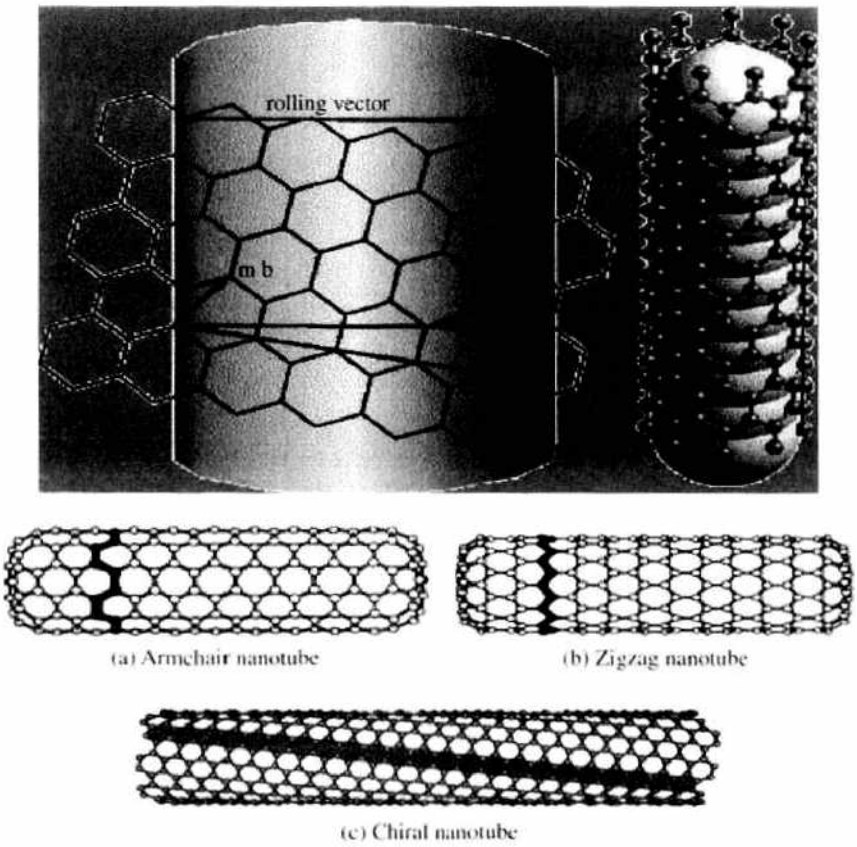


Figure 6. A schematic model for a single-wall carbon nanotube with the nanotube axis normal to (a) the direction $\phi = \pi/6$ (an armchair nanotube), (b) the direction $\phi = 0$ (a zigzag nanotube), and (c) a general direction $0 < \phi < \pi/6$ (a chiral nanotube). A chiral vector $\mathbf{T}_h = m\mathbf{b} + n\mathbf{a}$ is defined on the honeycomb lattice of carbon atoms by the unit vectors **a** and **b** and the chiral angle ϕ with respect to the zigzag axis **a** ($\phi = 0$).

along the zigzag edges and inserting i rows of nine zigzag hexagons, one gets a $C_{60+i\times 18}$ molecule (zigzag tubule [22, 57–61]). Figure 6b shows an example of a zigzag tubule. These tubules are of scientific interest as carbon fibers, which are today commercially important for their extraordinary high modulus and strength. In addition to the armchair and zigzag tubules, a large number of chiral carbon nanotubes, as shown in Fig. 6c, can be formed with a screw axis along the axis of the tubule and with a variety of "hemispherical"-like caps. These general carbon nanotubes can be specified mathematically in terms of the tubule diameter d_t and chiral angle ϕ (see Fig. 6), where a chiral vector (rolling vector) $\mathbf{T}_h$ [22]

$$\mathbf{T}_h = n\mathbf{a} + m\mathbf{b} \tag{1}$$

is shown. The vector $\mathbf{T}_h$ connects two crystallographically equivalent sites on a two-dimensional graphene structure. The construction in Fig. 6 shows the chiral angle ϕ of the tubule with respect to the zigzag direction **a** ($\phi = 0$) and two units, **a** and **b**, of the hexagonal honeycomb lattice. An ensemble of possible chiral vectors can be specified by $\mathbf{T}_h$ in terms of pairs of integers (m, n) [22, 57–61]. Each pair of integers (m, n) defines a different way of rolling the graphene sheet to form a carbon nanotube. In detail, the cylinder connecting the two hemispherical caps of Fig. 6 is formed by superimposing the two ends of the vector $\mathbf{T}_h$. The cylinder joint shown in Fig. 6 is made by joining the vertical and parallel lines. The chiral tubule thus generated has no distortion of bond angles other than distortions caused by the cylindrical curvature of the tubule. Differences in chiral angle ϕ and in the tubule diameter d_t give rise to differences in the properties of the various carbon nanotubes. In the (m, n) notation for specifying the chiral vector $\mathbf{T}_h$, the vectors $(n, 0)$ denote zigzag tubules, the vectors (m, m) denote armchair tubules, and all other vectors (n, m) correspond to chiral tubules [22, 57–61]. Since both right- and left-handed chirality is possible for chiral tubules, it is expected that chiral tubules are optically active to either right or left circularly polarized

light propagating along the tubule axis. In terms of the integers (n, m), the tubule diameter d_t is given by [22]

$$d_t = T_h/\pi = \frac{\sqrt{3(m^2 + mn + n^2)}}{\pi} a_{c-c} \tag{2}$$

where a_{c-c} is the nearest-neighbor c–c distance (= 1.421 Å in graphite [22]), T_h is the length of the chiral vector $\mathbf{T}_h$, and the chiral angle ϕ is given by [22]

$$\phi = \tan^{-1}\left(\frac{\sqrt{3}m}{m + 2n}\right) \tag{3}$$

For example, the armchair tubule specified by (5,5) in Fig. 6a has $d_t = 6.88$ Å and $\phi = \pi/6$, the zigzag tubule specified by (9,0) in Fig. 6b has a theoretical tubule diameter of $d_t = 7.15$ Å and $\phi = 0$, and the chiral tubule specified by (10,5) in Fig. 6c has $d_t = 10.36$ Å and $\phi = 0.7137$, all derived from hemispherical caps for the C_{60} molecule and assuming an average $a_{c-c} = 1.44$ Å appropriate for C_{60}. Let q be the largest common divisor in m and n. Then, the atom number N per unit cell is equal to [57–61]

$$N = \frac{4(m^2 + mn + n^2)}{q} \tag{4}$$

if $(m - n)$ is not a multiple of $3q$ or [57–61]

$$N = \frac{4(m^2 + mn + n^2)}{3q} \tag{5}$$

if $(m - n)$ is a multiple of $3q$. Similar to fullerenes, carbon nanotubes can be doped inside, on, and outside the surface.

Since 1991, carbon nanotubes have been extensively studied, both theoretically and experimentally, because of their unique geometric structures and remarkable mechanical, chemical, electronic, magnetic, and transport properties. The small diameter (at scale of nanometers) and the long length (at the order of microns) lead to such large aspect ratios that the carbon nanotubes act as ideal one-dimensional systems. Their rope crystallites offer host lattice for intercalation and storage. All these characteristics make carbon nanotubes the focus of extensive studies in nanoscale science and technology, with potential applications in various materials and devices. The possibility of filling the hollow cores of carbon nanotubes with selected metals opens up exciting possibilities with regard to the physics of low-dimensional transport, magnetism, and superconductivity [4]. The possibility of connecting carbon nanotubes of different diameters and chirality has generated considerable interest because these nanotube junctions can serve as potential building blocks for nanoscale electronic devices [4]. Nanostructured template channels have been used to grow individual Y-junction carbon nanotube heterostructures [62]. In summary, carbon nanotubes and their derivatives have been recognized as a fascinating material about to trigger a revolution in nanodevices, optical computing, optical communication, carbon chemistry, and new functional structural materials. Readers can find more details about the synthesis, physical and chemical properties, and technological applications of carbon nanotubes in recent books [22, 63–70] and review articles [71–77].

1.2.6. Other Carbon Nanostructures

Other carbon nanostructures include (1) carbon nanoparallelepipeds [78], produced from arc evaporation of carbon with alkaline earth metals and containing 5 to 20 layers of multiwalled graphitic carbon and 20 to 100 nm edge length; (2) rings of single-walled carbon nanotubes [79] or multiwalled carbon nanotubes [80]; (3) graphitic nanocones such as disks (no pentagons), five types of cones (one to five pentagons), and open tubes (six pentagons) produced by pyrolyzing a continuous flow of heavy oil inside an industrial-grade carbon arc plasma generator [81]; (4) carbon nanohorns [82]; (5) carbon beaded cone microstructures [83]; and (6) carbon microtrees [84].

1.3. Outline

Recently, Bianchetti et al. [85] presented and reviewed the most complete set of calculations to date of the ground-state electronic properties and of the linear optical response of linear carbon nanostructures C_N using *ab initio* methods based on local density and on time-dependent local density approximations. In this chapter, we focus on reviewing recent studies on the second-order optical nonlinearities (characterized by the first-order hyperpolarizability β) and the third-order optical nonlinearities (characterized by the second-order hyperpolarizability γ) of carbon nanostructures, mainly spherically and rod-shaped structures (note: until now, there is no study on the nonlinear optical (NLO) properties of carbon onions, carbon nanoparticles, carbon nanoparallelepipeds, *s* of single-walled carbon nanotubes, graphitic nanocones, carbon nanohorns, carbon-beaded cone microstructures, and carbon microtrees), in Sections 3 and 4, respectively. For further understanding of the contents presented in Sections 3 and 4, we introduce in Section 2 the nonlinear optical processes and several theoretical tools related to our discussions.

On the other hand, a lot of laser sources—for example, dye lasers and Raman lasers—are widely used in many areas of science and technology. Naturally, a potential hazard for human eyes and other optical sensors arises because of the laser power. Hence, protecting all sensors and human eyes against the hazard is a very important task that has stimulated many research groups to search for good optical limiters. This kind of device is designed to keep the power, irradiance, energy, or fluence transmitted by an optical system below some specified maximum value, regardless of the magnitude of the input. Also, it must do this while maintaining high transmittance at low input powers. In general, there are two obvious ways. One achieves optical limiting by active control. In this way, a sensor is used to monitor input light levels, activating through some processor a modulator or shutter that in turn limits the transmitted light. The iris and blink response of our human eyes are good examples. However, the response speed in this way is limited to about 0.1 seconds, and thus any intense pulse of lasers shorter than 0.1 second would get through our defense systems of eyes and damage our retinas before the eyes are able to make a response. Another way to realize optical limiting uses narrow-line spectral filters. In this way, as long as the laser wavelength is known—for example, our laboratory laser safety goggles—the limiting works very well. However, it would not be effective against a tunable laser. To our delight, we can develop passive systems that use NLO materials. These NLO materials function as combined sensors, processor and modulator. Obviously, this would lead to potential optical limiters with high speed, simplicity, compactness, and low cost. Many materials exhibit these types of effects that produce optical limiting, but these effects are usually too small for practical optical limiting applications. Hence, research to date on optical limiting has mainly focused on the search for new or modified materials with stronger nonlinearities and on how to optimally use the best NLO materials available. Extensive studies have shown that carbon nanostructures would be good candidates for optical limiters. Therefore, Section 5 reviews recent developments in optical limiting studies of carbon nanostructures.

2. NONLINEAR OPTICAL INTERACTIONS AND THEORETICAL TOOLS

2.1. Nonlinear Optical Interactions

In the decades since the invention of the laser and the first observations of NLO phenomena by Franken et al. [86], interest in this topic has continuously increased, and the field of nonlinear optics was born. It was recognized that many NLO effects—for example, frequency doubling and tripling—stimulated Raman scattering, and sum- and difference-frequency mixing can have useful practical applications in information and telecommunication technologies [87–101], such as all-optical switching, modulating and computing devices, optical bistable devices, and optical limiters. To introduce the concept of optical nonlinearity, one can express the field-induced dipole moment per unit volume, that is, the polarization $\mathbf{P}(t)$, in a power series of the strength of the applied optical field $\mathbf{F}(t)$ [95]. Assuming that this optical field

vector can be represented as the discrete sum of a number of frequency components as the compact form [95]

$$\mathbf{F}(t) = \sum_n \mathbf{F}(\omega_n) e^{-i\omega_n t} \tag{6}$$

one can write the polarization $\mathbf{P}(t)$ as a function of $F_i(\omega_n)$, that is, the Fourier components of $\mathbf{F}(t)$, in the Gaussian system of units

$$\mathbf{P}(t) = \sum_{m=1} \sum_n \sum_{j=x,y,z} P_j^{(m)}(\omega_n) e^{-i\omega_n t} \tag{7}$$

with

$$P_i^{(1)}(\omega_n) = \sum_{j=x,y,z} \chi_{ij}^{(1)} F_j(\omega_n) \tag{8}$$

$$P_i^{(2)}(\omega_n) = \sum_{j,k=x,y,z} \sum_{(mo)} \chi_{ijk}^{(2)}(\omega_n = \omega_m + \omega_o; \omega_m, \omega_o) F_j(\omega_m) F_k(\omega_o) \tag{9}$$

$$P_i^{(3)}(\omega_n) = \sum_{j,k,l=x,y,z} \sum_{(mop)} \chi_{ijkl}^{(3)}(\omega_n = \omega_m + \omega_o + \omega_p; \omega_m, \omega_o, \omega_p) \cdot F_j(\omega_m) F_k(\omega_o) F_l(\omega_p) \tag{10}$$

where the indices $ijkl$ refer to the Cartesian components of the applied optical field, the notation (mo) or (mop) indicates that the sum $\omega_m + \omega_o$ or $\omega_m + \omega_o + \omega_p$ is to be held fixed in performing the summation over m, o, and p. The coefficients $\chi^{(1)}$, $\chi^{(2)}$, and $\chi^{(3)}$ are known as the linear optical and second- and third-order nonlinear optical susceptibilities of the material system.

Microscopically, the polarization of a material such as a molecular crystal is attributed to the molecular polarization. The molecular polarizabilities are defined as the expansion coefficients of the molecular polarization in terms of the power series of the external field. Two types of definition are used in general: one type defined in terms of a Taylor expansion of the molecular polarization induced by the external field, and another type defined in terms of a perturbation expansion. We adopt the later definition because the susceptibilities are defined similarly. That is, for atoms or molecules they refer to the linear polarizability α $(= \chi^{(1)}/\rho)$, the first-order hyperpolarizability β $(= \chi^{(2)}/\rho)$, and the second-order hyperpolarizability γ $(= \chi^{(3)}/\rho)$, respectively, where ρ is the number density of molecules in a material. Assuming that the molecule-field coupling system is initially prepared in the gth state (for example, ground state), we can derive α, β, and γ by using a perturbation theory [95]. The linear optical polarizability $\alpha(\omega_p)$ of molecule at optical frequency ω_p can be [24]

$$\alpha_{ij}(\omega_p) = \hbar^{-1} \sum_m \left\{ \frac{U_{gm}^i U_{mg}^j}{\omega_{mg} - \omega_p} + \frac{U_{gm}^j U_{mg}^i}{\omega_{mg}^* + \omega_p} \right\} \tag{11}$$

where U_{mn} is the transition-dipole matrix element between the nth and mth states and ω_{mn} is the corresponding transition frequency. The first and second terms in this equation can be interpreted as the resonant and antiresonant contributions to the linear optical polarizability. $\beta(\omega_\sigma)$ of molecule at optical frequency $\omega_\sigma = \omega_q + \omega_p$ is given by [24]

$$\beta_{ijk}(\omega_\sigma; \omega_q, \omega_p) = \hbar^{-2} \Theta \sum_{mn} \left\{ \frac{U_{gn}^i U_{nm}^j U_{mg}^k}{(\omega_{ng} - \omega_p - \omega_q)(\omega_{mg} - \omega_p)} + \frac{U_{gn}^j U_{nm}^i U_{mg}^k}{(\omega_{ng}^* + \omega_q)(\omega_{mg} - \omega_p)} + \frac{U_{gn}^j U_{nm}^k U_{mg}^i}{(\omega_{ng}^* + \omega_q)(\omega_{mg}^* + \omega_p + \omega_q)} \right\} \tag{12}$$

where Θ denotes the intrinsic permutation operator (this operator tells us to average the expression that follows it over both permutations of frequencies ω_p and ω_q of the applied

field. The Cartesian indices j and k should be permuted simultaneously). The second-order hyperpolarizability $\gamma(\omega_\sigma)$ of molecule at $\omega_\sigma = \omega_r + \omega_q + \omega_p$ can be [24]

$$\gamma_{ijkl}(\omega_\sigma; \omega_r, \omega_q, \omega_p) = h^{-3}\Theta \sum_{nmg} \left\{ \frac{U_{gn}^k U_{nm}^j U_{mg}^i U_{mg}^h}{(\omega_{ng} - \omega_r - \omega_q - \omega_p)(\omega_{mg} - \omega_p - \omega_q)(\omega_{mg} - \omega_p)} \right.$$
$$+ \frac{U_{gn}^j U_{nm}^k U_{mg}^i U_{mg}^h}{(\omega_{ng}^* + \omega_r)(\omega_{ng} - \omega_q - \omega_p)(\omega_{mg} - \omega_p)}$$
$$+ \frac{U_{gn}^j U_{nm}^i U_{mg}^k U_{mg}^h}{(\omega_{ng}^* + \omega_r)(\omega_{mg}^* + \omega_r + \omega_q)(\omega_{mg} - \omega_p)}$$
$$\left. + \frac{U_{gn}^j U_{nm}^i U_{mn}^h U_{mg}^k}{(\omega_{ng}^* + \omega_r)(\omega_{mg}^* + \omega_r + \omega_q)(\omega_{mg}^* + \omega_r + \omega_q + \omega_p)} \right\} \quad (13)$$

Available studies [87–90, 102] have shown that the second-order NLO interaction can occur only in noncentrosymmetric crystals that do not display inversion symmetry, whereas the third-order NLO interactions can occur for both centrosymmetric and noncentrosymmetric media. Assuming that the two input waves are at ω_n and ω_m, we can have from Eq. (9) second-harmonic generation (SHG), sum-frequency generation (SFG2), difference-frequency generation (DFG), and optical rectification (OR), respectively. See the examples in Table 1. If the nonlinear crystal used in a DFG process is placed inside an optical resonantor, the frequency of the output signal or idler field can build up to large values, resulting in the device of optical parametric oscillator (OPO), which is used primarily at infrared wavelengths, where other sources of tunable radiation are not readily available. Because the nonlinear polarization can efficiently produce an output signal only if a certain phase-matching condition is satisfied, and this condition cannot be satisfied for more than one frequency component of the nonlinear polarization, no more than one of these frequency components shown in Eq. (12) will be presented with any appreciable intensity in the radiation generated by the nonlinear optical interaction. Operationally, one often chooses which frequency component will be radiated by properly selecting the polarization of the input radiation and orientation of the nonlinear crystal.

Based on Eq. (10), we can get many third-order NLO processes, for example, third-harmonic generation (THG) if $\omega_m = \omega_n = \omega_p = \omega$. Other third-order polarizations can be reached by varying the three frequencies as shown in Table 1. (For more details, refer to standard textbooks about nonlinear optics, such as Refs. [87–101].) Here we briefly describe

Table 1. The second- and third-order nonlinear processes and the corresponding frequency terms arising from the nonlinear polarization.

Nonlinear Optical Process	Polarization $P^{(n)}(\omega_\sigma)$	Input Example
Second harmonic generation (SHG)	$P^{(2)}(\omega_\sigma = 2\omega_n)$	$\omega_n = \omega_m$
Sum-frequency generation (SFG2)	$P^{(2)}(\omega_q = \omega_n + \omega_m)$	$\omega_n \neq \omega_m$
Difference-frequency generation (DFG3)	$P^{(2)}(\omega_n = \omega_n - \omega_m)$	$\omega_n > \omega_m$
Optical rectification (OR)	$P^{(2)}(\omega_n = 0)$	$\omega_n = -\omega_m$
Third harmonic generation (THG)	$P^{(3)}(\omega_n = 3\omega_p)$	$\omega_p = \omega_n = \omega_m$
Sum-frequency generation (SFG3)	$P^{(3)}(\omega_n = 2\omega_n - \omega_m)$	$\omega_p = \omega_n \neq \omega_m$
Frequency mixing (FM)	$P^{(3)}(\omega_n = \omega_r + \omega_p + \omega_m)$	$\omega_p \neq \omega_n \neq \omega_m$
Parametric amplification (PA)	$P^{(3)}(\omega_n = 2\omega_n - \omega_m)$	$\omega_p = \omega_n \neq \omega_m$
Coherent stokes Raman scattering (CSRS)	$P^{(3)}(\omega_n = \omega_n + \omega_p - \omega_m)$	$\omega_p \neq \omega_n \neq \omega_m$
Coherent anti-stokes Raman scattering (CARS)	$P^{(3)}(\omega_n = -\omega_n - \omega_p + \omega_m)$	$\omega_p \neq \omega_n \neq \omega_m$
Bound-electronic optical effect (BEOKE), Raman induced Kerr effect (RIKE), molecular orientational Kerr effect (MOKE), two-photon absorption (TPA), AC Stark effect (ACSE), stimulating Raman scattering (SRS), stimulating Rayleigh-Wing scattering (SRWS)	$P^{(3)}(\omega_n = \omega_m)$	$\omega_p = \omega_n = 0 \neq \omega_m$

the process of degenerate four-wave mixing (DFWM). In the DFWM experiment, a polarization is induced in the sample by the interaction between three light waves and the electrons in the nonlinear medium characterized by a third-order nonlinear optical susceptibility $\chi^{(3)}$ through Eq. (10), but two wave beams are incident at the same region on the sample and have equal and opposite wave vectors and the same frequency ω. The two wave beams from, for example, an He-Ne laser, set up an interference grating, which interacts with the third light beam coming from the back side of a transparent sample. Then, the fourth beam, resulting from this interaction with wave vector equal and opposite to that of the third beam, is detected. So, if the frequency of the third wave is ω, the fourth wave will be generated with the same frequency ω, but its complex amplitude everywhere will be the complex conjugate of the third wave.

In nonlinear optics, there are several different systems of units [95] that are commonly used in the community. In this chapter, we use the Gaussian system of units. In the Gaussian system, both the field **E** and polarization **P** have the same units statvolt/cm. Thus, $\chi^{(1)}$ is dimensionless, and the dimensions of $\chi^{(2)}$ and $\chi^{(3)}$ are given by cm/statvolt ($\equiv esu = 3.3354 \times 10^{-2}$ m/V) cm^2/statvolt^2 ($\equiv esu = 1.1125 \times 10^{-11}$ m^2/V^2), respectively. The units of $\chi^{(2)}$ and $\chi^{(3)}$ are not usually stated explicitly in the Gaussian system of units, but instead quote values simply in electrostatic units (esu). For different systems of units, reader may refer to Appendix A of Boyd's book [95]. In addition, there are different conventions for defining hyperpolarizabilities, for example, Taylor series, perturbation series, electric field–induced SHG (EFISHG), phenomenological; see Ref. [103] for details.

Theoretical guidance is highly desirable prior to the synthesis of new NLO materials. Thus, to study theoretically the NLO properties of carbon nanostructures, it is necessary to calculate and understand the electronic structures of carbon clusters and nanostructures. There are diverse theoretical tools with varied degrees of sophistication for doing electronic structure calculations. Some of these tools are complicated and time-consuming—for example, very accurate *ab initio* calculations (but only on atoms or very small molecules)—and some do not even require a calculator, for example, semiempirical methods, which intrinsically account for electron correlation and are of interest in the NLO study of large molecules, although their accuracy has not been carefully tested. All in all, from the computational point of view, the trade-off between the cost of calculations and the accuracy of the computed data has been the most important consideration in making a final decision to choose one ideal theoretical tool to perform the computation task. For this, this section reviews several key tools for performing electronic structure and (hyper)polarizability calculations.

2.2. Born-Oppenheimer Approximation

The physical systems, such as atoms, molecules, and nanostructures, consist not only of electrons but also of nuclei, and each of these particles moves in the field generated by the others. The total Hamiltonian of such a system of n electrons and N nuclei can be written as follows [104]

$$H_{\text{total}} = \sum_{i=1}^{n} \frac{p_i^2}{2m} + \sum_{\alpha=1}^{N} \frac{P_\alpha^2}{2M_\alpha} + \frac{1}{8\pi\epsilon_0} \sum_{i=1, i\neq j}^{n} \sum_{j=1}^{n} \frac{e^2}{|\mathbf{r}_i - \mathbf{r}_j|} - \frac{1}{4\pi\epsilon_0} \sum_{\alpha=1}^{N} \sum_{i=1}^{n} \frac{Z_\alpha e^2}{|\mathbf{r}_i - \mathbf{R}_\alpha|} + \frac{1}{8\pi\epsilon_0} \sum_{\alpha=1, \alpha\neq\beta}^{N} \sum_{\beta=1}^{N} \frac{Z_\alpha Z_\beta e^2}{|\mathbf{R}_\alpha - \mathbf{R}_\beta|} \tag{14}$$

where the index i and j refer to the electrons and α and β to the nuclei, m is the electron mass, and M_α is the mass of nuclei α. This Hamiltonian consists of five terms: the first two terms represent the kinetic energy operators for the electrons and nuclei; the third term is the potential-energy (Coulomb repulsion) operator for electron–electron interactions; the fourth term shows the Coulomb attraction between electrons and nuclei; the last term is the potential-energy (Coulomb repulsion) for nucleus–nucleus interactions. Also, $\mathbf{r}_i$ and $\mathbf{R}_\alpha$ are the positions of the ith electrons and the αth nuclei, respectively, and Z_α is the charge of

the αth nuclei. This Hamiltonian looks quite complicated. If the number of electrons and nuclei is not extremely small, it is impossible to solve the stationary Schrödinger equation for this Hamiltonian directly on even the largest and fastest computer available [105]. A first approximation is to consider the nuclei as being fixed and to solve the Schrödinger equation for the electronic system in the field of the static nuclei. This approach is called the Born-Oppenheimer approximation [104] and is justified by the nuclei being much heavier than the electrons so that they move at much slower speeds. The Born-Oppenheimer Hamiltonian for the electrons reads [104]

$$H_e = \sum_{i=1}^{n} \frac{p_i^2}{2m} + \frac{1}{8\pi\epsilon_0} \sum_{i=1,\, i\neq j}^{n} \sum_{j=1}^{n} \frac{e^2}{|\mathbf{r}_i - \mathbf{r}_j|} - \frac{1}{4\pi\epsilon_0} \sum_{\alpha=1}^{N} \sum_{i=1}^{n} \frac{Z_\alpha e^2}{|\mathbf{r}_i - \mathbf{R}_\alpha|} \tag{15}$$

The total energy is the sum of the energy E_e of the electrons and the energy resulting from the following Schrödinger equation satisfied by the nuclei [104]:

$$H_n = \sum_{\alpha=1}^{N} \frac{P_\alpha^2}{2M_\alpha} + \frac{1}{8\pi\epsilon_0} \sum_{\alpha=1,\, \alpha\neq\beta}^{N} \sum_{\beta=1}^{N} \frac{Z_\alpha Z_\beta e^2}{|\mathbf{R}_\alpha - \mathbf{R}_\beta|} + E_e \tag{16}$$

In a further approximation, the motion of the nuclei is neglected, and only the electrostatic energy of the nuclei should be added to the energy of the electrons to arrive at the total energy. Thus, it remains to solve for the electron structure. The Born-Oppenheimer approximation factorizes the total wavefunction into the electronic and nuclear components and works extremely well in most examples, except for the systems that have degenerate or almost degenerate electronic energy levels. In the former case, a Jahn-Teller distortion ensues, breaking molecular symmetry so that the degeneracy is lifted. In the latter case, symmetry lowering is observed. In the following, we briefly review different approaches to the many-electron problem.

2.3. *Ab Initio* Methods and Basis Sets

Ab initio methods try to obtain information by solving the Schrödinger equation without fitting parameters to experimental data, while experimental data guide the selection of *ab initio* methods rather than directly entering the computational procedures. From the viewpoint of a theoretician, fullerenes have been challenging molecules in *ab initio* calculations because of their sizes [21, 106]. Recent advances in *ab initio* electronic structure methods and parallel computing have brought a substantial improvement in the capabilities of supercomputers to predict and study the properties of large molecules. The coupled cluster method [105] has been used to predict phenomena in C_{20} [107]. Other *ab initio* methods, which are less demanding in terms of computation cost than the coupled cluster method, have been used for much larger fullerenes and carbon nanotubes, for example, C_{60} [108, 109] with self-consistent field and Moller-Plesset second-order theory (MP2), C_{240} [110] and carbon nanotubes [111] with density functional theory (DFT), and C_{540} with the Hartree-Fock (HF) method [112]. It is known that the major expense in HF and DFT calculations arises from solving the electronic quantum Coulomb problem. The effective Hamiltonian diagonalization, which is a procedure that scales as N^3 (N is the number of basis functions), represents only a minor portion of the computational time in calculations of molecular clusters containing up to several hundred atoms [113]. DFT requires an additional three-dimensional numerical quadrature to obtain the exchange and correlation energies [114]. For clarity, we review both HF and DFT methods here.

On the other hand, many different approximations [105] exist for solving Schrödinger equations. One of the approximations inherent in all *ab initio* methods is the introduction of a basis set [105]. If the basis set is complete, expanding an unknown function—for example, a molecule orbital (MO)—in a set of known functions is not an approximation. However, a complete basis set means that an infinite number of functions must be used, which is impossible in actual calculations. One may think of an unknown MO as a function in an infinite coordinate system spanned by a complete basis set [105]. If a finite basis set is

used, only the parts of the MO along those coordinate axes that correspond to the selected basis can be represented. The smaller the basis set, the poorer the representation. Also the type of basis functions used in calculations influence the accuracy. The better a single basis function is able to reproduce the unknown function, the fewer basis functions are necessary for achieving a given level of accuracy. Because the computational effort of *ab initio* methods scales formally as at least N^4 [105], it is of importance to make the basis set as small as possible without compromising the accuracy. Hence, for convenience, we also briefly introduce several basis sets in this section.

2.3.1. Hartree-Fock Method

The Hartree-Fock method can be viewed as a variational method, where the wave functions of the many-electron system have the form of an antisymmetried product of one-electron wave functions. This restriction leads to an effective Schrödinger equation for the individual one-electron wave functions (also called *orbitals*) with a potential determined by the orbitals occupied by the other electrons. This coupling between the orbitals via the potentials causes the resulting equations to become nonlinear in the orbitals, and the solution must be found iteratively in a self-consistency procedure. The Hartree-Fock procedure is close in spirit to the mean-field approach in statistical mechanics.

The electron Hamiltonian H_e can be written as

$$H_e = \sum_{i=1}^{N} h_1(r_1) + \frac{1}{2} \sum_{i \neq j=1}^{N} h_2(r_i, r_j) \tag{17}$$

which contains the one-electron operator (one-electron core Hamiltonian) h_1 and the two-electron operator (electron repulsion integral) h_2

$$h_1 = \frac{p_1^2}{2m} - \frac{1}{4\pi\epsilon_0} \sum_{\alpha=1}^{N} \frac{Z_\alpha e^2}{|\mathbf{r}_1 - \mathbf{R}_\alpha|} \tag{18}$$

$$h_2 = \frac{1}{4\pi\epsilon_0} \frac{e^2}{|\mathbf{r}_i - \mathbf{r}_j|} \tag{19}$$

The total electronic energy E_e is given by

$$E_e = \sum_{k=1}^{N} \epsilon_k - \frac{1}{2} \sum_{k,l=1}^{N} [\langle \phi_k \phi_l | h_2 | \phi_k \phi_l \rangle - \langle \phi_l \phi_k | h_2 | \phi_k \phi_l \rangle] \tag{20}$$

where

$$\epsilon_k = \langle \phi_k | h_1 | \phi_k \rangle + \sum_{l=1}^{N} [\langle \phi_k \phi_l | h_2 | \phi_k \phi_l \rangle - \langle \phi_l \phi_k | h_2 | \phi_k \phi_l \rangle] \tag{21}$$

The eigenvalues ϵ_k appear accordingly as individual orbital energies whose sum gives one large contribution to the total electronic energy but that has to be modified by a double-counting term because of electron–electron interactions. All the individual orbitals ϕ_k are orthonormal; that is, $\langle \phi_i | \phi_j \rangle = \delta_{ij}$. Then, the total energy E_t is given by

$$E_t = E_e + \frac{1}{8\pi\epsilon_0} \sum_{\alpha=1, \alpha \neq \beta}^{N} \sum_{\beta=1}^{N} \frac{Z_\alpha Z_\beta e^2}{|\mathbf{R}_\alpha - \mathbf{R}_\beta|} \tag{22}$$

In actual Hartree-Fock calculations, the wavefunctions ϕ_i are expressed in terms of K basis functions Φ_ν ($\nu = 1, 2, \ldots, K$):

$$\phi_i = \sum_{\nu=1}^{K} C_{\nu,i} \Phi_\nu \tag{23}$$

where $K \geq N$. The optimization leads to a set of nonlinear equations:

$$\sum_{\nu}\left[h_1^{\mu,\nu} + \sum_{\lambda\sigma} D_{\lambda,\sigma}(\langle\phi_\mu\phi_\sigma|\phi_\nu\phi_\lambda\rangle - \langle\phi_\mu\phi_\sigma|\phi_\lambda\phi_\nu\rangle)\right]C_{\nu,k} = \epsilon_k \sum_{\nu} S_{\mu,\nu}C_{\nu,k} \tag{24}$$

where $h_1^{\mu,\nu} = \langle\phi_\mu|h_1|\phi_\nu\rangle$, the overlap integrals $S_{\mu,\nu} = \langle\phi_\mu|\phi_\nu\rangle$ are collectively known as the one-electron integral, and $D_{\mu,\nu} = \sum_{i=\text{occ}} C^*_{\nu,i}C_{\mu,i}$ are the elements of the density matrix.

In the Hartree-Fock (HF) method, correlations between the electrons are neglected to some extent. In particular, the Coulomb repulsion between the electrons is represented in an averaged way. However, the effective interaction, which is caused by the fact that the electrons are fermions (obeying Pauli's principle) and want to keep apart if they have the same spin, is accurately included in the HF approach.

Sometimes, for example in the case of hydrogen dissociation, the HF approximation yields unsatisfactory results. This is, of course, due to Coulomb correlations not taken into account in the HF formalism. Actually, there exists a systematic way [105] to improve on HF by constructing a many-electron state as a linear combination of Slater determinants, which are constructed from the ground state by excitation: the first being the HF ground state, the second one being the first excited state, and so on. The resulting energy will be lower than the HF ground state energy. This is a time-consuming procedure so that only a limited number of determinants can be taken into account for systems containing many electrons. This is the configuration interaction (CI) method [105]. For most systems, full CI is impossible because of the large number of Slater determinants needed, but it is sometimes possible to obtain an estimate for the full CI result by extrapolating results for larger and larger number of Slater determinants. Several other methods [105]—for example, the Moller-Plesset perturbation method—improve on the approximations made in the HF method. These methods have also been implemented into many computational programs, for example, Gaussian 98, NWchem, Qchem, and GAMESS.

2.3.2. Density Functional Theory

Today's density functional theory (DFT), based on a strategy of modeling electron correlation via a general functional of the electron density, starts with the two theorems of Hohenberg and Kohn [115] for searching the electronic ground state of an isolated system of N interacting electrons in an external potential: the first theorem establishes that the external potential is a functional of the charge density, and the second one establishes the energy variational principle for the density. These two theorems show that the problem of solving the many-body Schrödinger equation for the ground state can be exactly recast into the variational problem of minimizing the Hohenberg-Kohn functional with respect to the ground-state electron density. Compared with conventional quantum-chemistry methods, the DFT approach is particularly appealing because it does not rely on a complete knowledge of the N-electron wave functions but only on the ground-state electron density. Of course, although the DFT theory is exact, the energy functional is unknown and has to be approximated in practical implementations. The available approximations for kinetic energy and exchange-correlation density-based functionals give moderate quantitative agreement with experimental data. A much improved strategy has been presented by Kohn and Sham [116] in terms of a set of self-consistent Kohn-Sham equations, and the total electronic energy is given by

$$E_{\text{total}} = E_{\text{k}} + E_{\text{V}} + E_{\text{ee}} + E_{\text{xc}} \tag{25}$$

The first three terms are exactly the same as those of Hartree-Fock, namely, the kinetic energy E_{k}, the energy E_{V} of the electrostatic attraction between the nuclear and electron and the repulsion between pairs of nuclei, and the electrostatic energy E_{ee} of the electron in the field generated by the total electron density. The last term E_{xc} contains the many-body effects, lumped together in an exchange-correlational potential, including the remaining part of electron–electron interactions. Hohenberg and Kohn [115] demonstrated that E_{xc} is determined entirely by the electron density. In practice, E_{xc} is usually approximated as an integral involving only the spin densities and possibly their gradients. Usually, E_{xc} is separated into

two parts: a pure exchange E_x and a correlational part E_c corresponding to same-spin and mixed-spin interactions. The functionals defining the exchange and correlation parts of E_{xc} are termed as exchange functionals and correlation functionals. There are two distinct functional types. One is the local functionals that depend on only the electron density. Another one is gradient-corrected functionals that depend on both the electron density and its gradient. The related methods are usually termed as LDA (local density approximation) and GGA (generalized gradient approximation). The local exchange functional is virtually always defined as [105]

$$E_{\mathrm{LDA}}^{X} = -\frac{3}{2}\left(\frac{3}{4\pi}\right)^{1/3} \int \rho^{4/3} d^3\mathbf{r} \tag{26}$$

Becke formulated the following gradient-corrected exchange functional based on the LDA exchange functional [117]

$$E_{\mathrm{Becke88}}^{X} = E_{\mathrm{LDA}}^{X} - \gamma \int \frac{\rho^{4/3}\chi^2}{(1 + 6\gamma\chi \sinh^{-1}\chi)} d^3\mathbf{r} \tag{27}$$

where $\chi = \rho^{-4/3}|\Delta\rho|$, and γ is a parameter chosen to fit the known exchange energies of the inert gas atoms and Becke defines its value as 0.0042 Hartrees. Similarly, there are local and gradient-corrected correlation functionals, for example, Perdew and Wang's formulation of the local part of their 1991 correlation functional [118], and Vosko, Wilk, and Nusair's local correlation functional [119]. Pure DFT methods are defined by pairing an exchange functional with a correlation functional. For example, the well-known BLYP functional pairs Becke's gradient-corrected exchange functional [117] with the gradient-corrected correlation functional of Lee, Yang, and Parr [120].

Becke has formulated functionals that include a mixture of Hartree-Fock and DFT exchange, along with DFT correlation, conceptually defining E_{xc} as

$$E_{xc}^{\mathrm{hybrid}} = C_{\mathrm{HF}} E_x^{\mathrm{HF}} + C_{\mathrm{DFT}} E_{xc}^{\mathrm{DFT}} \tag{28}$$

where the Cs are constants. It is termed a hybrid method. In this chapter, we consider B3LYP case [121], that is, a Becke-style three-parameter functional

$$E_{xc}^{\mathrm{B3LYP}} = AE_x^{\mathrm{slater}} + (1 - A)E_x^{\mathrm{HF}} + B\Delta E_x^{\mathrm{Becke88}} + E_c^{\mathrm{VWN3}} + C(E_c^{\mathrm{LYP}} - E_c^{\mathrm{vwn3}}) \tag{29}$$

In this hybrid functional, Hartree-Fock (exact) and Slater local exchanges [122] are included. In addition, Becke's gradient correction to LDA exchange is included, scaled by the parameter B. Similarly, the VWN3 local correlation functional is used, and it may be optionally corrected via the parameter C. In the B3LYP functional, the parameter's values are those specified by Becke, which he determined by fitting to the atomization energies, ionization potentials, proton affinities, and first-row atomic energies in the G_1 molecule set [121]. Different functionals can be constructed in the same way by varying the component functionals.

Many programs (for example, Gaussian 98, NWchem, Qchem, ABINIT, Crystal 98, and PWSCF) have implemented various pure DFT models given by combining the names for the exchange, correlation, and hybrid functionals, for example, HFB, PW91, VWN5, B3LYP, B3P86, B3P86, B1LYP, BP86, BLYP, BPW91, and PW91LYP.

In both DFT and HF theory, the electrons move in a background composed of the Hartree and external potentials. In addition to this, the exchange term in Hartree-Fock accounts for the fact that electrons with parallel spin avoid each other as a result of the exclusion principle (exchange hole), whereas the opposite spin pairs do not feel this interaction. In DFT, the exchange correlation potential includes not only exchange effects but also correlation effects due to the Coulomb repulsion between the electrons (dynamic correlation effects). In Hartree-Fock theory, the exchange interaction is treated exactly, but dynamic correlations are neglected. In principle, DFT is exact, but we do not know the exact form of the exchange correlation potential; that is, both exchange and dynamic correlation effects are in practice treated approximately.

2.3.3. Time-Dependent Density Functional Theory

For the calculation of the ground-state properties of many interacting electron systems, DFT is a popular and efficient method. Currently, there is a great deal of interest in extending DFT to study excited-state properties of molecules, clusters, and nanostructures. There exist several extensions of the basic formalism. One such extension uses the time-dependent formalism that provides a promising way to calculate the frequency-dependent response functions, for example, polarizability. The excitation energies are characterized as the poles of the frequency-dependent response functions. In detail, the time-dependent Kohn-Sham equation can be obtained, assuming the existence of a potential $V_{\text{eff}}(\mathbf{r}, t)$ for an independent particle system, whose orbitals $\Psi(\mathbf{r}, t)$ leads to the same charge density $\rho(\mathbf{r}, t)$ as for the interacting system:

$$\left[-\frac{1}{2}\nabla^2 + V_{\text{eff}}(\mathbf{r}, t)\right]\Psi(\mathbf{r}, t) = i\frac{\partial\Psi(\mathbf{r}, t)}{\partial t} \tag{30}$$

The potential has the following form

$$V_{\text{eff}}(\mathbf{r}, t) = V(t) + V_{\text{scf}}(\mathbf{r}, t) \tag{31}$$

where $V(t)$ is an applied field turned on slowly in the distant past and the self-consistent field (SCF) $V_{\text{scf}}(\mathbf{r}, t)$ is defined as

$$V_{\text{scf}}(\mathbf{r}, t) = \int \frac{\rho(\mathbf{r}, t)}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' + V_{\text{xc}}(\mathbf{r}, t) \tag{32}$$

where the exchange-correlation potential $V_{\text{xc}}(\mathbf{r}, t)$ is given as the functional derivative of the exchange-correlation action A_{xc}, represented by

$$V_{\text{xc}}[\rho](\mathbf{r}, t) = \frac{\delta A_{\text{xc}}[\rho]}{\delta\rho(\mathbf{r}, t)} \approx \frac{\delta E_{\text{xc}}[\rho_t]}{\delta\rho_t(\mathbf{r})} = V_{\text{xc}}[\rho_t](\mathbf{r}) \tag{33}$$

Here the unknown functional A_{xc} of ρ over both space and time is approximated by E_{xc} (the exchange-correlation functional of time-independent (TD) Kohn-Sham theory introduced previously), which is a function ρ_t at fixed time t. This local approximation in time is usually referred to as the adiabatic approximation. Many applications of TD-DFT have been reported (see Ref. [123–134] and several reviews [135–140]). The linear optical response TD-DFT is a well-developed amd matured technique that currently has become a method of choice for computing excited states of a particle. Application for the TD-DFT for calculating nonlinear optical responses is also reported [141–144].

The state-of-the-art TD-DFT calculations scale formally as $O(N^3)$, where N is the number of atoms involved. This makes TD-DFT a relatively expensive numerical method that cannot be employed to calculate the properties of very large systems. It is thus desirable to have linear-scaling TD-DFT whose computational time scales as $O(N)$. Yam, Yokojima, and Chen [145] have recently developed a linear-scaling first-principles TD-DFT method to evaluate the linear optical response of large molecular systems. Instead of a many-body wave function, the equation of motion is solved for the reduced single-electron density matrix in the time domain. The locality of the reduced single-electron density matrix is used to ensure that computational time scales linearly with system size. The two-electron Coulomb integrals are evaluated with the fast multipole method, and the calculation of exchange-correlation quadratures uses the locality of an exchange-correlation functional and the integral prescreening technique.

The tight-binding approach to TD-DFT (TD-DFT-TB) developed by Niehaus et al. [146] follows the TD-DFT route of Casida [135, 136] and uses a γ-approximation. In this approximation, the coupling matrix used in TD-DFT is treated by decomposing the transition density between different orbitals into atom-centered contributions. Because no integral evaluations need to be done, this scheme is numerically as efficient as the semiempirical tight-binding methods. On the other hand, a minimal basis set is used in the TD-DFT-TB. Thus, the optical properties for large systems, such as nanosystems, biological systems, and polymers with

hundreds of atoms, can be investigated easily. Especially for the calculations of the lowest excitation energies such as the optical gaps, one can obtain results at a highly reduced cost because the size of the coupling matrix is drastically reduced [146–150]. Application of the TD-DFT-TB to studying the NLO responses of large systems would be very interesting and is expected in the near future.

2.3.4. Time-Dependent Hartree Fock

Before TD-DFT was developed, a common method that most people used was the TD-HF method [151, 152], which is also known as the random phase approximation. This method can be derived along the same lines as TD-DFT. See the details in Refs. [151, 152]. TD-HF leads to an approximate treatment of electronic excitation energies with no electron correlation effects. However, electron correlation must be included in any predictive theory for an accurate treatment of excited states of molecules, clusters, and nanostructures.

2.3.5. Basis Sets

In the *ab initio* calculation of the electronic structure of a realistic system, one usually expands the single-electron orbital $|\phi_k\rangle$ in a set of predefined basis functions $|\chi_i\rangle$, known as a basis set, that is,

$$|\phi_k\rangle = \sum_{i=1}^{n} C_{ik}|\chi_i\rangle \tag{34}$$

which approximates the total electronic wavefunctions within the system. Here we do not discuss how these basis functions are chosen. There are two types of basis functions commonly used in electronic structure calculations: Slater type orbital (STO) [153] and Gaussian type orbital (GTO, also known as Gaussians) [154]. STOs have the functional form

$$|\chi_{i=(\zeta,n,l,m)}\rangle \equiv N Y_{l,m}(\theta,\phi) r^{n-1} e^{-\zeta r} \tag{35}$$

where N is a normalization constant and $Y_{l,m}$ is the usual spherical harmonic functions. The STOs are primarily used for atomic and diatomic systems where high accuracy is required and in semiempirical methods where all three- and four-center integrals are neglected. Primitive GTOs (PGTOs) can be written as

$$|\chi^{\mathrm{PGTO}}_{i=(\zeta,n,l,m)}\rangle \equiv N Y_{l,m}(\theta,\phi) r^{2(n-1)-l} e^{-\zeta r^2} \tag{36}$$

in term of polar coordinates or

$$|\chi^{\mathrm{PGTO}}_{i=(\zeta,l_x,l_y,l_z)}\rangle \equiv N x^{l_x} y^{l_y} z^{l_z} e^{-\zeta r^2} \tag{37}$$

in term of Cartesian coordinates, where the sum of l_x, l_y, and l_z determines the type of orbital. Both STOs and GTOs can be chosen to form a complete basis. However, the basis functions of choices in *ab initio* calculations have been GTOs because simple analytic formulas exist for evaluating the two-electron integrals that appear in the Coulomb problem [154, 155]. Because the inner (core) electrons require a relatively large number of functions for representing the wavefunction cusp near the nucleus and are largely independent of the environment, contracted GTOs (CGTOs) are especially useful for orbitals describing those electrons. The CGTOs are formed by combining a full set of PGTOs, that is,

$$|\chi_i\rangle_{CGTO} = \sum_j a_j |\chi_j^{\mathrm{PGTO}}\rangle \tag{38}$$

Contracting a basis set always increases the energy because it is a restriction of the number of variational parameters. In the following, we briefly present the classification of basis sets and introduce the Pople-style basis sets [156–160].

2.3.5.1. Minimum Basis Set The smallest number of functions is a minimum basis set. Only enough functions are used to contain all the electrons of the neutral atoms(s). For example, three s-functions (1s, 2s, 3s) and two sets of p-functions (2p and 3p) are used.

2.3.5.2. Double Zeta (DZ) Basis The improvement in the basis set is a doubling of all basis functions, which allows for a much better description of the fact that the electron distribution is different in different directions. The term *zeta* stems from the exponent ζ of STO basis functions. For example, six s-functions (1s, 1s′, 2s, 2s′, 3s, 3s′) and four p-functions (2p, 2p′, 3p, 3p′) are used.

2.3.5.3. Split Valence Basis A variation of the DZ type basis only doubles the number of the core orbitals.

2.3.5.4. Triple Zeta (TZ) Basis This basis contains three times as many functions as the minimum basis, namely, nine s-functions and six sets of p-functions for the second row elements.

2.3.5.5. Double Zeta Plus Polarization (DZP) Basis Adding a single set of polarization functions (p-functions on hydrogens and d-functions on heavy atoms) to the DZ basis forms a DZP type basis.

2.3.5.6. Pople-Style Basis Sets The STO-nG basis set is STOs consisting of n PGTOs and is a minimum type basis where the exponents of the PGTO are determined by fitting to the STO. STO-3G is a widely used minimum basis. The $k - nlm$G basis sets are of the split valence type, where k indicates how many PGTOs are used for representing the inner (core) orbitals, and nlm indicates how many functions the valence orbitals are split into and how may PGTOs are used for each function. Polarization functions are placed after G by using a star * on only heavy atoms or double stars ** on both heavy atoms and hydrogens. Diffuse functions are denoted by + (on only heavy atoms) or ++ (on both heavy atoms and hydrogens) before G. For example, 3-21G means that the core orbitals are a contraction of 3 PGTOs, the inner part of the valence orbitals is a contraction of 2 PGTOs, and the outer part of the valence orbitals is represented by 1 PGTO [157]. In addition, 6-31G is also a split-valence basis, where the core orbitals are a contraction of 6 PGTOs, the inner part of VOs is a contraction of 3 PGTOs, and the outer part of the VOs is represented by 1 PGTO [158]; 6-311G is a triple split valence basis, where the core orbitals are a contraction of 6 PGTOs and the VOs split into three functions, represented by 3, 1, and 1 PGTOs, respectively [159]; 6-31G* or 6-31G(d) is a split valence basis with a single d-type polarization function on heavy atom; 6-31+G* or 6-31+G(d) is a split valence basis with one set of diffuse sp-functions on heavy atoms only and a single d-type polarization function on heavy atoms [160].

2.4. Empirical Methods

2.4.1. Hückel Theory

In this theory, only the π electrons of conjugated unsaturated hydrocarbons are taken into account. The π-electron orbital energies are obtained by diagonalizing a Hückel matrix **H** [105]

$$\mathbf{H} = \beta\mathbf{A} + \alpha\mathbf{1} \tag{39}$$

where $A_{i,j} = 1$ if the ith and jth carbon atoms are linked through a bond and otherwise $A_{i,j} = 0$; α and β are negative and called the Coulomb and resonance integrals, respectively. This method is quick but crude, and explicit diagonalization of **A** is often possible because the adjacency matrix lends itself easily to mathematical analysis.

2.4.2. PPP Approximation

This method is based on two assumptions [161–163]. The first is σ-π separability, which allows one to construct a π-electron Hamiltonian, which is then expressed in the basis of p_z orbitals. The second assumption is the zero differential overlap (ZDO) approximation; that is, $\phi_\nu^* \phi_\mu \approx 0$ for $\nu \neq \mu$, applied to the electron repulsion integrals. The PPP method is usually used in conjunction with the configuration interaction formalism that involves excited determinants to calculate the $\pi \rightarrow \pi^*$ singlet and triplet excitation energies for planar conjugated

hydrocarbons and heterocyclic compounds. The PPP method using curvature-corrected resonance integrals was also applied successfully to three-dimensional carbon systems, including fullerenes.

2.4.3. Hubbard Approximation

This approximation [164] is made by setting all electron repulsion integrals to zero in the PPP method. In other words, only the one-center repulsion integrals are kept and are all equated to a parameter U, which is often called the on-site electron repulsion. All resonance integrals are set equal to a single parameter W, which is usually called the hopping integral. Sometimes, the first-neighbor two-electron integrals are retained and set to V, which is called the nearest-neighbor repulsion. The three parameters U, W, and V do not refer specifically to the p_z orbitals but are empirical parameters that describe interactions between sites in solids. Each site, which can be a single atom, an ion, or an entire molecule, is presumed to contribute one electron to the system under consideration. The advantages of using Hubbard approximation result from the fact that it can be diagonalized in many cases.

2.4.4. Tight-Binding Approximation

A straightforward generalization of the Hückel theory to all valence electrons is provided by the extended Hückel theory [165]. In this method, an effective Hamiltonian matrix is established within the basis of all valence orbitals of a given molecule. It is an example of the all-valence-electron semiempirical method. In this formalism, the orbital energies and coefficients are obtained by solving the following equations

$$\sum_{\nu} h_1^{\mu,\nu} C_{\nu,i} = \epsilon_i \sum_{\nu} S_{\mu,\nu} C_{\nu,i} \tag{40}$$

where $h_1^{\mu,\nu}$ are usually approximated by [166]

$$h_1^{\mu,\nu} = \frac{K}{2}(h_1^{\mu,\mu} + h_1^{\nu,\nu}) S_{\mu,\nu} \tag{41}$$

where $1 < K < 3$, and $h_1^{\mu,\mu}$ are respective ionization potentials. For very large systems with little or no symmetry, even this method is also too expensive. In the tight-binding (TB) Hamiltonian approximation [167], the extended Hückel matrix elements between basis functions residing on atoms that are not directly bonded are set to zero. The TB approach offers substantial savings in computational effort and is able to produce surprisingly accurate results, provided the remaining matrix elements are properly parameterized. A set of TB Hamiltonian parameters for carbon has been established by Xu et al. [168] and used with encouraging results in a variety of calculations on fullerenes.

2.4.5. CNDO/S Methods

Application of the ZDO approximation to all valence orbitals results in a variety of semiempirical methods. The most drastic variant of the ZDO approximation is the complete neglect of differential overlap (CNDO). In this formalism, the ZDO approximation is applied to all electron repulsion integrals, including those corresponding to pairs of orbitals residing on the same nucleus. The CNDO for spectroscopy (CNDO/S) method is an all-valence analog of the PPP formalism. In this approach, resonance integrals are calculated within the context of local coordinate systems. However, overlaps between the locally defined π orbitals are multiplied by $0.5k$ with $k = 0.585$ instead of 0.5. CNDO/S calculations are usually carried out in conjunction with a (limited) CI scheme involving singly excited determinants. Unfortunately, the original CNDO/S method leads to vanishing transition moments and singlet-triplet energy splittings for the $\sigma \rightarrow \pi^*$ and $n \rightarrow \pi^*$ excitations. Although the former deficiency can be easily remedied, the latter one cannot be eliminated without giving up the CNDO approximation.

2.4.6. INDO Method

In the intermediate neglect of differential overlap (INDO) approximation [105], the ZDO approximation is not applied to one-center electron repulsion integrals. INDO/S can yield nonvanishing singlet-triplet energy splittings for all excited states and has been successfully applied to describe fullerenes.

2.4.7. MNDO, AM1, and PM3 Methods

The neglect of diatomic differential overlap (NDDO) [105] constitutes one of the least severe ZDO-like approximations. In this formalism, only the electron repulsion integrals involving pairs of orbitals centered at two different nuclei are neglected. The heavily parameterized variant of NDDO, known as modified neglect of differential overlap (MNDO) [105], has been a mainstay of empirical quantum chemistry. The crucial innovation of MNDO lies in the treatment two-center electron repulsion integrals, which are efficiently computed with the help of a parameterized multipole interaction model. The resonance integrals, the matrix elements, and the core–core repulsion energy are all calculated from parametric expressions. The parameters are obtained by fitting the standard enthalpies of formations, geometries, first ionization potentials, and dipole moments for a molecule.

However, MNDO parameterization is far from optimal. Attempts to palliate such a weakness of MNDO led to a new parameterization called Austin Model 1 (AM1) [105]. In this method, the semiempirical formulas for the effective Fock matrix elements are the same as in MNDO, but the expressions for the core–core repulsions are more elaborate, giving rise to a substantially increased number of adjustable parameters.

The MNDO parametric method 3 (MNDO-PM3, or simply PM3) [105] stems from a massive reparameterization. PM3 parameters for H, C, N, O, F, Al, Si, P, S, Cl, Br, and I atoms have been evaluated.

2.5. Molecular Mechanics

Although these empirical methods, for example, PPP, are able to predict bond lengths in conjugated hydrocarbons with reasonable accuracy, full geometry optimizations that include bond and dihedral angles are possible only when the σ–electron interactions are taken into consideration. The simplest approach to augmentation of the total energy with these interactions is provided in the formalism of molecular mechanics [105], where molecules are treated as collections of atoms connected by flexible bonds. In the quantum consistent force field (QCFF/PI) method [169], the total energy is expressed as the sum of the π-electron energy calculated within the PPP approximation and the σ-electron contribution including the sp^3-hybridized carbon atoms and other atoms, the sp^2-hybridized carbons, and the mixed sp^2–sp^3 interactions.

2.6. Computational Techniques for (Hyper)polarizability

There are several theoretical techniques for studying the low-lying excited states and optical properties of carbon nanostructures. We summarize them here.

2.6.1. Sum-Over-States Approach

The sum-over-states (SOS) approach is based on the result of the self-consistent field solution of the Hartree-Fock Hamiltonian in the absence of an external field. Perturbation theory is then used to derive the (hyper)polarizabilities of atoms or molecules. In this approach, there is a lack of self-consistency between the eigenfunctions of the Hartree-Fock operator and the orbitals defining that operator when an electric field is applied. Details about SOS can be found in the original paper [170], where derivations of the hyperpolarizabilities β and γ and modeling of the SHG, THG, and other NLO processes were presented.

2.6.2. Finite Field Technique

Kurtz, Stewart, and Dieter [171] implemented a procedure for the calculation of molecular polarizabilities by using finite fields (FF). In this approach, an estimate of the uncertainty of a given calculation can be achieved by comparing the results obtained from using both the polarization and the energy versus the static field strength, that is, using the energy expansion and dipole expansion to calculate the response of the charge density to the field. Axial equations for FF calculations based on energy were originally derived by Bartlett and Purvis [172] and nonaxial equations by Kurtz, Stewart, and Dieter [171]. Similar equations for FF calculations based on polarization are given by Williams [173]. In this approach, the Hartree-Fock, self-consistent-field solution of the Hamiltonian is determined for various strengths of the external electric field. The results can be improved by treating in perturbation theory the difference between the exact Hamiltonian and the corresponding Hamiltonian in the so-called Moller-Plesset perturbation theory [105]. Usually, the Moller-Plesset correction in the second-order (MP2) is used. Details about FF calculations refer to Ref. [171–173].

2.6.3. SECI and SDCI

In the semiempirical CNDO/S approximation, several theoretical groups are able to obtain all the eigensolutions of the complete singly excited configuration interaction (SECI) [174] and those of truncated singly and doubly excited configuration interaction (SDCI) [175]. These solutions are used to calculate the molecular (hyper)polarizabilities with the sum-over-state method.

2.6.4. Spherical Shell Approximation

In this approximation [176], there is a free electron gas with N noninteracting electrons of mass m confined to a three-dimensional spherical shell of radius r. The (hyper)polarizabilities are calculated by using perturbation theory for a static electric field of strength F, and these results are valid for optical frequencies well below any resonance. Because there is no radial dependence, the lowest-order eigenfunctions are the spherical harmonics Y_{lm} with energies of $E_0 = \hbar^2 l(l+1)/(2mr^2)$. Then, the total energy can be expanded in a power series in the electric field and one gets the (hyper)polarizability.

2.6.5. TD-DFT and TD-HF

Ab initio and semiempirical methods have been extensively employed for NLO study. However, more efficient methods are needed for a quantitative theory of large systems. These methods should take account of electronic correlations, frequency dispersion, and an appropriate choice of the basis functions. For this, TD-HF and TD-DFT have been one of the time-dependent approaches to study NLO [144]. Compared with the TD-HF, the TD-DFT is not an approximation to the many-electron wavefunction. Rather, the level of an approximation is limited by our knowledge of the density functional.

In principle, an exact solution for excited states of many-electron system is possible within the nonadiabatic TD-DFT approach, and the TD-DFT gives a simple description with reasonable accuracy because of the inclusion of the electronic correlation effects in the excitation of molecules and condensed matters. Although TD-DFT and TD-HF approaches use a time-dependent many-body wavefunction in a form of a single Slater determinant, the meaning of this many-body wavefunction is very different in both approaches. Although the single Slater determinant in the TD-HF constitutes an approximation to the wavefunctions of a driven system, that in the TD-DFT is rather an auxiliary object that represents a system of fictitious Kohn-Sham noninteracting particles that are, however, able to reproduce the exact values of the electron density of the driven system at all times. As reviewed previously, the linear response TD-DFT is a well-developed and matured technique.

Meanwhile, an application of the TD-DFT for calculating nonlinear responses is a subject of great interest. The first application of the TD-DFT to NLO response was made by Senatore and Subbaswamy [177], who calculated the hyperpolarizability of rare gas atoms. The NLO susceptibilities of bulk semiconductors were calculated next [178, 179]. Recently, van Gisbergen, Snijders, and Baerends [180] extensively studied the dynamic hyperpolarizability

of molecules and showed that the TD-DFT is a promising tool to investigate NLO properties of molecules. Nevertheless, a systematic study of the nonlinear response in the TD-DFT method was not done until the recent work of Tretiak and Chernyak [144]. This work calculated the response of the density matrix to an external field in the adiabatic TD-DFT by mapping the equation of motion for the driven single-electron density matrix into the dynamics of coupled harmonic oscillators and then derived the resulting nonlinear response functions and the closed expressions for nonlinear frequency-dependent (hyper)polarizabilities.

2.6.6. Valence Effective Hamiltonian (VEH)

The VEH method was originally developed for molecules by Nicolas and Durand [181] and largely exploited for polymers by Andre et al. [182] and Bredas and colleagues [183–185]. It is based on the use of an effective Fock Hamiltonian F_{eff}, which combines a kinetic term and a summation over atomic potential V_a [181, 184, 185]:

$$F_{\text{eff}} = -\Delta/2 + \sum_a V_a \tag{42}$$

$$V_a = \sum_{l,m,i,j} C^a_{i,j,l,m} |\chi^a_{i,l,m}\rangle\langle\chi^a_{j,l,m}| \tag{43}$$

where the summation over l and m define the angular dependence of V_a. The coefficients are independent of m in the case of spherical symmetry. Usually, one 1s and 2p Gaussian Cartesian functions are used. The atomic potentials V_a are optimized in model molecules (for carbon atoms, the model molecules include ethane, butadiene, and acetylene) in order to minimize the difference between F_{eff} and the Fock Hamiltonian built from Hartree-Fock *ab initio* double-ζ calculations. In this way, the VEH method is completely nonempirical. Only valence electrons are explicitly considered and one-electron integrals need to be evaluated, which allows one to carry out calculations on large-size molecules.

2.6.7. Atomic Dipole Interaction Model

An alternate approach to estimate the (hyper)polarizability of a molecule is based on representing the molecule as a set of interacting point polarizabilities [186–188]. The atomic induced dipole moment of a system consisting of N interacting point polarizabilities can be written as

$$\mu_{\alpha,i} = \alpha_{\alpha,ij}\left(F^{\text{ext}}_{\alpha,j} + \sum_{\beta\neq\alpha} T_{\alpha\beta,jk}\mu_{\beta,k}\right) \tag{44}$$

where $F^{\text{ext}}_{\alpha,j}$ is the external electric field at atom α, and $T\mu$ term is the electric field of the induced dipole moment at site β calculated at atom α. The ijk subscripts denote the Cartesian coordinates, and the Einstein summation convention is used for repeated indices. If the atoms are spherically symmetric (that is, $\alpha_{\alpha,ij} = \alpha_\alpha\delta_{ij}$, where α_α is the isotropic atomic polarizabilities, a atom-type parameter), accurate parameters can be obtained if they are parameterized from a training set of molecular polarizabilities, for example, obtained from quantum chemical calculations instead of experimental data [189]. This is the advantage of this approach since experimental data are taken from various sources, inconsistent quality of the experimental data may have effects on the parameterization, and experimental polarizabilities are often taken from refractive indices that include local field factors as well as solvent and temperature effects. The resulting molecular polarizability α and molecular second hyperpolarizability γ can be expressed in terms of two-atom and four-atom relay tensors, respectively. The detailed theoretical backgrounds can be found in Ref. [190]. It should be mentioned that the dipole interaction model has been extended by atomic monopole polarizabilities [191, 192]. The monopole terms have been shown to be of significance [193]. Surely, it would be of interest to extend the atomic interaction model for hyperpolarizabilities with this kind of terms.

3. SECOND-ORDER OPTICAL NONLINEARITIES OF CARBON NANOSTRUCTURES

Second-order NLO measurements include techniques such as the Maker Fringe method, EFISHG, hyper-Rayleigh scattering, and the second-order NLO processes mentioned before. Detailed descriptions of the Maker Fringe method and hyper-Rayleigh scattering can be found in Refs. [87–101]. The EFISHG technique involves mixing three optical fields and one static field of parallel polarizations; details can be found in Ref. [194]. In the following, we review recent works on the second-order NLO properties of carbon nanostructures. Some results of fullerenes and derivatives are summarized in Tables 2 and 3.

3.1. Spherically Shaped Carbon Nanostructures

3.1.1. Experimental Measurement

From symmetry consideration, both C_{60} and C_{70} molecules possess inversion symmetry. Thus, within a dipolar approximation, no SHG activity should be detected for single C_{60} or C_{70} molecule. However, for their solutions or films, SHG signals were observed by several groups [195–199].

Hoshi et al. [195] were the first to investigate the second-order NLO response of C_{60} film deposited on a silica substrate by impinging an unfocused polarized Nd:YAG laser (1064 nm) on the film and measuring the dependence of the SHG intensities on the polarization of light and on the angle of incidence in the transmission geometry. No detectable SHG was emitted from the blank substrate. The presence of substrate is the only cause of the asymmetry. Later on, Kajzar et al. [198] also observed the incidence angle dependence of SHG intensity from a sublimed C_{60} thin film: the maximum of the SHG signal is at $\theta = 55°$. Koopmans et al. [197] performed more systematic studies on SHG from thin C_{60} films by combining frequency-, rotational-, angular-, and film-thickness-dependent measurement and gave the first SHG spectrum of pure C_{60} films or C_{60} surfaces. They found C_{60} films to show a high SHG efficiency and a strong and sharp resonance at $2\hbar\omega = 3.60$ eV, which is close to an allowed optical transition. Meanwhile, Wang et al. [196] also found that the square root of SHG intensity of C_{60} film increases linearly with the film thickness and varies with the change of temperature (reaching a maximum at 415 K), showing a bulk second-order optical nonlinearity. All these experiments demonstrated that the SHG signal from sublimed C_{60} thin films comes from the bulk and is not only an interface effect [200]. The large second-order NLO response of C_{60} films is due to small amounts of impurities (such as C_{60} isomers, other fullerenes or oxides that have noncentral symmetry), or electric quadrupole or magnetic dipole contributions that are allowed in centrosymmetric materials.

Kuhnke, Epple, and Kern [199] presented a wide SHG spectrum of C_{60} films: three pronounced resonances at 1.18 eV (1046 nm), 1.82 eV (678 nm), and 2.02 eV (0.611 nm) and a weak one at 1.35 eV (914 nm). The resonance peak observed at 1.82 eV by Kuhnke, Epple, and Kern [199] confirms the position and width of the resonance first observed by Koopmans et al. [197]. Kuhnke, Epple, and Kern [199] found a significant SHG signal at 1.165 eV (1059 nm), the energy of the study at fixed wavelength by Wilk et al. [201]. Later on, Janner [202] observed the resonance at 2.02 eV in an SHG spectrum. More recently,

Table 2. Second-order optical nonlinearities of C_{60} and derivatives measured by different experimental techniques (adapted from the work of Xie et al. [25]).

Fullerenes	β (10^{-29} esu)	$\chi^{(2)}$ (10^{-8} esu)	Method	State of Material	Ref.
C_{60}	0.11	0.13	ATR	Film	[203]
C_{60}/DEA	7 ± 2		EFISHG	Solution	[218]
$C_{60}(C_4H_8N_2)$	3.6 ± 1.3	18 ± 8	SHG	LB film	[224]
Carborane/C_{60} hybrid (1a)	35		SHG	Solution	[222]
Carborane/C_{60} hybrid (1b)	48		SHG	Solution	[222]
Carborane/C_{60} hybrid (1c)	119		SHG	Solution	[222]
CTP	8	3	SHG	LB film	[230, 231]

Table 3. The first hyperpolarizability (β, in 10^{-30} esu) and dipole (μ, in 10^{-18} esu) calculated for amino-ethylene (ME)-substituted C_{60} fullerene. The doped fullerene has C_{2v} symmetry. N is the total ME number. C_{60} has three unique sites ($S = 1, 2, 3$). $\beta^{(oct)}$ and $\beta^{(dip)}$ denote the octopolar and dipole parts contributed to the β tensor. The z axis is collinear to that of the undoped C_{60} which has D_{6h} symmetry. The value $\beta_{//z}$ corresponds to the sum over components in the z direction (adapted from Ref. [25, 248]).

N	Unique Site	β_{xxz}	β_{xyz}	β_{zzz}	$\beta_{//z}$	$\mu_{//z}$	$\beta^{(oct)}_{xxz}$	$\beta^{(oct)}_{xyz}$	$\beta^{(oct)}_{zzz}$	$\beta^{(dip)}_{//z}$
2	1	−13	−7	−75	−95	22	17	36	−18	−57
	2	−6	−7	0.2	−13	20				
	3	3	18	47	67	−6	−32	12	7	40
4	2	−285	2	3	280	28	−686	172	171	−168
	3	112	−576	−2	52	−32	304	−204	−33	31

using the experimental technique of electro-optically modulated attenuated total reflection (ATR) spectroscopy, Wang et al. [203] measured the second-order NLO response of of Langumir-Blodgett (LB) films of C_{60} by electro-optical (Pockels) characterization for the first time. Their observed electro-optical response was the Pockels effect, and the second-order NLO susceptibility $\chi^{(2)}$ of C_{60} film was determined to be 1.3×10^{-9} esu, in agreement with that of Wang et al. [196]. Note that although β is zero for isolated C_{60}, C_{60} film may exhibit nonzero value if it does not display centrosymmetry. Based on the work of Rashing et al. [204], the corresponding molecular hyperpolarizability β of the C_{60} molecule is about 1.1×10^{-30} esu [203]. Such a large electro-optical response of C_{60} shows that it is a promising NLO material, regardless of the molecular inversion symmetry.

The C_{60} film can be crystallized in an fcc structure at room temperature [205, 206] and is therefore centrosymmetric. As mentioned before, within the electric-dipole approximation, the C_{60} film would not show SHG signal, so that only the surface contribution could be detected. Nevertheless, the experiment of Wang and colleagues [196] showed that the surface contribution to SHG should be very small. As discussed previously, the origin of the SHG signal from a C_{60} film can be only an electric quadrupole, a magnetic dipole contribution, or from an impurity. Work by Koopmans et al. [207] has convincingly shown that a fullerene's SHG at 1.8 eV comes from a magnetic-dipole interaction. However, a study by Munn, Shuai, and Bredas [208] showed that although the SHG signal at 1.8 eV is dominated by the magnetic dipole response, this is not the case at other frequencies. Liu et al. [209] have observed a change of the SHG signal of C_{60} film with the structural phase transition of C_{60} at 245 K. Because the impurity cannot display the properties of C_{60}, their experimental results implied that the SHG in the fcc and sc phases was not induced simply by the impurity in the C_{60} material (or, say, the SHG response of impurity in C_{60} films cannot explain the change in SHG at 245 K). Thus, the SHG signal from C_{60} films cannot be attributed solely to the impurity.

Kuhnke, Epple, and Kern [199] were the first to measure the SHG response of higher fullerene C_{70}. They observed a weak resonance peak at 1.26 eV (979 nm) and a steep signal at 1.85 eV (667 nm) showing a broad resonance feature without resolved structure. In an Ne matrix et al. [210] assigned the lowest singlet (S_1) and triplet (T_1) excitons for isolated molecules to 1.93 eV and 1.56 eV, respectively. If a shift of about 0.1 eV between matrix isolated C_{70} and the solid exists, the lowest singlet exciton may correspond very well to the onset of the SHG intensity of C_{70} film. If so, the proximity of the next higher singlet exciton (S_2) and higher excitonic states may result in the broad resonance structure above 1.85 eV. However, the resonance peak at 1.26 eV is too low in energy to be assigned to the triplet exciton.

3.1.2. Theoretical Calculations

Theoretically, Qin, You, and Su [211] studied the dispersion relation of the quadrupole response for a single C_{60} molecule. They used a tight-binding approximation (see the Su-Schrieffer-Heeger (SSH) and extended SSH models reviewed in Section 4) to compute the

single-particle states for π-electron systems of C_{60} molecule. The Hamiltonian takes the form

$$H = -\sum_{\langle i,j\rangle,s}[t_0 - \alpha(l_{i,j} - l_1)](c_{i,s}^{\dagger}c_{js} + c.c.) + \frac{K_1}{2}\sum_{\langle i,j\rangle}(l_{i,j} - l_1)^2 + \frac{K_2}{2}\sum_{i} d\theta_{i,5}^2 + \frac{K_3}{2}\sum_{i}(d\theta_{i,6,1}^2 + d\theta_{i,6,3}^2) \tag{45}$$

where $c_{is}^{\dagger}$ (c_{is}) are electron creation (annihilation) operators at site i of the carbon atom with spin s; $\langle i, j\rangle$ means summation over the nearest-neighboring carbon atoms. The hopping integral is expanded to the linear term proportional to the length deviation from l_1. $l_{i,j}$ is the distance between atoms i and j, and α describes the strength of intersite electron-phonon coupling. The remaining terms describe elastic potential energies: K_1 is the stiffness constant of the bond-stretching spring; l_1 is the length at which all springs are relaxed; K_2 and K_3 are the stiffness constants of bond-bending springs; $\theta_{i,5}$ is the angle between two adjacent bonds of the pentagons; $\theta_{i,6,1}$ and $\theta_{i,6,2}$ are the angles between one bond of a pentagon and one adjacent bond of a hexagon; $d\theta_{i,5}$, $d\theta_{i,6,1}$, and $d\theta_{i,6,2}$ are angle deviations. In this model, the mutual interaction between the π electrons is accounted for only in the mean-field sense. And in the basis of atomiclike orbitals for π electrons, all matrix elements are neglected except the nearest neighbors. Based on this model, they found two resonant peaks located at 1.09 eV and 1.86 eV, respectively. The peak at 1.09 eV is only about 0.1 eV lower than the lowest peak observed by Kuhnke, Epple, and Kern [199], whereas the one at 1.86 eV is in good agreement with experiment [199]. Although their calculated peak is of asymmetric shape, no separate feature appears around the energy 2 eV. Hence, their calculation showed that the peak observed at 2.02 eV by Kuhnke, Epple, and Kern [199] is not caused by the electric quadrupole transition. The sharp minimum [211] between the two peaks seems to be caused by a change of amplitude sign. In the SHG spectrum of Kuhnke, Epple, and Kern [199], a similarly sharp minimum at 1.57 eV was observed. Hence, their theoretical calculation is in overall agreement with the experiment. Finally, it should be mentioned that no SHG peak is detected at the spin-forbidden excitation of the lowest triplet exciton at 1.55 eV (0.796 nm) [212]. The theoretical study of Qin, You, and Su [211] also demonstrated that the second-order quadrupole nonlinear contribution is large enough to explain the experimental results obtained by Wang and colleagues. [196].

Shuai and Bredas [213] developed the valence effective Hamiltonian (VEH) method to calculate the magnetic dipole (MD) and electric quadrupole (EQ) contributions to the hyperpolarizability β of molecular systems. For both C_{60} and C_{70}, they found that the MD contribution leads to a β value that is several times larger than that of the EQ contribution. Taking into account the local field correction factor for C_{60} in solid state, they calculated $\chi^{(2)}$ susceptibility values that are in excellent agreement with experiments [213]. The calculation of Shuai and Bredas [213] indicated that the SHG spectra for C_{70} come from the contribution of the electric quadrupole, having a sharp peak at 1.2 eV and a broad resonance structure between 2.0 and 2.4 eV, which is in good agreement with the SHG spectrum of Kuhnke, Epple, and Kern [199].

Because C_{70} has a lower symmetry than C_{60}, the low-energy part of the absorption spectrum of the C_{70} solid may be rich of dipole-allowed transitions. However, only two features were observed in the SHG spectrum of C_{70} film; that is, a lot of dipole allowed transitions are absent. Anyhow, this and the agreement with theoretical calculation suggest that dipole-forbidden transitions in C_{70} may be observed most easily. Indeed, Kajzar and colleagues [198, 214] predicted a dipole-forbidden transition at 2.49 eV.

3.2. Functionalized Carbon Nanostructures

3.2.1. Experimental Measurements

Charge-transfer is demonstrated to be the most effective mechanism to enhance the second-order optical nonlinearity of the organic molecule [215, 216]. Because fullerenes are excellent electron acceptors [217], forming charge-transfer complexes with appropriate donors would

break the centrosymmetry of C_{60} or C_{70} and thus induce significant second-order optical nonlinearity.

Wang and Cheng [218] measured the second-order NLO property of C_{60} charge-transfer complexes formed with *N,N*-diethylaniline (DEA), which is a well-known electron donor, to form charge-transfer complexes with various aromatic acceptor molecules [219], by using EFISHG technique [220, 221]. Indeed, the formation of charge-transfer complexes with DEA breaks the centrosymmetry in C_{60} and induces its second-order optical nonlinearity [218]. The value of β_μ in the dipolar direction for C_{60}/DEA is $(6.7 \pm 2) \times 10^{-29}$ esu.

Lamrani et al. [222] found experimentally that a combination of seemingly attractive carboranes and fullerene through an ethynyl π-system produced high β values. The β values found were 34.6×10^{-29}, 48.3×10^{-29}, and 118.9×10^{-29} esu for the three carborane-fullerene hybrids investigated.

The monofunctionized derivative C_{60}-($C_4H_8N_2$) has no centrosymmetry [223]. Gan et al. [224] have measured its SHG spectrum. The second-order nonlinear optical susceptibility $\chi^{(2)}$ and molecular hyperpolarizability β of C_{60}-($C_4H_8N_2$) are determined to be $(1.8 \pm 0.8) \times 10^{-7}$ esu and $(3.6 \pm 1.2) \times 10^{-29}$ esu, respectively [224].

Kajzar et al. [225] performed SHG measurements on C_{60}-based composites and multilayered charge-transfer structures with electron donors, 5, 10, 15, 20-tetrapheny-21*H*, 23*H*-porphine (TPP), and 5, 6, 11, 12-tetraphenylnaphthacene (rubrene) (TPN). In thin films that contain C_{60} molecules only is observed the SHG signal, and a significant enhancement in SHG signals is observed in the multilayered structures. The enhancement is due to a ground-state permanent electron transfer from the electron-donating molecules to C_{60} [225].

The charge separation in substituted C_{60} leading to enhancement of the second-order nonlinear optical susceptibility was also demonstrated theoretically [226, 227]. Further charge transfer can take place during the excitation process from the ground state to the excited state [227]. This kind of strong intramolecular charge transfer causes large second-order nonlinear optical susceptibility in the substituted fullerene. Among C_{60}, C_{60}/DEA, C_{60}/AN, and $C_{59}X$ (X = B, N), C_{60} has the smallest β value and $C_{59}X$ has the largest one [226, 227]. As discussed later, enhanced β value is also obtained for Si-substituted C_{60}. This means that substitute doping is a good means to achieving a large second-order nonlinear optical susceptibility of pure C_{60}.

The crown ether–modified C_{60} will be typical supramolecules that combine the unique properties of C_{60} and crown ethers. Surely, the solubility and amphiphathy of C_{60} can be improved because of the good hydrophilic properties of the crown ether moiety. This makes crowned C_{60} a kind of competitive compound to form structurally ordered fullerene thin films, and it may find practical applications as active surface layers in microsensors or optoelectronic devices. Several kinds of crowned fullerene derivatives were synthesized [228], based on which Langmuir and Langmuir-Blodgett (LB) films [229] were made. Recently, Guo and colleagues [230, 231] investigated macroscopic SHG of an LB film of a novel crowned [60] fulleropyrrolidine (CFP). They observed a strong SHG signal whose intensity depends linearly on on the number of film layers, implying that the LB films of CFP are a promising material for SHG. The second-order molecular susceptibility $\chi^{(2)}$ and hyperpolarizability β were determined to be 3.2×10^{-8} esu and 8.3×10^{-29} esu, respectively. The mechanism is still the intramolecular electron transfer.

The second-order nonlinearity of C_{60} endohedral doped with Li, Li@C_{60}, has been investigated both theoretical and experimentally by Campbell et al. [232]. They found, theoretically, that displacement of the Li off the idealized molecular center can induce a large first hyperpolarizability. An enhanced nonlinear response was also found experimentally by considering the SHG from thin films containing ~30% Li@C_{60}. From the results, it was suggested that $\chi^{(2)}$ of Li@C_{60} would be one or two orders of magnitude larger than the magnetically induced $\chi^{(2)}$ of pristine C_{60}.

Very recently, Fuks et al. [233] observed that fullerene molecules such as C_{60} can be modified by substituting fullerene with Si atoms to enhance the second-order photoinduced nonlinear optical properties. The experimental results are [233]: $\chi^{(2)}_{xxx} = 0.21, 0.87$ pm/V for $C_{59}Si$ and $C_{58}Si_2$, respectively; $\chi^{(2)}_{zzz} = 0.66, 0.67$ pm/V for $C_{59}Si$ and $C_{58}Si_2$, respectively. The substitution by two Si atoms does favor the larger second-order susceptibility in comparison

with that for only one-atom substitution [233]. This indicates that the photoinduced second-order NLO susceptibilities for $C_{58}Si_2$ are more isotropic than that for $C_{59}Si$.

Lin et al. [234] successfully incorporated C_{60}/C_{70} molecules into inorganic optical glasses with high melting temperatures. Fullerene doping leads to changes in glass properties [234]. Considering that optical glass without fullerene doping has centrosymmetry supporting no second-order optical nonlinearities, Zeng et al. [235] have recently shown that fullerene dopants can form ordered structures in the vitreous glass matrix. The fullerene doping broke the centrosymmetry of glass, leading to macroscopic second-order optical nonlinearities. An enhancement for the SHG of nearly two orders of magnitude was observed near the self-assembled islands [235].

3.2.2. Theoretical Calculations

Based on the intermediate neglect of differential overlap/configuration interaction plus sum-over-state approach (INDO/CI-SOS) method [236–238], Li, Feng, and Sun [239] calculated the second-order nonlinear optical susceptibilities β_{ijk} and β_μ of C_{60}/aniline (AN) (a theoretical model of the experimentally studied C_{60}/DEA [218]). The calculated value of β_μ at 1910 nm is 3.217×10^{-29} esu, which is in good agreement with experiment [218]. Later on, using this theory, Liu et al. [240] found that *N*-methyl-2-(2′-thiophene)-pyrrolo[3,4]C_{60} *N*-methyl-pyrrolo[3,4]C_{60}, and its derivatives exhibit good second-order optical nonlinearities. Both theoretically demonstrated the polarization effects, induced by the charge–transfer interaction, on the large number of hyperpolarizable π electrons, which have a significant contribution to the higher value of β_μ obtained by Wang and Cheng [218]. Recently, Fu et al. [241] found, using INDO/CI-SOS, that spiroannelated quinone-type methanofullerenes show enhanced second-order nonlinearity, compared with C_{60}/AN.

Rustagi, Ramaniah, and Nair [242] studied the first- and second-order hyperpolarizability of fullerene substitutionally doped with one B or one N atom by using a tight-binding method and found large enhancement of the nonlinearities in the doped fullerenes. Also, heterofullerenes substitutionally doped with silicon atoms [243] are of great interest from many aspects. One is the possibility of increasing second-order optical susceptibility. Recently, Fuks et al. [233] have shown, both theoretically and experimentally, that fullerene molecules such as C_{60} can be modified by substituting fullerene with Si atoms to enhance the second-order photoinduced nonlinear optical properties. The experimental results are [233]: $\chi^{(2)}_{xxx} = 0.21, 0.87$ pm/V for $C_{59}Si$ and $C_{58}Si_2$, respectively; $\chi^{(2)}_{zzz} = 0.66, 0.67$ pm/V for $C_{59}Si$ and $C_{58}Si_2$, respectively. The substitution by two Si atoms does favor the larger second-order susceptibility in comparison with that for only one-atom substitution [233]. This indicates that the photoinduced second-order NLO susceptibilities for $C_{58}Si_2$ are more isotropic than that for $C_{59}Si$.

It is known that high values of the first hyperpolarizability β tensor components for a given molecule imply optical nonlinearities as the molecule responds to applied optical fields. The so-called donor–acceptor model shows one approach to search the structure–property relationships of β and design our needed NLO materials. Based on this model, electronic excitation would result in charge migration from the donor to the acceptor group, showing a large dipole moment along the direction bounding the two groups. Indeed, many of donor–acceptor systems exhibit high values of β. This has also been demonstrated theoretically for B,N-substituted C_{60} by Jensen et al. [244] and for push-pull derivatives of C_{60} by Fanti, Orlandi, and Zerbetto [245], where also a relation between the conjugation path length and β was found.

The first hyperpolarizability is made up of its dipolar and octopolar contributions. It is possible to cancel the dipolar part by symmetry and enhance the octopolar part of the β tensor [246, 247]. This gives us another approach to designing NLO materials by exploring octopolar architectures of a number of possible molecules [247]. Such an octopolar-enhanced approach has two advantages [246]: preventing the molecules from losing their optical properties in bulk by dipole cancellation on antisymmetric crystallization and raising the possibility for the system to have more than one optical axis. Recently, Barbosa and Nascimento [248] proposed a new approach, incorporating an octopolar-enhanced approach into the donor–acceptor model, for designing molecules with large hyperpolarizabilities.

In their approach, a central acceptor unit with a high value of electron affinity is used and bounded to several donor groups, maximizing either the dipolar or the octopolar components of the β tensor.

Recently, C_{36} fullerene [249, 250] has been synthesized. This molecule with D_{6h} and D_{2d} symmetries [see Fig. 4(q) and Fig. 4(r)] exhibits rich structural and electronic properties [251–261]. The crystal D_{6h} structure has a very high electron affinity (>3 eV) and many symmetry elements [248]. By adding two equivalent donor amino-ethylene (ME) groups to one of the three unique symmetric positions of the central acceptor group C_{36} fullerene, Barbosa and Nascimento [248] explored the second-order NLO properties of such system. The nuclear frameworks of the new system with C_{2h} point group cannot present second-order NLO properties. The framework with C_{2v} symmetry, as listed in Table 3, may exhibit high values of the $\beta_{(z)}$ component, more than 10 times larger than those for the p-nitroaniline [262]. Because of the symmetry, most of the components of the β tensor vanish. Their results showed that a symmetric positioning of a great number of donor groups is more important to enhance the octopolar component of β than an augmentation of chain length, which significantly influences only the dipolar component. According to their proposed approach, some of the designed molecules can be good candidates for building second-order NLO materials.

4. THIRD-ORDER OPTICAL NONLINEARITIES OF CARBON NANOSTRUCTURES

There are several popular NLO measurement techniques to obtain the third-order optical susceptibility $\chi^{(3)}$ of a material: DFWM, THG, EFISHG, Z-scan, optical Kerr effect, optically heterodyned OKE (OHD-OKE), and coherent anti-Stokes Raman spectroscopy (CARS) vibrational lineshape analysis. In Section 1.3, we introduced the THG and DFWM techniques. Actually, the THG process occurs in every medium, and consequently the harmonic intensity is a result of the superposition of harmonic fields generated in the separate media. And the DFWM is a powerful tool for measuring $\chi^{(3)}$ of a material. From the strength of the signal, one can determine the magnitude of $\chi^{(3)}$. However, the time-dependent behavior of the DFWM signal is related to the development and persistence of the grating and therefore provides information on the ultrafast physics of the material. The Z-scan technique [263] measures the real and imaginary parts of $\chi^{(3)}$ at the same time and is based on the variation of transmitted radiation intensity by alteration of the geometrical parameters of the interaction region. Explicitly, Z-scan is achieved by gradually moving a sample along the Z-axis of propagation of a focused Gaussian beam through its focal plane and measuring the transmission of the sample for each Z position. The recording of the transmission as a function of the Z coordinate provides accurate information about the nonlinear effect present as the sample experiences different electric field strengths at different positions. For an ordinary OKE technique, the detected signal is proportional to the quadrature of $\chi^{(3)}$, and one can acquire only the modulus of $\chi^{(3)}$, which should be related to the excitation light because the imaginary component of $\chi^{(3)}$ is determined by the absorption of the sample at the wavelength of the excitation light. In most cases, the real component of $\chi^{(3)}$ is, in fact, more concerned because it is a more important parameter for the application in photonic devices. OHD-OKE determines the magnitudes and signs of $\chi^{(3)}$, and this method greatly improves the signal-to-noise ratio of experimental data. In the CARS technique, one fits the dispersive lineshape caused by interference between the unknown nonlinearity and a Raman resonance of a standard like benzene. This CARS procedure is well accepted and has been shown to provide accurate values for γ. Details about CARS techniques can be found in Ref. [264].

Photonic applications (for example, data processing, eye and sensor protection, and all-optical switching) need molecules with large third-order optical nonlinearities [92–94, 96–101]. However, the $\chi^{(3)}$ or γ magnitudes of most materials are usually smaller than those needed for photonic devices. Hence, searching for potential third-order NLO materials with a large NLO response has been an interesting and important issue in physics, chemistry, and material science. Quantum dots and conjugated π-electron organic systems (for example,

polydiacetylenes, polyacetylenes, and polythiophenes) have been shown to have large third-order optical nonlinearities [92–94, 98, 100, 265–276]. However, because of the overtones of high-energy C—H and O—H vibrations, those organic materials show strong absorption in the near-infrared region. Obviously, this would limit the application of organic materials in the infrared region, for example, telecommunications. The advent of the technology for production of bulk quantities of spherically and rod-shaped carbon nanostructures [9–15, 22] provides us with another class of completely conjugated materials having quantum-dot nature, as well as a large number of delocalized π electrons. Because these carbon nanostructures are uniquely composed of carbon atoms, there is no C—H or O—H bond present in these structures. This implies that no absorption would be observed in the infrared region. On the other hand, the HOMO-LUMO energy gaps in these carbon nanostructures are narrower [277] than those in conjugated polymers (the HOMO-LUMO energy gaps in conjugated polymers are broadened because of the confirmations, polymer chain length distributions, and vibronic couplings to electronic levels). Thus, carbon nanostructures exhibit narrower resonances. These novel features naturally make carbon nanostructures appealing NLO materials for photonic applications and stimulate the investigation of the third-order optical nonlinearities of carbon nanostructures. In the following, we review the experimental and theoretical studies on the third-order optical nonlinearities of carbon nanostructures. The results are summarized in Tables 4 through 11.

4.1. Spherically Shaped Carbon Nanostructures

4.1.1. Experimental Measurements

The third-order optical nonlinearity of C_{60} was first measured by Blau et al. [278], using DFWM technique. Their measured third-order optical susceptibility has an error by more than three orders of magnitude than the accurate value determined later [279, 280]. Nevertheless, their experiment did attract extensive studies of the third-order optical nonlinearities of C_{60} in the world. Table 4 collects some of reported optical susceptibilities $\chi^{(3)}_{xxxx}$ and hyperpolarizability γ of C_{60} at a few selected wavelengths measured by DFWM, THG, and electric-field-induced second-harmonic generation (EFISHG). The differences and errors of the measured results were quite indicative of the later developments for obtaining the value $\chi^{(3)}$ or γ of C_{60} [195, 218, 279–301]. The third-order NLO responses are very sensitive to a lot of experimental factors, for example, the incident laser power, pulse duration, the measurement techniques adopted, and the state of materials (sample preparation methods). Thus, it is not easy to directly compare these experimental results obtained from different groups by using different techniques. But, in general, there is a reasonable agreement between the values measured by THG and EFISHG. From Table 4, two interesting features are observed from the wavelength (λ)-dispersed $\chi^{(3)}$ spectrum for C_{60}: (1) a strong resonancement in $\chi^{(3)}$ with two peaks at 1064 nm and 1210 nm and (2) a sharp decrease of $\chi^{(3)}$ at a shorter wavelength. As reviewed in the theoretical part of this section, a three-level model proposed by Kajzar et al. [198] explained that the resonances observed at 1.064 and 1.210 μm are three-photon and two-photon processes, respectively. In spite of the obvious discrepancy for the $\chi^{(3)}$ or γ value of C_{60} in different experimental groups, the most recent experiments have demonstrated that the third-order optical nonlinearity of C_{60} has a small value [302, 303]: the lower limit of γ is 3.7×10^{-35} esu measured by nondegenerate four-wave mixing [302], whereas the upper limit of γ is 9.0×10^{-35} esu determined by the femtosecond OKE [303].

Compared with C_{60}, higher fullerenes have attracted less attention on their third-order optical nonlinearities because large amounts of higher fullerenes are not available. This is also why the synthesis and isolation of higher fullerenes is still an active research field. Some experimental results [218, 289–291, 295, 297, 300, 304, 305] and theoretical calculations [293, 294, 296, 306–327] have shown that higher fullerenes possess larger third-order optical nonlinearity than C_{60}. The most easily purified higher fullerene, C_{70}, was experimentally proven to have a γ magnitude of 1.6 to 3 times larger than that of C_{60} [218, 289–291, 295, 297, 300]. Tables 4 and 6 list the third-order nonlinear optical susceptibilities $\chi^{(3)}$ of C_{70} and higher fullerenes studied by different experimental or theoretical techniques. In the following, we briefly review recent experimental studies on the third-order

Table 4. Third-order optical nonlinearities of spherically shaped carbon nanostructures measured by different experimental techniques. λ and τ are the wavelength and pulse duration of the laser source, respectively. Some data are adapted from Nalwa [99], Xie [24], Xie et al. [25], Kajzar et al. [198], Rustagi et al. [328], and Belousov et al. [329].

Carbon Structure	Wavelength (μm)	τ	$\chi^{(3)}$ (10^{-12} esu)	γ (10^{-32} esu)	Method	State of Material	Ref.
C_{60}	0.532	70 ps	0.1	2200 ± 600	DFWM	Solution	[305]
	0.356	50 ps	15		THG	Thin film	[287]
	0.405	50 ps	21		THG	Thin film	[287]
	0.494	50 ps	27		THG	Thin film	[287]
	0.497	0.5 ps	-0.7 ± 0.2	-130 ± 30	z-scan, DFWM	Solution	[382]
	0.588	50 ps	22		THG	Thin film	[287]
	0.597	1.2 ps	380		DFWM	Thin film	[291]
	0.602	400 fs		100 ± 20	DFWM	Solution	[286]
		50 ps	11		THG	Thin film	[287]
	0.620	60 fs	220		OKE	Solution	[286]
	0.633	100 ps	200		DFWM	Thin film	[295]
	0.640	120 fs		1600	OKE	Solution	[378]
	0.647	150 fs		1600	OKE	Solution	[300]
		165 fs	-0.3		OHD-OKE	Film	[387]
	0.675	1.2 ps	82		DFWM	Thin film	[291]
	0.686	50 ps	9		THG	Thin film	[287]
	0.796	50 ps	4.5		THG	Thin film	[287]
	0.810	120 fs		<1	OKE	Solution	[303]
	0.816	13 ns	13		THG	Thin film	[198]
	0.825	13 ns	7		THG	Thin film	[198]
	0.834	13 ns	10		THG	Thin film	[198]
	0.843	13 ns	7		THG	Thin film	[198]
	0.850	10 ns	15		THG	Thin film	[281, 282]
	0.852	13 ns	19		THG	Thin film	[198]
	0.861	13 ns	15		THG	Thin film	[198]
	0.870	13 ns	15		THG	Thin film	[198]
	0.882	13 ns	13		THG	Thin film	[198]
	0.891	13 ns	18		THG	Thin film	[198]
	0.900	13 ns	15		THG	Thin film	[198]
	0.909	13 ns	21		THG	Thin film	[198]
	1.022	13 ns	71		THG	Thin film	[198]
	1.030	13 ns	73		THG	Thin film	[198]
	1.039	13 ns	73		THG	Thin film	[198]
	1.056	13 ns	74		THG	Thin film	[198]
	1.064	35 ps	7		DFWM	Thin film	[285]
		50 ps	60000		DFWM	Solution	[278]
			3		DFWM	Solid	[284]
			3	1600	DFWM	Solution	[284]
		50 ps	14		THG	Thin film	[287]
			200		THG	Thin film	[195]
		10 ns	82		THG	Thin film	[281, 282]
		13 ns	82		THG	Thin film	[198]
		10 ns	72	47	THG	Solution	[289]
		35 ps	7	3	DFWM	Thin film	[297]
	1.074	13 ns	87		THG	Thin film	[198]
	1.083	13 ns	76		THG	Thin film	[198]
	1.092	13 ns	74		THG	Thin film	[198]
	1.138	13 ns	63		THG	Thin film	[198]
	1.158	13 ns	67		THG	Thin film	[198]
	1.165	13 ns	78		THG	Thin film	[198]
	1.177	13 ns	51		THG	Thin film	[198]
	1.236	13 ns	53		THG	Thin film	[198]
	1.245	13 ns	51		THG	Thin film	[198]
	1.254	13 ns	52		THG	Thin film	[198]
	1.263	13 ns	54		THG	Thin film	[198]
	1.269	13 ns	52		THG	Thin film	[198]
	1.278	13 ns	52		THG	Thin film	[198]

continued

Table 4. Continued.

Carbon Structure	Wavelength (μm)	τ	$\chi^{(3)}$ (10^{-12} esu)	γ (10^{-34} esu)	Method	State of Material	Ref.
	1.287	13 ns	57		THG	Thin film	[198]
	1.291	13 ns	60		THG	Thin film	[198]
	1.296	13 ns	57		THG	Thin film	[198]
	1.305	13 ns	56		THG	Thin film	[198]
	1.323	13 ns	61		THG	Thin film	[198]
	1.330	50 ps	30		THG	Thin film	[287]
		10 ns	61		THG	Thin film	[281, 282]
	1.332	13 ns	57		THG	Thin film	[198]
	1.344	13 ns	60		THG	Thin film	[198]
	1.356	13 ns	60		THG	Thin film	[198]
	1.368	13 ns	58		THG	Thin film	[198]
	1.381	13 ns	57		THG	Thin film	[198]
	1.413	13 ns	57		THG	Thin film	[198]
	1.437	13 ns	55		THG	Thin film	[198]
	1.456	13 ns	56		THG	Thin film	[198]
	1.500	10 ns	30	-13 ± 8	THG	Solution	[289]
	1.815	50 ps	4		THG	Thin film	[287]
	1.907	13 ns	32		THG	Thin film	[198]
		13 ns	32		THG	Thin film	[283]
		13 ns	22		THG	Thin film	[380]
			20		THG	Thin film	[196]
	1.910		16	7.5 ± 2	EFISHG	Solution	[218]
		50 ps	9		THG	Thin film	[287]
		10 ns	32		THG	Thin film	[281, 282]
		10 ns	20		THG	Thin film	[283]
	2.000	10 ns	37	16 ± 8	THG	Solution	[289]
	2.373	50 ps	4		THG	Thin film	[287]
	2.380	50 ps	4		THG	Thin film	[287]
C_{70}	0.532	70 ps	0.4	13000 ± 4000	DFWM	Solution	[305]
	0.597	12 ps	2100		DFWM	Thin film	[291]
	0.633	100 ps	300		DFWM	Thin film	[295]
	0.647	150 fs		4700	OKE	Solution	[300]
	0.675	1.2 ps	64		DFWM	Thin film	[291]
	0.799	13 ns	8		THG	Thin film	[198]
	0.810	120 fs		5	OKE	Solution	[383]
	0.816	13 ns	11		THG	Thin film	[198]
	0.825	13 ns	13		THG	Thin film	[198]
	0.834	13 ns	6		THG	Thin film	[198]
	0.843	13 ns	6		THG	Thin film	[198]
	0.852	13 ns	7		THG	Thin film	[198]
	0.861	13 ns	10		THG	Thin film	[198]
	0.870	13 ns	6		THG	Thin film	[198]
	0.882	13 ns	7		THG	Thin film	[198]
	0.891	13 ns	12		THG	Thin film	[198]
	0.900	13 ns	11		THG	Thin film	[198]
	1.013	13 ns	30		THG	Thin film	[198]
	1.022	13 ns	29		THG	Thin film	[198]
	1.039	13 ns	29		THG	Thin film	[198]
	1.047	13 ns	29		THG	Thin film	[198]
	1.056	13 ns	26		THG	Thin film	[198]
	1.064	35 ps	12		DFWM	Thin film	[297]
		10 ns	6	12000	DFWM	Soution	[290]
		10 ns	1400	570	THG	Soution	[289]
		13 ns	26		THG	Thin film	[214]
	1.074	13 ns	26		THG	Thin film	[198]
	1.129	13 ns	36		THG	Thin film	[198]
	1.131	13 ns	35		THG	Thin film	[198]
	1.138	13 ns	26		THG	Thin film	[198]
	1.148	13 ns	34		THG	Thin film	[198]
	1.158	13 ns	39		THG	Thin film	[198]

continued

Table 4. Continued

Carbon Structure	Wavelength (μm)	τ	$\chi^{(3)}$ (10^{-12} esu)	γ (10^{-34} esu)	Method	State of Material	Ref.
	1.167	13 ns	43		THG	Thin film	[198]
	1.177	13 ns	37		THG	Thin film	[198]
	1.187	13 ns	30		THG	Thin film	[198]
	1.216	13 ns	47		THG	Thin film	[198]
	1.226	13 ns	44		THG	Thin film	[198]
	1.247	13 ns	43		THG	Thin film	[198]
	1.267	13 ns	19		THG	Thin film	[198]
	1.278	13 ns	29		THG	Thin film	[198]
	1.289	13 ns	51		THG	Thin film	[198]
	1.300	13 ns	67		THG	Thin film	[198]
	1.322	13 ns	69		THG	Thin film	[198]
	1.333	13 ns	71		THG	Thin film	[198]
	1.344	13 ns	68		THG	Thin film	[198]
	1.368	13 ns	83		THG	Thin film	[198]
	1.378	13 ns	66		THG	Thin film	[198]
	1.390	13 ns	77		THG	Thin film	[198]
	1.401	13 ns	90		THG	Thin film	[198]
	1.413	13 ns	76		THG	Thin film	[198]
	1.420	10 ns	90		THG	Thin film	[214]
	1.425	13 ns	72		THG	Thin film	[198]
	1.437	13 ns	78		THG	Thin film	[198]
	1.449	13 ns	83		THG	Thin film	[198]
	1.500	10 ns	54	220	THG	Solution	[289]
	1.907	13 ns	24	110 ± 10	THG	Thin film	[198]
	1.910		44	13 ± 3	EFISHG	Solution	[218]
		10 ns	24		THG	Thin film	[214]
	2.000	10 ns	91	−38	THG	Solution	[289]
C_{76}	0.532	70 ps	0.3 ± 0.1	8000 ± 3000	DFWM	Solution	[305]
C_{78}	0.532	70 ps	0.6 ± 0.1	15000 ± 3000	DFWM	Solution	[305]
C_{84}	0.532	70 ps	0.4 ± 0.1	12000 ± 3000	DFWM	Solution	[305]
	0.647	150 fs		5200	OKE	Solution	[300]
C_{86}	0.532	70 ps	0.5 ± 0.2	13000 ± 5000	DFWM	Solution	[305]
C_{90}	0.532	70 ps	0.7 ± 0.2	18000 ± 6000	DFWM	Solution	[304, 305]
C_{94}	0.532	70 ps	0.6 ± 0.2	19000 ± 6000	DFWM	Solution	[305]
C_{96}	0.532	70 ps	0.7 ± 0.2	21000 ± 6000	DFWM	Solution	[305]

optical nonlinearities of higher fullerenes. Other reviews on the nonlinear optical properties of fullerene (mainly C_{60} and C_{70} molecules) are also available in the recent literatures by Nalwa [99], Xie [24], Kajzar et al. [198], Rustagi, Nair, and Ramaniah [328], and Belousov et al. [329].

Neher et al. [289] measured the optical susceptibilities $\chi^{(3)}$ of C_{70} in the infrared region (1.064, 1.500, and 2.000 μm) by using the THG technique in a toluene solution. Compared with C_{60}, C_{70} exhibits enhanced third-order optical nonlinearity. In particular, a strong effect with γ exceeding 5×10^{-32} esu was observed in the three-photon resonant regime, and a negative real hyperpolarizability in C_{70} was found in the nonresonant measurements. Later on, Kajzar et al. [198, 214] reported a broad THG spectrum for C_{70}. The overall THG spectrum shows a broad resonance enhancement in $\chi^{(3)}$ with a maximum value of $\chi^{(3)} = (0.9 \pm 0.1) \times 10^{-10}$ esu located at 1.420 μm and a dramatic decrease at short wavelengths [198, 214]. Yang et al. [290] were the first to measure the third-order NLO susceptibility of C_{70} in a toluene solution by using the DWFM technique with 10 ns laser pulses at 1.064 nm. The third-order nonlinear optical susceptibility $\chi^{(3)}$ is determined to be 5.6×10^{-12} esu for a C_{70} toluene solution at a concentration of 0.476 g/l. Later on, Flom et al. [291]; Lindle et al. [297], and Rosker et al. [295] performed time-resolved DFWM measurement on the third-order optical nonlinearities of C_{70} film by using a picosecond laser pulse. The C_{70} was found to exhibit a two-photon resonantly enhanced third-order optical response [297]. The dynamics of C_{70} shows wavelength and fluence dependence [291]. To our interest, at high laser intensities, their experiment showed a fifth-order component to the NLO signal due to

Table 5. Second-order hyperpolarizabilities γ of C_{60} calculated by different theoretical techniques (adapted from Xie et al. [25]).

Technique	Wavelength (μm)	Process	γ (10^{-34} esu)	Ref.
INDO/SDCI-SOS	∞	Static	11	[350]
INDO/SDCI-SOS	1.064	DFWN	49	[350]
INDO/SDCI-SOS	1.064	EFISH	149	[350]
INDO/SDCI-SOS	1.064	THG	71	[350]
INDO/SDCI-SOS	1.910	DFWN	37	[350]
INDO/SDCI-SOS	1.910	EFISH	79	[350]
INDO/SDCI-SOS	1.910	THG	15	[350]
INDO/SCI-SOS	1.064	DFWM	53	[292]
INDO/SDCI-SOS	1.064	DFWM	44	[292]
INDO/SDCI-SOS	0.532	DFWM	19	[292]
INDO/SDCI-SOS	1.910	EFISH	41	[292]
CNDO/S-SOS	∞	Static	5	[344]
CNDO/S-SOS	∞	Static	33	[344]
CNDO/S-SOS	1.900	THG	36	[344]
CNDO/S-SOS	1.830	THG	36	[344]
CNDO/S-SOS	1.320	THG	39	[344]
CNDO/S-SOS	1.064	THG	44	[344]
CNDO/S-SOS	∞	Static	8	[349]
CNDO/S-SOS	1.900	THG	9	[349]
CNDO/S-SOS	1.830	THG	9	[349]
CNDO/S-SOS	1.320	THG	10	[349]
CNDO/S-SOS	1.064	THG	11	[349]
MNDO-FF	∞	Static	0.3	[348]
MNDO-SOS	∞	Static	0.5	[348]
NDDO/PM3-FF	∞	Static	0.3	[348]
NDDO/PM3-SOS	∞	Static	0.5	[348]
NDDO/AM1-FF	∞	Static	0.3	[348]
NDDO/AM1-SOS	∞	Static	0.4	[348]
MNDO/PM3-FF	∞	Static	0.4	[294]
AM1-FF	∞	Static	2	[347]
AM1-FF	∞	Static	1	[322, 346]
AM1/Valence-FF	∞	Static	0.2	[311]
INDO-TDHF	∞	Static	0.3	[286]
INDO-TDHF	1.370	OKE	0.3	[286]
INDO-TDHF	1.370	EFISH	0.3	[286]
INDO-TDHF	1.370	DFWM	0.3	[286]
Hückel	∞	Static	25	[343]
Spherical shell	∞	Static	−3	[306]
VEH-SOS	∞	Static	2	[312, 313]
Tight-binding	∞	Static	176	[336]
Tight-binding	∞	Static	0.1	[342]
Tight-binding	∞	Static	31	[242]
SSH-SOS	∞	Static	6	[307]
SSH-SOS	∞	Static	2	[324]
PPP-FF	∞	Static	0.3	[343]
PPP-SOS	∞	Static	1	[343]
PPP/MP2-F	∞	Static	0.4	[343]
PPP-TDHF	∞	Static	1	[337]
PPP-TDHF	∞	Static	0.2	[338]
PPP-TDHF	1.880	EFISH	0.3	[338]
PPP-TDHF	1.879	THG	0.4	[339]
PPP-TDHF	1.320	EFISH	0.4	[338]
PPP-TDHF	1.320	THG	0.7	[339]
PPP-TDHF	1.060	EFISH	0.5	[338]
PPP-TDHF	1.060	THG	3	[339]
Coupled HF	∞	Static	0.08	[341]
HF-SOMO	∞	Static	2	[315]
HF-RPA	∞	Static	0.6	[301]
HF-RPA	∞	Static	0.6	[296]
HF/STO-3G-FF	∞	Static	0.1	[333]

continued

Table 5. Continued

Technique	Wavelength (μm)	Process	γ (10^{-34} esu)	Ref.
HF/3-21G-FF	∞	Static	0.1	[333]
LDA	∞	Static	0.6	[340]
LDA-RPA	∞	Static	0.3	[345]
LDA-FF	∞	Static	0.2	[334]
LDA-FF	∞	Static	0.4	[335]
LDA-TDDFT	∞	Static	0.4	[330]
LDA-TDDFT	∞	Static	0.7	[244]
LDA-TDDFT	∞	Static	0.6	[331]
LDA-TDDFT	0.830	OKE	0.7	[331]
LDA-TDDFT	1.064	DFWM	0.8	[331]
LDA-TDDFT	1.910	EFISH	0.7	[331]
LDA-TDDFT	2.380	THG	0.7	[331]
BLYP-TDDFT	∞	Static	0.8	[331]
BLYP-TDDFT	0.830	OKE	0.9	[331]
BLYP-TDDFT	1.064	DFWM	1.0	[331]
BLYP-TDDFT	1.910	EFISH	0.8	[331]
BLYP-TDDFT	2.380	THG	0.9	[331]
LB94-TDDFT	∞	Static	0.5	[331]
LB94-TDDFT	0.830	OKE	0.6	[331]
LB94-TDDFT	1.064	DFWM	0.6	[331]
LB94-TDDFT	1.910	EFISH	0.5	[331]
LB94-TDDFT	2.380	THG	0.5	[331]
LB94-TDDFT	∞	Static	0.3	[330]
LB94-TDDFT	0.830	OKE	0.4	[330]
LB94-TDDFT	1.910	EFISH	0.4	[330]

a two-photon excited-state transient grating [297]. As well as THG and DFWM measurements, Wang and Cheng [218] determined the optical susceptibility $\chi^{(3)}$ of C_{70} by using the EFISHG technique with a 1.910 μm radiation. It concluded that the optical nonlinearities of C_{70} are comparable to those of linearly conjugated organics with due consideration to the molecular size [218]. This is consistent with THF and DFWM measurements. Similar to C_{60}, C_{70} spectrum exhibits several interesting features: (1) a sharp decrease of χ^3 at a shorter wavelength; (2) a strong resonance enhancement in $\chi^{(3)}$ at 1.410 μm; (3) a shoulder around 1.064 μm. The observed resonance at 1410 nm is caused by a three-photon resonance with a one-photon allowed transition lying at 470 nm, while the shoulder at 1064 nm is due a two-photon resonance [198, 214].

Unlike C_{60} and C_{70}, less work has been done on the NLO of higher fullerenes. Sun et al. [300] were the first to measure the third-order optical nonlinearity of high fullerenes C_{84} by using the time-resolved OKE technique with 150 fs laser pulse at 0.647 μm. A large instantaneous NLO response was observed for C_{84}. In comparison with C_{60} and C_{70}, the γ value of C_{84} is enhanced a little. Such a small accretion of the optical Kerr response of C_{84} is due to the sample impurity. The experiment of Huang et al. [304] indicated that a purity of C_{84} is higher than 85%, but the impurities include C_{78}, C_{82}, C_{86} and C_{90}. Later on, the third-order optical nonlinearities of C_{90} (>97% purity) at 0.532 μm were measured by Huang et al. [304], using the DFWM technique. Their measurement determined the second-order hyperpolarizability γ of C_{90} to be $(1.8 \pm 0.6) \times 10^{-30}$ esu (the γ value of C_{60} was $(2.2 \pm 0.6) \times 10^{-31}$ esu in the same condition). Compared with C_{60}, the γ value of C_{90} was enhanced by about one order of magnitude. This is expected because there are more highly delocalized π-conjugated electrons over the spherical-like surface in C_{90} that in C_{60}. Furthermore, the systematical work of Huang et al. [305] on other higher fullerenes (for example, C_{76}, C_{78}, C_{84}, C_{86}, C_{94}, and C_{96}) performed by using DFWM with 70 ps laser pulses at 0.532 μm under optimized experimental conditions, found an eightfold increase in the γ values from C_{60} to C_{96} (see Table 4). This is consistent with the trend predicted by theory discussed below [314–321].

Table 6. Static second-order hyperpolarizabilities γ of higher fullerenes calculated by different theoretical techniques. Some data are adapted from Xie [24], Xi et al. [25].

Molecule	Technique	γ (10^{-31} esu)	Ref.
C_{70}	MNDO/PM3-FF	0.7	[294]
	NDDO/PM3-FF	0.6	[348]
	NDDO/PM3-SOS	5	[348]
	NDDO/AM1-FF	0.5	[348]
	NDDO/AM1-SOS	5	[348]
	INDO/SDCI-SOS	27	[350]
	INDO/SDCI-SOS	50	[293]
	CNDO/S-SOS	9	[314]
	AM1/valence-FF	0.5	[311]
	AM1-FF	12	[322, 346]
	Spherical shell	−4	[306]
	VEH-SOS	9	[312, 323]
	SSH-SOS	6	[307]
	SSH-SOS	11	[324]
	HF-RPA	0.8	[296]
	HF-SOMO	3	[315]
C_{76}	CNDO/S-SOS	12	[314]
	HF-SOMO	4	[315]
$C_{78}(D_3)$	HF-SOMO	4	[315]
	SSH-SOS	7	[324]
	AM1-FF	2	[322, 346]
	AM1/valence-FF	3	[327]
$C_{78}(C_{2v}(1))$	HF-SOMO	4	[315]
	SSH-SOS	20	[324]
	AM1-FF	2	[322, 346]
	AM1/valence-FF	4	[327]
$C_{78}[C_{2v}(2)]$	HF-SOMO	4	[315]
	SSH-SOS	5	[324]
	AM1-FF	2	[322, 346]
	AM1/valence-FF	3	[327]
$C_{78}[D_{3h}(1)]$	SSH-SOS	5	[324]
	AM1-FF	2	[322, 346]
	AM1/valence-FF	5	[327]
$C_{78}[D_{3h}(2)]$	SSH-SOS	21	[324]
	AM1-FF	2	[322, 346]
	AM1/valence-FF	3	[327]
$C_{84}(D_2)$	CNDO/S-SOS	18	[314]
	AM1/valence-FF	0.6	[311]
	AM1-FF	2	[322, 346]
	HF-SOMO	4	[315]
	HF-RPA	0.8	[296]
$C_{84}(D_{2d})$	CNDO/S-SOS	18	[314]
	AM1/valence-FF	0.6	[311]
	AM1-FF	2	[322, 346]
	HF-SOMO	5	[315]
C_{100}	HF-SOMO	9	[315]

4.1.2. Theoretical Calculations

In Tables 5 and 6, we summarize the second hyperpolarizabilities γ of several spherically shaped carbon nanostructures, predicted by different theoretical methods. The theoretical results have a large spread. However, if one considers the *ab initio* methods, such as HF and DFT, the calculated results are more consistent. The low value of γ for C_{60} is confirmed by *ab initio* calculations [244, 296, 330, 331], although the experimental results tend in general to be larger than the theoretical results. On the other hand, because of the presence of medium effects, a direct comparison between experiments and theoretical predictions is difficult. That is, the experimental results refer to the condensed phase whereas the theoretical results usually are for the gas phase. By considering the intermolecular interaction and local-field corrections on the NLO properties of C_{60}, Luo et al. [332] found a good agreement with

Table 7. The averaged second-order hyperpolarizability $\{\bar{\gamma} = [\gamma_{xxxx} + \gamma_{yyyy} + \gamma_{zzzz} + 2(\gamma_{xxyy} + \gamma_{xxzz} + \gamma_{yyzz})]/5\}$ and the averaged pyramidalization angle $\bar{\theta}_p = \sum \theta_p / n$, n is the total number of carbon atoms for the isolated-pentagon rule isomers of higher fullerenes (adapted from Refs. [25, 326]).

Fullerene	Symmetry	$\bar{\gamma}$ (10^{-34} esu)	$\bar{\theta}_p$
C_{76}	T_d	1.21	0.1810
	D_2	0.55	0.1822
C_{78}	$D_{3h}(1)$	0.82	0.1784
	$C_{2v}(1)$	0.68	0.1791
	$C_{2v}(2)$	0.56	0.1800
	D_3	0.55	0.1803
	$D_{3h}(2)$	0.53	0.1810
C_{80}	C_s	2.05	0.1766
	C_{2v}	1.22	0.1772
	D_2	0.59	0.1781
	D_{5d}	0.57	0.1788

Table 8. The third-order optical nonlinearities of functionalized carbon nanostructures. MTMDA = monotrimethylenediamine. C_{60}/Si = $C_{60}[NH_2\text{-}(CH_2)_3\text{-}Si(OC_2H_5)_3]_5$. C_{60}/Fe1 = $C_{60}[Fe(NH_2CN)_5](NO_3)_3$. C_{60}/Fe2 = $C_{60}[Fe((NH)_2CNCN)_5](NO_3)_3$. C_{60}/Ru1 = $C_{60}[Ru_2(NH_2CN)_8]Cl_6$. C_{60}/Ru2 = $C_{60}[Ru_2((NH)_2CNCN)_8]Cl_6$. C_{60}/γ-CD = C_{60}/γ-cyclodextrin.

Material	Method	λ (μm)	τ	Material of State	$\chi^{(3)}$ (10^{-12} esu)	γ (10^{-34} esu)	Ref.
C_{60}	CARS			Solution		24 ± 10	[351]
C_{70}	CARS			Solution		40 ± 10	[352]
C_{84}	CARS			Solution		76 ± 5	[352]
$C_{60}Pt_2$	OHD-OKE	0.647	165 fs	Film	−0.5		[387]
$C_{60}Pd_2$	OHD-OKE	0.647	165 fs	Film	−0.8		[387]
$C_{60}Sm_2$	OHD-OKE	0.647	165 fs	Film	−2.6		[387]
Li@C_{60}	DFWM	0.497	0.5 ps	Solution	0.07 ± 0.02	-1800 ± 450	[382]
TPN:C_{60} multilayer	THG	1.907	13 ns	Film	7		[380]
TPN:C_{60} composite	THG	1.907	13 ns	Film	5		[380]
TPP:C_{60} multilayer	THG	1.907	13 ns	Film	8		[380]
TPP:C_{60} composite	THG	1.907	13 ns	Film	12		[380]
Poly-C_{60}	THG	1.907	13 ns	Film	48		[380]
C_{60}:O_2	THG	1.907	13 ns	Solution	46		[380]
C_{60}-poly-aminonitrile	OKE	0.810	120 fs	Solution		32	[303]
$C_{60}(NH_2CN)_5$	OKE	0.830	120 fs	Solution	0.06	100	[390]
$C_{60}[(NH)_2CNCN]_5$	OKE	0.830	120 fs	Solution		350	[390]
C_{60}/Fe1	OKE	0.830	120 fs	Solution	0.03	72	[391]
C_{60}/Fe2	OKE	0.830	120 fs	Solution		250	[391]
C_{60}/Ru1	OKE	0.830	120 fs	Solution		2400	[391]
C_{60}/Ru2	OKE	0.830	120 fs	Solution		6500	[391]
C_{60}/Si	OKE	0.820	200 fs	Sol	0.1		[402]
C_{60}/Si	OKE	0.820	200 fs	Gel	0.2		[402]
C_{60}/γ-CD (1 day)	Z-scan	0.532	10 ns	Solution		13000	[384]
C_{60}/γ-CD (3 days)	Z-scan	0.532	10 ns	Solution		286000	[384]
C_{60}/γ-Cd (7 days)	Z-scan	0.532	10 ns	Solution		507000	[384]
CuPc-C_{60}	OKE	0.830	120 fs	Solution		5400	[385]
C_{60}-TTF-α	DFWM	0.532	ps	Solution	3		[392]
C_{60}-TTF-β	DFWM	0.532	ps	Solution	5		[392]
C_{60}-(*PS*)$_n$ (900 Da)	OKE	0.800	100	Solution		2400	[389]
C_{60}-(*PS*)$_n$ (17800 Da)	OKE	0.800	100	Solution		880	[389]
C_{60}-(*PS*)$_n$ (4800 Da)	OKE	0.800	100	Solution		660	[389]
C_{60}-PTHFMA	OKE	0.640	fs	Solution		8900	[378]
MTMDA-C_{60}	DFWM	0.532	15 ns	Solution	53.9	2480	[381]
C_{60}(*biphenyl*)$_{x-3}$	Z-scan			Solution	10^3		[386]
C_{70}-poly-aminonitrile	OKE	0.810	120 fs	Solution		160	[383]
$C_{70}(NH_2CN)_5$	OKE	0.810	120 fs	Solution		410	[388]
$C_{70}[(NH)_2CNCN]_5$	OKE	0.810	120 fs	Solution		580	[388]
Dy@C_{82}	DFWM	0.532	70 ps	Solution		30000 ± 5000	[379]
Er_2@C_{82}	DFWM	1.064				−870	[377]

Table 9. SSH-SOS theoretical static γ (in the unit of 10^{-32} esu) tensor components for C_{70}, $C_{60+(9,0)@10}$, $C_{60+(5,5)@10}$, and C_{144} (uncapped zigzag tubule). TD-DFT/B3LYP/3-21G results for finite and open (4,4) and (5,5) carbon nanotubes are listed (adapted from Refs. [25, 370]).

γ	C_{60}	C_{70}	$C_{60+(9,0)@10}$	$C_{60+(5,5)@10}$	C_{144}	(4,4)	(5,5)
γ_{xxxx}	4.8	5.1	13.3	14.8	141.3	0.059	0.099
γ_{yyyy}	4.8	5.1	13.3	14.8	141.3	0.059	0.098
γ_{zzzz}	5.3	6.3	312.0	278.3	2566.8	0.856	0.658
γ_{xxyy}	1.3	1.6	4.6	5.0	44.7		
γ_{xxzz}	2.7	3.1	19.4	6.2	132.3		
γ_{zzxx}	2.7	3.1	19.4	6.2	132.3		
γ	5.6	6.4	85.0	68.6	697.6	0.307	0.494

Table 10. The static γ value and average contribution Γ of 17 chiral carbon nanotubes and seven well-characterized polyenic polymers, where S-SWNT$_{(n,m)}$ and M-SWNT$_{(n,m)}$ denote semiconducting and metallic SWNT, respectively, d_t and ϕ are the diameter and chiral angle of carbon nanotubes; N is the total number of carbon atoms in a given material; Γ is the average contribution of a carbon atom to the third-order optical nonlinearity of the material; ζ is the ratio between the Γ values of a material and C_{60} (some data are adapted from Xie [24], Xie et al. [25]).

Materials	N	d_t (Å)	ϕ (degree)	γ (10^{-32} esu)	Γ (10^{-32} esu)	ζ
C_{60}	60	7.0	—	561.2	0.09	1.0
S-SWNT$_{(16,5)}$	364	7.5	27	63.56	0.175	1.9
S-SWNT$_{(9,3)}$	364	7.5	5	52.00	0.143	1.5
S-SWNT$_{(8,3)}$	388	7.7	15	54.01	0.139	1.5
S-SWNT$_{(9,2)}$	412	8.0	10	57.20	0.139	1.5
S-SWNT$_{(7,5)}$	436	8.2	25	60.17	0.138	1.5
S-SWNT$_{(8,4)}$	448	8.3	19	61.77	0.138	1.5
S-SWNT$_{(10,2)}$	496	8.7	9	67.05	0.135	1.4
S-SWNT$_{(7,6)}$	508	8.8	28	68.50	0.135	1.4
S-SWNT$_{(9,4)}$	532	9.0	18	69.16	0.130	1.4
S-SWNT$_{(10,3)}$	556	9.2	13	68.81	0.124	1.3
S-SWNT$_{(8,6)}$	592	9.5	25	68.16	0.115	1.2
S-SWNT$_{(9,5)}$	604	9.7	21	67.64	0.112	1.2
M-SWNT$_{(7,4)}$	372	7.6	21	230.00	0.618	6.6
M-SWNT$_{(10,1)}$	444	8.3	5	250.40	0.564	6.0
M-SWNT$_{(9,3)}$	468	8.5	14	221.88	0.474	5.1
M-SWNT$_{(8,5)}$	516	8.9	22	181.60	0.352	3.8
M-SWNT$_{(10,4)}$	624	9.8	16	148.83	0.239	2.6
Polyenic polymer	230	—	—	35	0.15	1.6
Polyenic polymer	340	—	—	54	0.16	1.7
Polyenic polymer	450	—	—	78	0.17	1.8
Polyenic polymer	620	—	—	155	0.25	2.7
Polyenic polymer	830	—	—	267	0.32	3.4
Polyenic polymer	1460	—	—	629	0.43	4.6
Polyenic polymer	2320	—	—	854	0.37	4.0

Table 11. Third-order optical nonlinearities of MWNTs and SWNTs. λ and τ are the wavelength and pulse duration of laser source; Γ is the averaged second-order hyperpolarizability γ contributed by one carbon atom in carbon nanotubes; PPCE denotes the mixing of polypyrrol and *m*-cresol with ethanol; DMF = dimethylformamide (adapted from Xie et al. [25]).

Material	Method	λ (μm)	τ	State	$\chi^{(3)}$ (10^{-12} esu)	Γ (10^{-34} esu)	Ref.
MWNT	DFWM	1.064	30 ps	PPCE solution	0.07	0.06	[355, 356]
	DFWM	0.532	30 ps	PPCE solution	0.06	—	[355, 356]
	DFWM	1.064	8 ns	PPCE solution	12	—	[355, 356]
	DFWM	1.064	8 ns	PPCE solution	3	—	[355, 356]
SWNT	OKE	0.82	120 fs	DMF solution	0.4	77	[357, 358]
	DFWM	0.532		Thick film		9 ± 5	[372]

experiments if care was taken to select experimental results without significant resonant contributions. Also, the vibrational contributions to the NLO properties of isolated C_{60} have been studied and shown to be significant for certain processes [333]. In Section 4.3, we shall see that a strong enhancement on third-order optical nonlinearity—for example, an increase from several decades to 100 times on the γ value—was even observed as C_{60} was chemically modified to form a charge-transfer complex [218, 303] or was chemically reduced to anions [351, 352].

A three-level model was proposed by Kajzar et al. [198] to explain the THG spectrum of C_{60} and C_{70}. Shuai and Bredas [312, 323] used the VEH approach to investigate the electronic structures of C_{60} and C_{70} and applied the SOS method to study the dynamical NLO spectrum of C_{60} and C_{70}. Their calculations are fully consistent with the EFISHG [218], THG [218], and DFWM [280, 285] measurements, but about three to four orders of magnitude lower than the data reported by Blau et al. [278] and Yang et al. [290]. The lowest two-photon and three-photon resonances in C_{70} were observed at almost the same frequencies for C_{60} because of the symmetry [312, 323].

The second hyperpolarizabilities γ of higher fullerenes were found by extensive theoretical studies to scale with the mass of the all-carbon molecule [306, 310, 314–321]. The free-electron gas model proposed by Knize [306] showed that the γ value of higher fullerene increases as the cube of the number of carbon atoms. Using the SOS-INDO/CI approach [314] or sum-over-molecular-orbitals (SOMO) approach, Fanti and colleagues [314, 315] found that the second-order hyperpolarizabilities of fullerenes are a nonlinear function of the number of carbon atoms. Later on, the extended Su-Schrieffer-Heeger model proposed by Xie and Rao [316] and Xie [317] indicated that the static γ values of armchair and zigzag tubular fullerenes of small size scale with about three powers of the number of carbon atoms. Using the SOS method, Harigaya [318–321] predicted that the off-resonant third-order NLO susceptibilities of higher fullerenes is nearly proportional to the fourth power of the carbon number when the on-site Coulomb repulsion is $2t$ or $4t$, t being the nearest-neighbor hopping integral. The work of both Xie and Rao [316] and Xie [317] and Harigaya [318–321] demonstrated the important roles of Coulomb interactions in higher fullerenes. Luo [310] pointed out that if C_{60} is excluded, a perfect power law dependence of the γ value on the number N of carbon atoms can be observed [310]: $\gamma = 1.1 \times 10^{-4} N^{2.4}$ for the data of Fanti et al. [315]; $\gamma = 3.2 \times 10^{-2} N^{0.75}$ for the data of Jonsson et al. [296]. Luo's work [310] reflected the important issue that C_{60} has the most exceptional electron localization among all fullerenes. In comparison with the experiment of Huang et al. [305], the theoretical predicted increase trend is somewhat lower than the measured one. This implies the existence of other important factors (for example, geometrical structure, π-π overlap discussed later, and resonance enhancement) contributed to the increased γ values for higher fullerenes.

Higher ($N > 70$) and lower ($N < 60$) fullerenes have many isomers with specific symmetries (see Fowler and Manlopoulos's *An Atlas of Fullerenes* [20]). It has been an interesting issue to discuss the effects of symmetries and atom arrangements on the NLO of fullerenes. As shown in Fig. 4(e–i), C_{78} has five topologically distinct structures: two with C_{2v} symmetry, two with D_{3h} symmetry, and one with D_3 symmetry. Using the SOS method, Wan, Dong, and Xing [324] demonstrated numerically the big effects of symmetry and atom arrangement on the γ values of five isomers of C_{78}. From a geometric point of view, Moore and colleagues [311, 322] analyzed the static γ values of C_{60}, C_{70}, five isomers of C_{78}, and two isomers of C_{84} in terms of the geometry symmetry (characterized by the molecular group order), aromaticity (characterized by the selection of six-member rings determined from the analysis of bond lengths), and molecular size (characterized by maximum interatomic distance and surface areas) by using the FF theoretical technique, a semiempirical Hamiltonian, and molecular structures obtained from DFT calculations. The γ value is affected by a combination of the number of aromatic rings, length, and group order, in decreasing importance.

The RPA method has been shown to give excellent agreement for optical band gaps and NLO properties of larger, conjugated molecules [3, 92, 93, 99, 100]. Correlation effects are therefore expected to be small for these conjugated molecules. Based on this method,

Jonsson et al. [296] calculated the static γ values of C_{60}, C_{70} and C_{84}, which are determined to be are 0.55, 0.75, 0.82×10^{-34} esu, respectively. Their results are of near Hartree-Fock limit quality.

An important factor governing the optical nonlinearity is the extended delocalized π-electron cloud along the effective conjugation length or sphere of the π-conjugated system. Recently, Lin, Sheu, and Lee [326] have demonstrated the crucial role of π-π overlap in determining the second-order hyperpolarizability γ in the π-conjugated framework, for example, by taking the semiempirical Austin model 1 (AM1) quantum chemical calculations on various isomers of higher fullerenes. For different isomers of the same mass of fullerene, different curvatures are exhibited and indicate that the π-π overlap is different among the isomeric fullerenes. Table 7 summarizes the orientationally averaged second-order hyperpolarizability $\bar{\gamma}$ and the averaged pyramidalization angle $\bar{\theta}_p$, which describes the deviation from the planar sp^2 carbon and is defined as $\theta_{\sigma\pi} - \pi/2$ ($\theta_{\sigma\pi} = \pi/2$ for planar sp^2 carbons and 0.608π for tetrahedral sp^3 centers). For isometric fullerenes of the same mass, they found an excellent correlation between $\bar{\gamma}$ and $\bar{\theta}_p$: $\bar{\gamma}$ decreases with the increase of $\bar{\theta}_p$. Hence, a minor change of average π-π overlaps in fullerenes has a large effect on the average second-order hyperpolarizability.

In summary, Table 6 collects the static γ values of some fullerenes calculated by different theoretical techniques. A direct comparison between these theoretical results shows significant differences. All of the SOS approaches give a second hyperpolarizability, which is about an order of magnitude larger than that predicted by the *ab initio* calculations. This discrepancy arises from the truncation of the expansion in excited states in the explicit summation of contributions to γ in SOS calculations [296], which appears to neglect important negative contributions to the second hyperpolarizability. The SOS methods predict larger ratios between the second hyperpolarizabilities for higher fullerenes than do both FF and *ab initio* calculations. The danger of this truncation of the sum over excited states is most clearly illustrated by this fact that γ_{zzzz} remains almost constant for the first 180 excited states included and then increased [296]. In contrast, the γ values obtained in the FF methods [294, 311] are of the same order of magnitude. The described problems are absent in the analytical RPA approach [296]. On the other hand, all these theoretical studies do not consider the screening effect induced by electron–electron interaction (in TD-DFT, this is to some extent taken into account). Actually, the screening effect should be considered because the polarization of the electron cloud can modify the charge density and therefore the self-consistent potential seen by each electron. Detailed analysis of the screen effect is expected.

4.2. Rod-Shaped Carbon Nanostructures

4.2.1. Theoretical Calculations

Compared with the extensive NLO studies of spherically shaped carbon nanostructures, less work [307–309, 316, 317, 353–366] has been done on the third-order optical nonlinearities (TOON) of rod-shaped carbon nanostructures. Xie [309] and Xie and Jiang [307, 308] were the first to study theoretically the TOON of SWNT by using a tight-binding method. In the following, we review this method.

The Coulomb interaction effect plays an important role in a physical understanding of the electronic structures and properties of carbon nanostructures [24]. By including the Coulomb interaction, Xie and Jiang [307–309] have extended the SSH model to describe carbon nanostructures such as higher fullerenes and carbon nanotubes. The total Hamiltonian can be written as follows [307–309]:

$$H = \sum_{\langle ij \rangle} \sum_{s} (-t_0 - \alpha_0 y_{ij})(c_{i,s}^{\dagger} c_{j,s} + h.c.) + \frac{k_0}{2} \sum_{\langle ij \rangle} y_{ij}^2 + u_0 \sum_{i} c_{i,\uparrow}^{\dagger} c_{i,\uparrow} c_{i,\downarrow}^{\dagger} c_{i,\downarrow} + v_0 \sum_{\langle ij \rangle} \sum_{s,s'} c_{i,s}^{\dagger} c_{i,s} c_{j,s'}^{\dagger} c_{j,s'} \tag{46}$$

where the sum $\langle ij \rangle$ is taken over the nearest neighbors for the c—c bond; t_0 represent the hopping integrals for the c—c bond; α_0 is the electron-phonon coupling constants related to

the c—c bond; k_0 is the spring constants corresponding to the c—c bond; y_{ij} is the change of the bond length between the ith and jth atoms; the operator $c_{i,s}$ ($c^{\dagger}_{i,s}$) annihilates (creates) a π electron at the ith atom with spin s ($s = \uparrow, \downarrow$); u_0 is the usual on-site Coulomb repulsion strength; and v_0 is the Coulomb interaction between the nearest and next-nearest atoms. Using the Hartree-Fock approximation, Xie and Jiang [307–309] transformed that equation into

$$H = \sum_{\langle ij \rangle}\sum_{s}(-t_0 - \alpha_0 y_{ij})(c^{\dagger}_{i,s}c_{j,s} + h.c.) + \frac{k_0}{2}\sum_{\langle ij \rangle} y_{ij}^2 + u_0\sum_{i}\left(\sum_{s}\rho_{i,s}c^{\dagger}_{i,s}c_{i,s} - \rho_{i,\uparrow}\rho_{i,\downarrow}\right)$$
$$+ v_0\sum_{ij}\sum_{s}\left(\sum_{s'}\rho_{j,s'}c^{\dagger}_{i,s}c_{i,s} - \rho_{i,s}\sum_{s'}\rho_{j,s'} - \tau_{ij,s}c^{\dagger}_{i,s}c_{j,s} + \tau^2_{ij,s}\right) \tag{47}$$

where $\rho_{i,s} = \langle c^{\dagger}_{i,s}c_{i,s}\rangle$ is the electron density and $\tau_{ij,s} = \langle c^{\dagger}_{i,s}c_{j,s}\rangle$ is the bond order parameter. This equation is solved by the adiabatic approximation for phonons. The Schrödinger equation for the π electron is

$$\epsilon_k Z_{k,s} = \sum_{\langle ij \rangle}(-t - \alpha_0 y_{ij} - v_0\tau_{ij,s})Z_{k,s}(j) + \left[u_0\rho_{i,s} + v_0\sum_{s'}\sum_{j}\rho_{j,s'}\right]Z_{k,s}(i) \tag{48}$$

where ϵ_k is the kth eigenvalue. The self-consistent equation for the lattice is

$$y_{ij} = -2\alpha k_0^{-1}\sum_{k,s}{}'\left\{Z_{k,s}(i)Z_{k,s}(j) - \Pi^{-1}\sum_{\langle ml \rangle} Z_{k,s}(m)Z_{k,s}(l)\right\} \tag{49}$$

where the prime denotes the sum over all occupied states, the second term originates from the constraint condition $\sum_{\langle ij \rangle} y_{ij} = 0$, and Π is the number of π bonds. Then, the electron eigenstates $z_{k,s}(i)$, eigenenergies ϵ_k, and bond variables y_{ij} can be obtained by performing the self-consistent iteration.

The third-order nonlinear optical polarizability γ can be expressed as follows: Within the independent electron approximation and the SOS approach discussed by Orr and Ward [170, 367], the second-order hyperpolarizability γ for the THG process can be rewritten (see Fig. 2 of Ref. [323], taking into account six diagrams in total) [323, 368, 369] as:

$$\gamma(-3\omega;\, \omega, \omega, \omega) = \gamma_2 + \gamma_3 + \gamma_4 + \gamma_7 + \gamma_8 \tag{50}$$

where γ_i is given in detail next [312]:

$$\gamma_2(-3\omega;\, \omega, \omega, \omega) = 2\sum_{a\in\text{unocc}}\sum_{i,j,k\in\text{occ}}\mu_{ai}\mu_{ij}\mu_{jk}\mu_{ka}S_2(\omega) \tag{51}$$

Hereafter, a, r, and s denote unoccupied molecular levels, and i, j, and k denote occupied ones. The $S_2(\omega)$ function is defined as [312]

$$S_2(\omega) = \frac{1}{(\epsilon_{ai} - 3\omega)(\epsilon_{aj} - 2\omega)(\epsilon_{ak} - \omega)} + \frac{1}{(\epsilon_{ai} + \omega)(\epsilon_{aj} + 2\omega)(\epsilon_{ak} - \omega)}$$
$$+ \frac{1}{(\epsilon_{ai} + \omega)(\epsilon_{aj} - 2\omega)(\epsilon_{ak} - \omega)} + \frac{1}{(\epsilon_{ai} + \omega)(\epsilon_{aj} + 2\omega)(\epsilon_{ak} + 3\omega)} \tag{52}$$

The other terms are given in detail as follows [312]:

$$\gamma_3(-3\omega;\, \omega, \omega, \omega) = 2\sum_{a,r,s\in\text{unocc}}\sum_{i\in\text{occ}}\mu_{ia}\mu_{ar}\mu_{rs}\mu_{si}S_3(\omega) \tag{53}$$

where $S_3(\omega)$ is similar to $S_2(\omega)$ with the substitution of ϵ_{aj} by ϵ_{rj}, and ϵ_{ak} by ϵ_{sj};

$$\gamma_1(-3\omega;\, \omega, \omega, \omega) = -2\sum_{a,r\in\text{unocc}}\sum_{i,j\in\text{occ}}\mu_{ia}\mu_{ar}\mu_{ji}\mu_{rj}S_1(\omega) \tag{54}$$

where $S_1(\omega)$ is analogous to $S_2(\omega)$ with the substitution of ϵ_{aj} by ϵ_{ri}, and ϵ_{ak} by ϵ_{rj}:

$$\gamma_4(-3\omega;\omega,\omega,\omega) = -2 \sum_{a,r\in\text{unocc}} \sum_{i,j\in\text{occ}} \mu_{ia}\mu_{ji}\mu_{ar}\mu_{rj}S_4(\omega) \tag{55}$$

where $S_4(\omega)$ is analogous to $S_2(\omega)$ with the substitution of ϵ_{ak} by ϵ_{rj}:

$$\gamma_7(-3\omega;\omega,\omega,\omega) = -2 \sum_{a,r\in\text{unocc}} \sum_{i,j\in\text{occ}} \mu_{ia}\mu_{rj}\mu_{aj}\mu_{ri}S_7(\omega) \tag{56}$$

where $S_7(\omega)$ is analogous to $S_2(\omega)$ with the substitution of ϵ_{aj} by $\epsilon_{ai}+\epsilon_{rj}$, and ϵ_{ak} by ϵ_{aj}

$$\gamma_8(-3\omega,\omega,\omega,\omega) = -2 \sum_{a,r\in\text{unocc}} \sum_{i,j\in\text{occ}} \mu_{ia}\mu_{rj}\mu_{aj}\mu_{ri}S_8(\omega) \tag{57}$$

where $S_3(\omega)$ is analogous to $S_2(\omega)$ with the substitution of ϵ_{aj} by $\epsilon_{ai}+\epsilon_{rj}$, and ϵ_{ak} by ϵ_{ri}. The double excitation channels $\gamma_7+\gamma_8$ can be cast into a single term after simple algebra by noting $\epsilon_{ai}+\epsilon_{rj}=\epsilon_{aj}+\epsilon_{ri}$ [312]:

$$\begin{aligned}\gamma_7+\gamma_8 = -2 \sum_{a,r\in\text{unocc}} \sum_{i,j\in\text{occ}} \mu_{ia}\mu_{rj}\mu_{aj}\mu_{ir} \\ \times\left\{\frac{1}{(\epsilon_{ai}-3\omega)(\epsilon_{aj}-\omega)(\epsilon_{ri}-\omega)}+\frac{1}{(\epsilon_{ai}+\omega)(\epsilon_{rj}+\omega)(\epsilon_{aj}+3\omega)}\right. \\ \left.+\frac{1}{(\epsilon_{ai}+\omega)(\epsilon_{rj}+\omega)(\epsilon_{ri}-\omega)}+\frac{1}{(\epsilon_{ai}+\omega)(\epsilon_{aj}-\omega)(\epsilon_{ri}-\omega)}\right\}\end{aligned} \tag{58}$$

In these formulae, ϵ_p is the one-electron energy, ϵ_{pn} $(=\epsilon_p-\epsilon_n)$ is the transition energy, and μ_{mn} is the transition matrix element between one-electron states

$$\mu_{mn} = \sum_{js} Z^*_{n,s}(j)(-e\mathbf{r})Z_{ms}(j) \tag{59}$$

In the numerical calculations, a lifetime broadening factor η $(=1.6\times10^{-2}$ eV) is included. Because the ratios between different components of γ are not known, a spatial average of γ is given by

$$\gamma = \frac{\gamma_{xxxx}+\gamma_{yyyy}+\gamma_{zzzz}+2\gamma_{xxyy}+2\gamma_{yyzz}+2\gamma_{zzxx}}{5} \tag{60}$$

Based upon the electronic structure obtained in the extended Su-Schrieffer-Heeger (ESSH) model, Xie [24] calculated the second-order hyperpolarizabilities of armchair and zigzag nanotubes. In the actual calculation, the z axis in the armchair tubule is taken along the direction from the bottom pentagon to the top one, and the z axis in zigzag tube is taken along the direction from the bottom hexagon to the top one. If the length of every bond is given, the coordinates of every atom will be obtained. However, the γ magnitude is not sensitive to small changes of atomic coordinates. So, for simplicity, the same bond length $(=1.4225$ Å, which is the average bond length of C_{70} given by the experiment for all rod-shaped carbon nanostructures) is assumed. The coulomb interaction is assumed to be $u_0=2v_0=t$, which is not strong, and thus the effective hopping integral is the same as that in the free-electron case. Because three parameters, t_0, α_0, and k_0, in the ESSH model do not sensitively depend on the shape and size of the nanostructure, Xie [24] took $t_0=2.5$ eV, $\alpha_0=6.31$ eV/Å, and $k_0=49.7$ eV/Å^2, which are the same as those in C_{60} and C_{70}.

The theoretical static γ magnitude of several armchair and zigzag tubular structures and C_{60} are listed in Table 12. For C_{60}, the difference between z and x (or y) components of γ is very small because C_{60} is almost a sphere. But this difference is pronounced for large tubular structures, which contain a large number of carbon atoms. Compared with C_{60}, SWNT has a larger TOON, which would have potential applications as an ultrafast

Table 12. The ratio $q = \gamma^{doped}/\gamma^{pure}$ of several doped armchair nanotubes $C_{60+10k}X$ and doped zigzag nanotubes $C_{6k}X$ ($X = B$, N), where γ^{doped} is the calculated static γ value of the doped nanotube and γ^{pure} is the static γ value of the corresponding pure nanotube and given by an empirical formula for armchair and zigzag nanotubes (adapted from Xie et al. [24, 469]).

	$C_{60}X$	Armchair Nanotube				Zigzag Nanotube			
X	$k = 0$	$k = 1$	$k = 2$	$k = 9$	$k = 18$	$k = 1$	$k = 2$	$k = 5$	$k = 10$
N	30.5	30.7	32.4	36.8	41.2	31.1	33.2	36.0	40.4
B	3.9	4.3	4.9	7.6	8.5	4.7	5.1	6.9	7.7

optical switch. Then, empirical relations for the static second-order hyperpolarizabilities γ of both armchair and zigzag nanotubes are established [316, 317, 353]: $\gamma = (1 + 0.3n)^{2.98}\gamma_{C_{60}}$ for the zigzag tube and $\gamma = (1 + 0.167n)^{3.15}\gamma_{C_{60}}$ for the armchair tube, where $\gamma_{C_{60}}$ is the static γ value of C_{60}. The average contribution Γ of one carbon atom to the third-order optical nonlinearity of each chiral carbon nanotube, as listed in Table 10, is examined by Xie and Rao [354]. The chiral effect on the TOON is clearly shown. This work demonstrated that the metallic tube favors larger γ values. In particular, the smaller the diameter of a chiral carbon nanotube, the larger the average contribution Γ. Compared with that of a well-characterized polyenic polymer listed in Table 10, chiral carbon nanotubes can compete with the conducting polymer, achieving a large γ value that is needed for photonic applications.

In the synthesis of carbon nanotubes, their caps may be destroyed partially or completely. Such effects will greatly influence the geometric and electronic structures of carbon nanotubes. Surely, the change of these structures will have a large effect on the NLO properties of carbon nanotubes. As an example, Xie [24] studied the static γ magnitude of uncapped zigzag tubular structure C_{144}. The results are shown in Table 12. The static γ magnitude is about 10 times larger than that of capped zigzag tubular structure $C_{60@5\times18}$. Compared with the symmetry effect, the cap effect is more obvious. The reason is addressed as follows: The NLO response for the C_{60} system is mainly produced by delocalized π electrons as in conjugated polymer chains. However, the three-dimensional character of C_{60} causes severe limitations on its NLO property and thus makes its γ magnitude become about two orders of magnitude smaller than those of linear polymers containing a similar number of carbon atoms. For a capped tubular structure, a π electron on a site can transfer to the site's three neighbors. If both caps are cut, a π electron on the site at the edge of a cylinder can transfer only to the site's two neighbors. Obviously, this kind of edge effect will reduce the effective space dimension of π electrons and thus enhance the static γ magnitude of tubular fullerenes.

In principle, *ab initio* quantum chemical calculations would be a valuable tool to compute the second hyperpolarizability of carbon nanostructures, but currently they are too computer-demanding to be used systematically to study the NLO of these structures. Recently, Wu et al. [370] optimized finite open SWNT (4,4) and (5,5) carbon nanotubes by using B3LYP/3-21G. Combined with TD-DFT with B3LYP hybrid functional and SOS approach, they studied the dynamical third-order optical polarizabilities in the THG, EFISHG, and DFWM optical processes. The static γ values for (4,4) and (5,5) SWNTs are listed in Table 12. The average γ values predicted by TD-DFT are about two orders of magnitude smaller than those predicted by the tight-binding methods. They found that the largest third-order polarizability is in the direction of the polarized and basic light along the tube axis. This is in agreement with the prediction of the tight-binding method.

The scaling law of γ predicted by the tight-binding method is based on small carbon nanotubes, but the saturation limit was not examined. Recently, based on the atomic dipole interaction model introduced in Section 2, Jensen, Astrand, and Mikkelsen [371] calculated the second hyperpolarizability of carbon nanotubes on a length up to 75 nm. This work demonstrated that an atomic representation of mesoscale systems—for example, carbon nanotubes—can be used to obtain a cubic response property up to a size of the system where the property scales linearly with increasing size. This model was demonstrated to be useful for designing nonlinear molecular materials, where local modifications may give large

macroscopic contributions. Jensen, Astrand, and Mikkelsen [371] found that carbon nanotubes are comparable with conjugated polymers with respect to the magnitude of the second hyperpolarizability γ. This conclusion is consistent with that predicted by the tight-binding approach.

4.2.2. Experimental Measurement

The extensive theoretical studies showed that higher fullerenes and carbon nanotubes can compete with polymers for third-order optical applications and stimulated many experimental groups to measure nonlinear optical properties of higher fullerenes and (doped) carbon nanotubes. As reviewed in this section, the experimental results are in accord with theoretical predictions.

Liu et al. [355] and Xie et al. [356] were the first to experimentally investigate the third-order optical nonlinearities of carbon nanotubes by using the picosecond and nanosecond Nd:YAG laser and the technique of backward DFWM. Their results are listed in Table 11. They found enhancement of the TOON in the carbon nanotubes, in agreement with the theoretical predictions of Xie and colleagues [307–309, 316, 317, 353, 354]. The two-photon and one-photon processes, as demonstrated in the theoretical work of Xie and colleagues, contribute to the TOON. The fast response process of the carbon nanotubes on the picosecond scale is an instantaneous electronic process because of the large polarizability arising mainly from the $\pi \rightarrow \pi^*$ virtual transition, whereas the slow process on the picosecond scale is associated with an excited-state population. On the nanosecond scale, similar results are obtained.

However, Liu and colleagues [355] did not obtain a real solution for carbon nanotubes. Thus, a large absorption was observed in the infrared region. On the other hand, their DFWM technique had limited ability to distinguish the NLO contribution from the π-conjugated electrons. Recently, Shi and colleagues [357, 358] performed NLO measurements on real SWNT solutions by using the femtosecond optical Kerr technique with a Ti:Sapphire laser operating at 820 nm. Their results are listed in Table 11. The magnitude of Γ for each carbon atom in the SWNT is about 7.7×10^{-33} esu, which is about three orders of magnitude larger than the carbon nanotube value reported by Liu and colleagues [355]. To improve NLO performance of SWNTs, it would be of interest to test the optical nonlinearity of SWNTs of various lengths.

The measurement of Liu and colleagues [355] was made on carbon nanotube solution. Recently, Botti et al. [372] studied the third-order NLO response of thick film of SWNTs, produced without catalyst and deposited onto quartz substrate, by performing DFWM measurement. The second-order hyperpolarizability γ was estimated to be as large as 1.6×10^{-29} esu, and the average contribution of each carbon atom in the carbon nanotube is $\Gamma = (9 \pm 5) \times 10^{-34}$ esu, about one order of magnitude larger than our theoretical calculations. This experiment indicates that further investigations are required to understand the inadequacy of current theories in modeling results of experiments performed at different wavelengths.

Recently, using ultrashort pulse generated by a Cr:Forsterite laser, at a wavelength of 1.250 μm, Stanciu et al. [373] studied THG generation from solid samples of carbon nanotubes. The results show an unusual nonperturbative behavior of the THG yield, for relatively low-input laser fields, of about 10^{10} W/cm^2. This strong nonlinearity of the laser interaction with carbon nanotubes is also confirmed in a full quantum-mechanical theory, developed as an improvement of the semiclassical models, for harmonic generation from a SWNT excited by intense linearly polarized radiation. Physically, this fact shows that the interaction of carbon nanotubes with an intense laser pulse cannot be described by a perturbation approach, even for relatively low laser intensities.

As reviewed previously, theoretical and experimental studies demonstrated that carbon nanotubes could have a very strong third-order optical nonlinearity and thus would have great potential in photonic applications, especially as an ultrafast optical switch, which is a crucial component for future high-bit-rate time-division-multiplexing optical communication systems or free-space optical-digital computing systems. So far, many types of ultrafast all-optical switches have been studied and demonstrated, using optical nonlinearities in optical

fibers and semiconductor materials. Because of the limits imposed by the properties of materials used for all-optical switches, it is very hard to achieve subpicosecond all-optical switches at the optical communication wavelength of 1550 nm [374, 375]. Recently, using standard time-resolved pump-probe photomodulation technique with a 150-fs fiber laser at the wavelength 1550 nm, Chen et al. [359] experimentally demonstrated that SWNTs have an exciton decay time of less than 1 ps and exhibit a high third-order optical nonlinearity. They also observed similar experimental results from MWNTs or a SWNT/polyvinyl alcohol composite. Their experiment suggested that SWNT/polymide composites have the potential to become an ultrafast waveguide switch and to develop high-quality subpicosecond all-optical switches. Later, Han et al. [376] measured the NLO of well-separated SWNTs grown into the voids of an ordered array of silica spheres and obtained the nonlinear decay time of a 0.8-μm probe at 320 fs as the same was pumped with 0.4-μm light. This experiment demonstrated again the strong and fast NLO effects of carbon nanotubes, making these nanostructures good candidates for all-optical switches.

4.3. Functionalized Carbon Nanostructures

4.3.1. Experimental Measurements

Spherically shaped carbon nanostructures such as C_{60}, C_{70}, and C_{82} are well-known electron acceptors, and thus they easily form charge transfer complexes or, say, fullerene derivatives, with other organic groups or metals, for example, complex formed by C_{60} and polymer matrix, or trapping metal to C_{82}. Over the past 10 years, third-order optical nonlinearities of fullerene derivatives have been studied by different research groups and different experimental techniques, for example, coherent anti-Stokes Raman spectroscopy (CARS) analysis technique, optical Kerr effect (OKE), THG, DFWM, Z-scan, and optically heterodyned OKE (OHD-OKE). These studies demonstrated that fullerene derivatives, because of the charge transfer, do exhibit enhanced third-order optical nonlinearities at nanosecond or picosecond scales [303, 351, 352, 377–392]. The third-order optical nonlinearities of some representative fullerene derivatives are summarized in Table 8. It is seen that the chemical-modified fullerenes all exhibit enhancement on their third-order optical nonlinearities compared with small host fullerene molecules. The enhancement is greatly influenced by the measuring wavelength and pulse duration. The negative or positive $\chi^{(3)}$ or γ can be understood as follows: There are two terms in the SOS representation of the nonlinearity γ_{ijkl} [393], the first one representing four-wave mixing pathways that involve two-photon transitions to or between excited electronic states, and the second one describing pathways that involve transitions that return to the ground state. The final value of $\chi^{(3)}$ or γ depends on the interference between the two terms. If the first term dominates, $\chi^{(3)}$ or γ will be positive; they will be negative if the second term dominates.

Lascola and Wright [351, 352] determined the second-order molecular hyperpolarizability γ for the charged species, C_{60}^{-}, C_{60}^{2-}, and C_{60}^{3-}, by using CARS vibrational lineshape analysis. Their measured γ values for those charged molecules are larger than those of the neutral C_{60} and are comparable to values of highly conjugated organic polymers, for example, polydiacetylene. Heflin et al. [377]; Gu et al. [379]; Campbell et al. [382]; Mavritsky et al. [394]; Qian et al. [387]; and Huang et al. [385] have reported $\chi^{(3)}$ measurement of metal-containing endohedral or exohedral fullerenes, for example, $Er_2@C_{82}$ [377] by DFWM at 1064 nm, DyC_{82} by DFWM at 532 nm, LiC_{60} by z-scan DFWM at 497 nm, $(Ph_3P)_2PtC_{60}$ [394] by time-resolved z-scan technique at 527.5 nm, $C_{60}M_2$ (M=Pd, Pt, Sm) by OHD-OKE at 647 nm, and CuPc-C_{60} by OKE at 830 nm, all obtaining enhanced third-order optical nonlinearity compared with that of pristine fullerene. Recently, Wang et al. [391] have investigated the third-order optical nonlinearities of iron (III) and ruthenium (III) derivatives of organofullerene with ultrafast optical Kerr gate technique at 830 nm, indicating that the electron-deficient metal ion iron (III) blocked charge transfer from the organic group to C_{60} and reduced the optical nonlinearity, while a ruthenium (III) compound, because of the chlorobridge built by the remaining uncoordinate orbits of ruthenium ions, showed strongly enhanced optical nonlinearity. These measured results suggested that the increased charge on the cage,

and not asymmetric charge distribution due to metal-cage interactions, is the primary reason for increased nonlinearity, encouraging implications for the use of charge-transfer and endohedral or exohedral fullerene complexes in future photonic devices.

The serial work of Gong's group [303, 383, 388, 390] showed the following order of the nonresonant γ value for C_{60}- or C_{70}-derivatived molecules: $C_{60} < C_{60}(NH_2CN)_5 < C_{60}[(NH_2)_2CNCN]_5$ and $C_{70} < C_{70}(NH_2CN)_5 < C_{70}[(NH_2)_2CNCN]_5$. This trend indicates that the enhancement of the third-order optical nonlinearity is proportional to the increase of charge-transfer strength. This is a very important rule for us to design and synthesize new molecules for third-order optical nonlinear applications.

The fabrication of stable and efficient NLO molecules is an interesting challenge because of a large class of possible applications of these molecules. Controlled-transfer multilayered structures do offer an alternative solution, well argued by the controlled fabrication and expected enhancement stability. The work of Kajzar et al. [380]; Gong et al. [378]; and Koudoumas et al. [389] confirmed this point by measuring the nonlinear optical properties of C_{60}-based composites and multilayered charge-transfer structures with 5,10,20-tetraphenyl-21*H*,23*H*-porphine (TPP) and 5,6,11,12-tetraphenylnaphthacene (rubrene) (TPN), poly tetrahydrofurfuryl methacrylate (PTHFMA), polystyrene *n*-arm star polymers $(PS)_n$ (n = 3,6) with molecular weights of 900, 17800, and 4800 Da.

Konstantaki et al. [384] examined the third-order NLO response of the water-soluble inclusion complex C_{60}/γ-cyclodextrin by using z-scan technique at 532 nm. Because aging of this complex solution results in the formation of aggregates, enhanced γ values were obtained (see Table 8).

A main direction for searching organic materials that have enhanced third-order optical nonlinearities have focused on adding π-conjugated bonds, for example, —C=C—C=C—. This leads to decreased transparency because of narrowing of HOMO-LUMO gaps. Very recently, Fuks-Janczarek et al. [392] reported DFWM measurements on the third-order NLO responses of novel dyads of C_{60}-TTF with saturated (—C—C—C) chemical bonds at 532 nm in picosecond time. The main importance of their work lies in their obtained results that the increase of the third-order optical nonlinearity is not connected with the increasing number of conjugated bonds (which increase the UV-Vis-transparency) but is due to the increasing number of saturated (—C—C—C—) bonds without a decrease of the effective HOMO-LUMO gap. Those C_{60}-derivative compounds have large optical limiting coefficients, which imply interesting applications in optical limiting systems.

Organic materials have two advantages: being optimized at the molecular engineering level for a specific application (for example, photorefractivity [395], holographic memory [396], electro-optic modulation [397]) and combining chemical functionality with their optical properties. Apart from showing NLO properties with the inherent electron-transfer properties between donor and acceptor groups, one of these groups can have redox or proton-transfer capability [398]. This was exploited to demonstrate the reversible switching of the first-order hyperpolarizability of a donor-acceptor molecule [399]. Very recently, Asselberghs et al. [400] investigated the proton-accepting capability of the dimethylamino electron donor group on substituted 5,6-open azafulleroids [401], obtained through ring-opening aryl azide with C_{60} in toluene, to demonstrate reversible switching based on proton transfer. Their open azafulleroid exhibits a clear charge-transfer absorption band around 710 nm, and the first-order hyperpolarizability was determined by using nanosecond hyper-Rayleigh scattering at 1064 nm in chloroform to be 1.4×10^{-28} esu, which is slightly resonantly enhanced because of the closeness of the charge-transfer absorption band to the second-harmonic wavelength. The side product, 6,6-closed aziridinofullerene, lacked the homoconjugation with the electron-accepting fullerene moiety and led to a lower hyperpolarizability. Their detailed studies demonstrated that the second-order NLO response was dramatically reduced upon protonation and completely restored after deprotonation. From the point of view of developing molecular photonic devices whose properties can be switched by modifying one of the component parts, the ability to switch the NLO response of a molecule on and off reversibly by a simple controllable perturbation would add significant value to the utility of NLO molecules.

Because C_{60} is not easy to mix homogeneously with organic-inorganic material, less work on the optical nonlinearity of C_{60} doped in a solid matrix is reported. Very recently, Yu et al. [402] used sol–gel technique and synthesized the solid compounds of C_{60} attached by three $NH_2—(CH_2)_3—Si(OC_2H_5)_3$ groups. The nonresonant third-order optical susceptibilities measured by OKE gate technique at 820 nm are 1.4×10^{-13} esu and 1.6×10^{-13} esu for $C_{60}[NH_2—(CH_2)_3—Si(OC_2H_5)_3]_3$ sol and gel, respectively. This off-resonant NLO response is also mainly derived from the electron transfer process. Their results suggest a potential application of C_{60} derivatives of solid state in ultrafast optical devices.

4.3.2. Theoretical Calculations

Theoretical investigations of the third-order optical nonlinearities of doped carbon nanostructures, for example, fullerene derivatives, have so far been done only for the substitute doped fullerenes [242, 244, 403–409]. Rustagi, Ramaniah, and Nair [242] were the first to consider the second hyperpolarizability of C_{60} substitutional doped with one B or one N atom. They used a tight-binding mehod for calculating the second hyperpolarizability and found a large enhancement in the doped fullerenes, especially for $C_{59}N$. Later, Dong, Jiang, Xing, and coworkers used the Su-Schrieffer-Heeger (SSH) model for calculating the second hyperpolarizability of $C_{60-n}X_n$ ($n = 1, 2$) where X = B or N [403, 404] and also found enhanced nonlinearities compared with C_{60}, especially for the doubly substituted fullerenes. However, for the monosubstituted fullerenes, they showed that by including Coulomb interactions in the SSH model they found a much smaller second hyperpolarizabilty [405]. Xie found that using an SSH model with the Coulomb interactions included enhanced third-order nonlinearity of C_{70} monosubstituted with either B or N [406]. See the methods reviewed in later section.

Recently, the static second hyperpolarizability of the double substitute-doped fullerenes $C_{58}NN$, $C_{58}BB$, and $C_{58}BN$ has been investigated by Jensen et al. [244] using TD-DFT. They found only small changes in the second hyperpolarizability for $C_{58}NN$ and $C_{58}BB$, but an enhanced second hyperpolarizability was found for the donor-accepted doped fullerene $C_{58}BN$. Also, using TD-DFT, Xie and colleagues [407, 408] investigated the second hyperpolarizability of the heterofullerene $C_{48}N_{12}$ and $C_{48}B_{12}$ and found an enhanced second hyperpolarizability, especially for $C_{48}B_{12}$, for which the average second hyperpolarizability was 180% larger than that of C_{60}.

Cheng et al. [409] studied the frequency-dependent second-order hyperpolarizability of the $C_{59}Si$ and $C_{58}Si_2$ heterofullerenes using the INDO/CI-SOS method. Using the results for the second-order hyperpolarizability of $C_{59}Si$ and $C_{58}Si_2$, they estimated the third-order susceptibility for the DFWM optical process for Si-doped films to be about 10^{-10} esu at the first excited state and 10^{-11} esu at the ground state.

Recently, doped carbon nanostructures have stimulated great interest from researchers in physics and chemistry to investigate their structural, electronic, optical, and other properties. Besides the alkali metal doping, there is another type of doping, substitute doping (SD), which is substituting one or more carbon atoms of fullerenes and carbon nanotubes by other atoms. For example, boron and nitrogen atoms have been successfully used to replace carbon atoms of fullerenes and carbon nanotubes. Available studies have shown that the lattice and electronic structures of fullerenes change with substituted doping; the band gaps between the highest occupied molecular orbitals (HOMO) and lowest unoccupied molecular orbitals (LUMO) and the electronic polarization of the substituted fullerenes vary greatly with different SD; the distribution of π electrons on the surface of fullerene is changed due to the SD effect; the original delocalized π electrons in pure fullerene become more localized around the substituted atoms. Obviously, these factors also have a large effect on the NLO properties of carbon tubules as well as fullerenes. Therefore, it would be interesting and useful to investigate theoretically the SD effect on the NLO properties of carbon nanostructures from the viewpoint of practical application. In this section, we review the model we developed for studying the third-order optical nonlinearity of doped carbon nanostructures.

The ESSH model has been used to describe pure carbon nanostructures. But it should be modified to include the effect of the dopant ions in order to describe the substituted

carbon nanostructures. This model refers to doped ESSH (DESSH) model. In detail, the total Hamiltonian for the single substituted tubules can be written as [24]

$$H = H_{c-c}^{(0)} + H_{X-c}^{(1)} \tag{61}$$

$$H_{c-c}^{(0)} = \sum_{\langle ij\rangle}\sum_{s}(-t_0 - \alpha_0 y_{ij})(c_{i,s}^{\dagger}c_{j,s} + h.c.) + \frac{k_0}{2}\sum_{\langle ij\rangle} y_{ij}^2 + u_0\sum_{i} c_{i,\uparrow}^{\dagger}c_{i,\uparrow}c_{i,\downarrow}^{\dagger}c_{i,\downarrow} + v_0\sum_{\langle ij\rangle}\sum_{s,s'} c_{i,s}^{\dagger}c_{i,s}c_{j,s'}^{\dagger}c_{j,s'} \tag{62}$$

$$H_{X-c}^{(1)} = \sum_{\langle ij\rangle}\sum_{s}(-t_1 - \alpha_1 y_{ij})(c_{i,s}^{\dagger}c_{j,s} + h.c.) + \frac{k_1}{2}\sum_{\langle ij\rangle} y_{ij}^2 + u_1\sum_{i} c_{i,\uparrow}^{\dagger}c_{i,\uparrow}c_{i,\downarrow}^{\dagger}c_{i,\downarrow} + v_1\sum_{\langle ij\rangle}\sum_{s,s'} c_{i,s}^{\dagger}c_{i,s}c_{j,s'}^{\dagger}c_{j,s'} \tag{63}$$

where X denotes the substituted atom, and the sum $\langle ij\rangle$ is taken over the nearest neighbors for both the c—c and X—c bonds; t_0 (or t_1), α_0 (or α_1), and k_0 (or k_1) represent the hopping integrals, the electron–phonon coupling constants, and the spring constants corresponding to the c—c (or X—c) bonds, respectively; u_0 or u_1 is the usual on-site Coulomb repulsion strength, and v_0 or v_1 is the Coulomb interaction between the nearest and next-nearest atoms. Because there is only one substituted impurity atom (namely, X) in tubules, $H_{X-c}^{(1)}$ plays a perturbational role, and as an approximation, the original empirical parameters (t_0, α_0, k_0, u_0, v_0) in $H_{c-c}^{(0)}$ are assumed not to change because of the substitute doping (taken to be the same in this numerical calculation as those in the pure cases studied in the previous section), $u_1 \approx u_0$, and $v_1 \approx v_0$.

As before, we use the Hartree-Fock approximation to transform the previous equations into [24]

$$H = H_{c-c}^{(0)} + H_{X-c}^{(1)}; \tag{64}$$

$$H_{c-c}^{(0)} = \sum_{\langle ij\rangle}\sum_{s}(-t_0 - \alpha_0 y_{ij})(c_{i,s}^{\dagger}c_{j,s} + h.c.) + \frac{k_0}{2}\sum_{\langle ij\rangle} y_{ij}^2 + u_0\sum_{i}\left(\sum_{s}\rho_{i,s}c_{i,s}^{\dagger}c_{i,s} - \rho_{i,\uparrow}\rho_{i,\downarrow}\right) + v_0\sum_{\langle ij\rangle}\sum_{s}\left(\sum_{s'}\rho_{j,s'}c_{i,s}^{\dagger}c_{i,s} - \rho_{i,s}\sum_{s'}\rho_{j,s'} - \tau_{ij,s}c_{i,s}^{\dagger}c_{j,s} + \tau_{ij,s}^2\right); \tag{65}$$

$$H_{X-c}^{(1)} = \sum_{\langle ij\rangle}\sum_{s}(-t_1 - \alpha_1 y_{ij})(c_{i,s}^{\dagger}c_{j,s} + h.c.) + \frac{k_1}{2}\sum_{\langle ij\rangle} y_{ij}^2 + u_1\sum_{i}\left(\sum_{s}\rho_{i,s}c_{i,s}^{\dagger}c_{i,s} - \rho_{i,\uparrow}\rho_{i,\downarrow}\right) + v_1\sum_{\langle ij\rangle}\sum_{s}\left(\sum_{s'}\rho_{j,s'}c_{i,s}^{\dagger}c_{i,s} - \rho_{i,s}\sum_{s'}\rho_{j,s'} - \tau_{ij,s}c_{i,s}^{\dagger}c_{j,s} + \tau_{ij,s}^2\right) \tag{66}$$

where $\rho_{i,s} = \langle c_{i,s}^{\dagger}c_{i,s}\rangle$ is the electron density, and $\tau_{ij,s} = \langle c_{i,s}^{\dagger}c_{j,s}\rangle$ is the bond order parameter. This equation is also solved by the standard adiabatic approximation method, which leads to the Schrödinger equation for the π electron

$$\epsilon_k Z_{k,s} = \sum_{\langle ij\rangle}(-t_0 - \alpha_0 y_{ij}^{(0)} - v_0\tau_{ij,s})Z_{k,s}(j) + \left[u_0\rho_{i,s} + v_0\sum_{s'}\sum_{j}\rho_{j,s'}\right]Z_{k,s}(i) + \sum_{\langle ij\rangle}(-t_1 - \alpha_1 y_{ij}^{(1)} - v_1\tau_{ij,s})Z_{k,s}(j) + \left[u_1\rho_{i,s} + v_1\sum_{s'}\sum_{j}\rho_{j,s'}\right]Z_{k,s}(i) \tag{67}$$

where $y_{ij}^{(0)}$ and $y_{ij}^{(1)}$ denote the y_{ij} for c—c and X—c bonds, respectively; ϵ_k is the eigenvalue of the kth eigenstate; $Z_{k,s}$ is the electronic wave function. The total energy of the system is a functional of the set of y_{ij}:

$$E = \sum_{k}\epsilon_k + \sum_{\langle ij\rangle}\frac{1}{2}k_0(y_{ij}^{(0)})^2 + \sum_{\langle ij\rangle}\frac{1}{2}k_1(y_{ij}^{(1)})^2 \tag{68}$$

where the first sum runs over only the occupied states. Minimizing the total energy E over y_{ij} and using the constraint condition

$$\sum_{(ij)} y_{ij}^{(0)} + \sum_{(ij)} y_{ij}^{(1)} = 0 \tag{69}$$

we are able to obtain the self-consistent equation for y_{ij}

$$y_{ij}^{(0)} = -2\alpha_0 k_0^{-1} \sum_{k,s} \left\{ Z_{k,s}(i) Z_{k,s}(j) - \Pi^{-1} \sum_{(ml)} Z_{k,s}(m) Z_{k,s}(l) \right\} \tag{70}$$

$$y_{ij}^{(1)} = -2\alpha_1 k_1^{-1} \sum_{k,s} \left\{ Z_{k,s}(i) Z_{k,s}(j) - \Pi^{-1} \sum_{(ml)} Z_{k,s}(m) Z_{k,s}(l) \right\} \tag{71}$$

where the first sum also runs over only the occupied states, and Π is the number of π bonds. These coupled equations can be solved iteratively, and final result should be independent of choosing the different initial values of the set y_{ij}.

The choice of the three parameters (t_1, α_1, k_1) for the X—c bonds is, of course, important [24]. The best way to do it is to determine them by comparison between theoretical calculations and experimental measurements. But to the best of our knowledge, there has not been an experimental measurement on the nonlinear optical properties of doped carbon tubules. Recently, by using a molecular orbital method with Harris functional and spin-restricted approximations [410], where the total electron density of the system can be approximated by a superposition of electron densities of the isolated atoms with a first-order energy correction of the density error and quadratic errors in the electron Coulomb repulsion and exchange-correlation energies are partially canceled, Kurita et al. [411] optimized the structures of $C_{59}N$ and $C_{59}B$ and at the same time investigated their electronic properties. They found that the optimized structures and binding energies for $C_{59}N$ and $C_{59}B$ were almost the same as those for C_{60}, and the energy levels near the Fermi level were remarkably changed by doping. Using the DESSH model shown previously, Xie [24] also investigated the structural and electronic properties of the same substituted doped fullerenes by carefully adjusting the values of the three parameters (t_1, α_1, and k_1). It was found that our numerical calculations can accurately reproduce the results obtained by Kurita et al. [411] if $t_1 = 1.17$ eV, $\alpha_1 = 6.04$ eV/Å, and $k_1 = 51.1$ eV/Å^2 for $C_{59}B$, $t_1 = 1.05$ eV, $\alpha_1 = 6.13$ eV/Å, and $k_1 = 49.6$ eV/Å^2 for $C_{59}N$. The total energies of the molecules and the excess electron densities at and around the positions of substitute impurity atoms are also calculated. Both $C_{59}N$ and $C_{59}B$ have nearly the same total energy, which means that they are equally stable. The excess electron density of the boron atom is -0.5128 but that of the nitrogen atom is 0.2745, which implies that electron deficiency is produced at the doped boron atom site and that the boron atom gives up its electronic charge to its neighbors and exists as a donor. It is obvious that electronic charge accumulates on the doped nitrogen site and the nitrogen atom exists as an acceptor. The signs of excess electron density for the boron and nitrogen atoms are different, which makes $C_{59}N$ and $C_{59}B$ have opposite electronic polarization. The distributions $D(y_{ij})$ of the bond variables in $C_{59}N$ and $C_{59}B$ exhibit a two-peak structure, which indicates the presence of the dimerization. The narrow peaks in the negative y_{ij} region correspond to the distortion part around the impurity ions. The area of the extended portion for $C_{59}B$ is greater than that of $C_{59}N$, which means that the doped boron atom will produce stronger distortions around it than that in the nitrogen atom. Finally, the site-occupying probabilities of the HOMO states for $C_{59}B$ and $C_{59}N$ show that there are two large peaks at the position of the impurity atom. The peak in $C_{59}N$ is higher than that in $C_{59}B$. This reflects that the localization effect coming from the nitrogen impurity atom is stronger than that from the boron atom. These conclusions are also well consistent with the experimental observations [412, 413]. Therefore, in numerical calculations, we adopt the same values of all the parameters for $C_{59}B$ and $C_{59}N$. Of course, after more experimental observations are made for doped carbon nanotubes, it will be possible to determine more accurately all the parameter values in our DESSH model.

Based on this model, Xie [24] has proposed a substitutional doping approach to achieve the large third-order optical nonlinearities of carbon nanotubes, which significantly enhances the γ value of nanotubes by about one order of magnitude with respect to C_{60} (see Table 9). Because of the distortion of π electron distribution in the substituted tubes, especially around the dopant atoms, the difference between the z and x (or y) components of γ for doped carbon nanotubes is much more pronounced than that for the parent "pure carbon nanotubes." The study of the dynamical NLO responses of pure carbon nanotubes indicates that the relatively large NLO responses for carbon nanotubes are mainly caused by the delocalized π electronics, as in the conjugated polymer chains. The three-dimensional character of the nanotube leads to severe limitations on their nonlinear optical properties and makes their γ values smaller than those of linear polymers containing the same number of carbon atoms. However, as shown by Xie [24], the substituted dopants, such as B and N atoms, could attract or repel electrons and thus introduce a local perturbation on the π electron distribution around the dopants, leading to the so-called inductive effect. On the other hand, the dopant ions would result in a stronger localization of the original delocalized π electrons around them and therefore may reduce the effective space dimensions of nanotubes, namely, reduction effect. Both inductive and reduction effects would make the NLO properties of doped carbon nanotubes superior to pure carbon nanotube.

Finally, the localization effect of the N dopant is stronger than that of the B, implying the stronger enhancement in N-doped carbon nanotubes. Thus, it would be interesting to study the third-order optical nonlinearities of carbon nanotubes with heavily N-substitutional doping in future experiments.

5. PASSIVE OPTICAL LIMITING OF CARBON NANOSTRUCTURES

The most important application of optical limiters is the protection of our human eyes, sensitive optical sensors, and components from laser damage [414–419]. Also, there are many other potential applications for optical limiters, such as laser power regulation, stabilization, restoration of signal levels in optical data transmission, and logic systems [414–416, 418, 419]. As mentioned in Section 1, to achieve the optical limiter devices, passive systems have been considered an ideal approach, although these systems place a severe requirement on the nonlinear medium that should have a large optical nonlinearity. To build an ideal passive optical limiter, an optical sensor needs a high linear transmittance $T_{\max}^{\text{linear}}$ at a low input light fluence W_{ILF} (also named *limiting threshold* for the limiter) for the transmission of images, and the optical limiter should be able to clamp the transmitted energy below a maximum value W_{OLF} (or, say, a saturated value) of the output light fluence (OLF) for higher inputs, up to the maximum input energy W_{damage} at which the optical limiter undergoes irreversible laser damage and a minimum transmittance $T_{\min}$ of the limiter is reached [414]. Usually, the performance of a optical limiting system is desirable if a large linear transmittance $T_{\max}^{\text{linear}}$ is combined with a low minimum transmittance $T_{\min}$. In Fig. 7, we show the response of

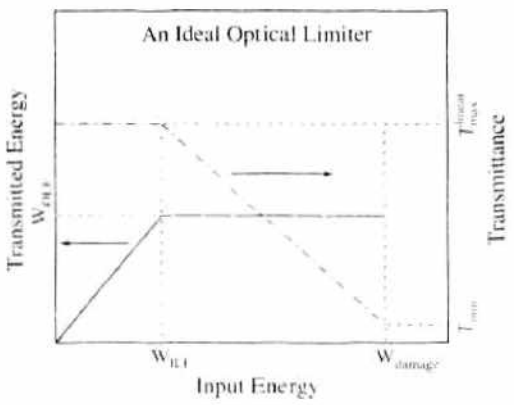


Figure 7. The relation between transmitted energy, transmittance, and input energy of an ideal optical limiter. See text for the definitions of quantities.

an ideal optical limiter. A good optical limiter should also be stable under ambient and photoirradiation conditions and efficient from a few picoseconds up to longer pulse duration from nanoseconds and millisecond on a broad range of wavelengths (visible and near infrared). Thus, the transmitted pulse energy never exceeds the damage threshold for the optical sensor. Next, we briefly review the basic principles of passive optical limiting.

5.1. Basic Principles of Passive Optical Limiting

Passive optical limiters rely on effective NLO processes [414, 415, 418, 419] to dissipate the incident light as a function of its intensity and to allow only a reduced transmission to the target area being protected. There are many schemes proposed for passive optical limiting. The common theme to all schemes is that the NLO materials are put in or near a focal plane in the optical device. Here we mention five main schemes that form the basis for the vast majority of practical limiting devices: (1) coherent NLO absorption (reverse saturable absorption (RSA), two-, three- or multi-photon absorption [420]), (2) nonlinear refraction (electronic effects [95] or thermal lensing [421]), (3) nonlinear scattering (solvent bubble formation and/or particle sublimation [422–424] or mismatched indices [425]), (4) photorefraction, and (5) optically induced phase transitions. Next, we briefy review how NLO property leads to limiting in each scheme.

5.1.1. Coherent NLO Absorption

In this scheme, the absorption increases with increasing incident pulse fluence or irradiance. In the two-photon absorption (TPA), the effective absorption coefficient α_{eff} can be written as follows [417]:

$$\alpha_{\text{eff}} = \alpha + \beta I + \sigma_{ex} = \sigma_g \rho_g + \beta I + \sigma_{ex} \rho_{ex} \tag{72}$$

where β is the TPA coefficient, σ_g and σ_{ex} are the ground and excited state absorption cross section, respectively, with the corresponding population density ρ_g and ρ_{ex}, and I is the strength of the incident irradiance. If $\sigma_{ex} > \sigma_g$, the absorption will increase with increasing ρ_{ex} and hence with increasing incident fluence. This is usually called *excited state absorption* (ESA) or *reverse-saturable absorption* (RSA). For $\sigma_{ex} < \sigma_g$, the absorption decreases with increasing fluence and referred to *saturable absorption*. An optical limiter is required to work under high levels of excitation with a large ratio of σ_{Ex}/σ_g but small linear absorption. However, σ_g cannot be too small because ρ_{ex} must become large enough to produce a strong limiting effect and a maximum $T_{\min}/T_{\max}^{\text{linear}} = \exp[-(\sigma_{ex} - \sigma_g)\rho L]$ is achieved when all molecules are promoted to the first excited state, where ρ is the total molecular density and L is the material thickness [416]. Among the NLO absorption, RSA is a primary mechanism for the nonlinear absorption optical limiting and is demonstrated to be one of the best processes to use for optical limiting because it reduces the total pulse energy rather than simply reducing the fluence or irradiance.

It is generally accepted that the RSA of a material system, for example, C_{60}, can be described by a five–level model [426–431]. If the absorption cross section for the singlet ground state is larger than that for the singlet excited state, the NLO material is more transparent and becomes a saturable absorber. However, when the absorption cross sections of the singlet or the triplet excited states are larger than that of the ground state and the lifetime of the excited state is long enough, the total absorption of the NLO material increases and the material is then the RSA process mentioned previously. For longer laser pulses, significant intersystem crossing (ISC) to a triplet state occurs for some kinds of materials (for example, an important feature of C_{60} is the fast and efficient ISC from the first singlet excited state to the first triplet excited state). In these cases, the excited state absorption appears from the lowest triplet to higher excited triplet states providing an increasing contribution to the total absorption of the system. It should be pointed out that the transition from the ground state to the first singlet electron state is forbidden because of the same parity but becomes partially allowed because of vibronic interactions responsible for the weak absorption band in the visible region extending from 450 nm to 650 nm. In this RSA mechanism, significant parameters are the excited state lifetimes and the ISC

time, compared with the duration of the laser pulse, plus the values of the absorption cross sections. Typical reverse saturable absorbers are molecules with weak ground state absorptions at the concerned wavelengths, such as metallophthalocyanines [418, 432, 433], mixed metal complexes [434–437], and fullerenes [419]. In a more general RSA mechanism, the strongly absorbing electronic excited states can be populated by other incoherent processes, for example, internal conversion and energy transfer.

5.1.2. Nonlinear Refraction

Kramers-Krönig relations show that all materials showing nonlinear absorption must accompany nonlinear refraction [438]

$$n_{\text{eff}} = n_0 + n_2 I + n_{ex}\rho_{ex} \tag{73}$$

where n_0 is the normal refractive index of material, n_2 is the instantaneous index change proportional to the incident irradiance, and n_{ex} is the index change because of the population of excited states; n_2 is related to the TPA coefficient β [438]; n_{ex} and σ_{ex} are related in a similar manner [417]. As mentioned before, nonlinear absorption devices need a large amount of energy, and thus thermal damage problems may occur. However, this problem does not exist on the nonlinear refraction mechanism for optical limiting. A potential problem is that inadvertent refocusing of the eye could reduce the defocusing effect of the optical limiter [417].

5.1.3. Nonlinear Scattering

Scattering can strongly decrease a transmitted light beam. As we switch on a laser on a material, a new scatter center may be created, and the refractive index difference between existing scatter centers and their surroundings may change correspondingly. Thus, nonlinear scattering is an important and unintentional mechanism in the operation of optical limiters.

5.1.4. Photorefraction

As a material is exposed to a light, its refractive index would change because of a complex process involving photoexcitation of charge carriers and diffusion of those carriers that leads to a space charge field [417], which in turn causes an index change via the electro-optic effect. In this mechanism, the limiting is coherence- and intensity-dependent. The requirement for charge diffusion leads to slow turn-on time. Thus, this type of optical limiter is suitable only for pulses of millisecond or longer duration [417].

5.1.5. Optically Induced Phase Change

This limiting scheme is characterized by showing a reversible, thermally induced semiconductor-metal phase transition as a material is illuminated by using a strong laser radiation [417]. To date, a lot of materials have been found to exhibit this kind of phase transition, for example, vanadium oxides, being transparent to the infrared radiation in their semiconducting state but highly reflective in their metallic state. Hence, in the infrared region, the optical limiters may be transparent for low powers, whereas weak optical absorption and subsequent heating may induce the strongly reflecting metallic phase at high powers blocking the transmitted light [417]. To get an effective limiter, the material should be stable in its transparent state, have a small latent heat related with the phase transition, and need a small temperature (about 100 K) to induce the phase transition [417].

In the following, we review extensive studies of the optical limiting properties of carbon nanostructures. Most of these results for these nanostructures are attributed to the RSA processes, whereas nonlinear scattering plays an important role in the optical limiting performances of carbon nanotubes. Some important results for those structures and others are summarized in Table 13.

Table 13. Optical limiting performance of fullerenes, fullerene derivatives, single-walled carbon nanotubes (SWNT), multiwalled carbon nanotubes (MWNTs), carbon black suspension (CBS), and some other materials. W_{ILT} denotes the ILT threshold; W_{OLT} is OLT saturation threshold; τ and λ are the duration and wavelength of the laser light; T_{min}^{linear} is the percentage of linear transmitting sample; The TCNEO-C_{60}/PC ratio for PCC1, PCC2, PCC3, PCCC4, PCC5 and PCC6 are 1.02, 1.06, 1.01, 1.96, 3.05, 5.03 wt%; K_m (in 10^{-3} M) is the solution concentration; DMF = dimethylformamide; TBM(60)CO = *tert*-buty methano-C_{60} carboxylate. C_{70}-H = η^2-C_{70})Mo(CO)$_2$(o-phen)(DBM); PMMA = poly(methyl methaerylate); PPEI = poly(propionylethyleneimine); PPEI-EI = poly(propionylethyleneimine-*co*-ethylenimine). PVDF = poly(vinylidenefluoride); OC = octadecylamine (adapted from Xie et al. [25]).

Material (K_m)	Solvent	τ (psec)	λ (μm)	T_{min}^{linear}	W_{ILT} (kJ/m^2)	W_{OLT} (kJ/m^2)	Ref.
C_{60}	Toluene	8 ns	0.532	63	1.1	0.65	[439]
	Toluene	8 ns	0.532	70	1.6	0.71	[439]
	Toluene	8 ns	0.532	80	5.1	2.4	[439]
	Toluene	5 ns	0.532	55	1.2	0.58	[443]
	Toluene	5 ns	0.532	70	1.8	1.1	[443]
	Toluene	5 ns	0.532	70	1.8	1	[456]
	Toluene	40 ps	0.532	76	3.2	1.2	[442]
	Toluene	15 ns	0.532	55	3	7.8	[466]
	Toluene	8–10 ns	0.532	66	6	1–1.2	[441]
	Toluene	8 ns	0.532	81	1.5	2.8	[489]
	CCl_4	8–10 ns	0.532	80	10	4.6	[441]
	CCl_4	8–10 ns	0.532	65	3	2	[441]
	CCl_4	8–10 ns	0.532	50	1	0.5	[441]
	Toluene	8–10 ns	0.308	25	3	3	[441]
	Toluene	8 ns	0.532	65	1	0.899	[481]
	DMF	10 ns	0.532	75	1	1.2	[490]
C_{60} (0.33)	Toluene	5 ns	0.532	70	0.1	2.3	[443]
C_{60} (1.64)	Toluene	5 ns	0.532	70	0.1	1.2	[443]
C_{60} (8.19)	Toluene	5 ns	0.532	70	0.1	1	[443]
C_{60} (0.5)	Toluene	45 ps	0.532	—	1.5	1	[478]
C_{60} (2.5)	Toluene	45 ps	0.532	—	0.5	0.5	[478]
C_{70}	Toluene	8 ns	0.532	70	1	3.5	[439]
	Toluene	10 ns	0.532	70	2	2.7	[440]
C_{60}/C_{70} mixture	Toluene	10 ns	0.532	14	4.1	0.4	[444]
	Toluene	10 ns	0.532	24	4.5	0.7	[444]
	Toluene	10 ns	0.532	44	6.2	3.2	[444]
	Toluene	10 ns	0.532	69	12.2	6.9	[444]
	Argon-degased	10 ns	0.532	44	1.1	1.2	[444]
	Not argon-degased	10 ns	0.532	44	4.8	3.1	[444]
C_{76}/C_{84} mixture	Toluene	10 ns	1.064	45	7	20	[441]
C_{70}-H	1,2-Dichlorobenzene	10 ns	0.532	30	1.8	0.25	[440]
	1,2-Dichlorobenzene	10 ns	0.532	50	0.94	1.1	[440]
	1,2-Dichlorobenzene	10 ns	0.532	70	0.37	1.6	[440]
CuPc-C_{60} (0.5)	Toluene	45 ps	0.532	—	1.5	0.7	[478]
CuPc-C_{60} (2.5)	Toluene	45 ps	0.532	—	0.1	0.2	[478]
DTC_{60}	Hexane	8 ns	0.532	65	1	1.099	[481]
DTC_{60}-Ag	Hexane	8 ns	0.532	65	1	0.716	[481]
$HDTC_{60}$	Hexane	8 ns	0.532	70	3	4.4	[485]
$HDTC_{60}$-Ag	Hexane	8 ns	0.532	70	0.8	2.2	[485]
C_{60}BPY	Chloroform	8 ns	0.532	51	2	1.3	[489]
C_{60}TPY	Chloroform	8 ns	0.532	51	2	1.3	[489]
C_{60}BPY-Au	Chloroform	8 ns	0.532	81	2	6.1	[489]
C_{60}TPY-Au	Chloroform	8 ns	0.532	81	2	2.1	[489]
C_{60}BPY-Au	Chloroform	40 ps	0.532	75	0.5	1.2	[489]
C_{60}TPY-Au	Chloroform	40 ps	0.532	75	0.5	0.8	[489]
$C_{60}(CN)_2$	DMF	10 ns	0.532	55	1	1.2	[490]
	DMF	40 ps	0.532	55	0.7	0.5	[490]
	DMF	23 ps	0.532	55	0.7	0.5	[490]
C_{60}-PC	Chloroform	8 ns	0.532	33	3	3	[491]
PCC1	Chloroform	8 ns	0.532	70	3	3.5	[491]
PCC2	Chloroform	8 ns	0.532	70	3	3	[491]

continued

Table 13. Continued.

Material (K_m)	Solvent	τ (psec)	λ (μm)	T_{max}^{linear}	W_{ELI} (kJ/m^2)	W_{OLE} (kJ/m^2)	Ref.
PCC3	Chloroform	8 ns	0.532	70	3	4.2	[491]
PCC4	Chloroform	8 ns	0.532	70	3	2.7	[491]
PCC5	Chloroform	8 ns	0.532	70	3	4	[491]
PCC6	Chloroform	8 ns	0.532	70	3	3.8	[491]
TBM[60]CO (0.123)	Toluene	5 ns	0.532	70	0.9	3.1	[443]
TBM[60]CO (0.246)	Toluene	5 ns	0.532	70	0.9	3	[443]
TBM[60]CO (0.616)	Toluene	5 ns	0.532	70	0.9	1.8	[443]
TBM[60]CO (1.23)	Toluene	5 ns	0.532	70	0.9	1.7	[443]
TBM[60]CO (5.92)	Toluene	5 ns	0.532	71	0.9	1.4	[443]
TBM[60]CO (6.16)	Toluene	5 ns	0.532	70	0.9	1.1	[443]
TBM[60]CO (6.4)	Toluene	5 ns	0.532	69	0.9	1	[443]
TBM[60]CO (10.4)	Toluene	5 ns	0.532	55	0.9	0.7	[443]
TBM[60]CO (12.3)	Toluene	5 ns	0.532	70	0.9	1	[443]
TBM[60]CO (5.92)	tol-PMMA	5 ns	0.532	71	0.9	1.7	[443]
TBM[60]CO (10.4)	tol-PMMA	5 ns	0.532	55	0.9	0.9	[443]
TBM[60]CO (0.123)	$CHCl_3$-PPEI	5 ns	0.532	69	0.9	1.6	[443]
SWNT suspension	Water	5 ns	0.532	70	1	1.1	[456]
	Water	15 ns	0.532	55	12	11	[466]
	Water	15 ns	0.532	42	11	7.6	[466]
	Ethanol	15 ns	0.532	42	4	4.5	[466]
	Ethylene glycol	15 ns	0.532	42	11	15	[466]
Short-SWNT suspension	Water	5 ns	0.532	70	1.8	1.6	[456]
Short-SWNT-PPEI-EI	Chloroform	5 ns	0.532	70	2.1	3.5	[456]
Short-MWNT suspension	Water	5 ns	0.532	70	1.6	1.4	[456]
Short-MWNT-PPEI-EI	Chloroform	5 ns	0.532	70	1.6	2.6	[456]
Short-MWNT-OC$^+$	Chloroform	5 ns	0.532	70	1.6	2.9	[456]
MWNT (large aspect ratio)	PVDF/DMF	6–8 ns	0.532	50	2	3.5	[457]
MWNT (small aspect ratio)	PVDF/DMF	6–8 ns	0.532	50	2	4.1	[457]
CBS	PVDF/DMF	6–8 ns	0.532	50	2	3.3	[457]
	Water	5 ns	0.532	70	1	1.2	[456]
DT-Ag	Hexane	8 ns	0.532	65	1	2.8	[481]
DT-Ag	Hexane	8 ns	0.532	70	3	3.3	[485]
CuPc (0.5)	Toluene	45 ps	0.532	—	1	0.75	[478]
CuPc (2.5)	Toluene	45 ps	0.532	—	0.6	0.22	[478]
$HFeCo_3(CO)_{10}[P(CH_3)_3]_2$	Methylene chloride	8 ns	0.532	70	4	3.9	[439]
$HFeCo_3(CO)_{12}$	Methylene chloride	8 ns	0.532	70	4	5.5	[439]
$HFeCo_3(CO_{10})(PPh_3)_2$	Methylene chloride	8 ns	0.532	70	8	7.1	[439]
indanthrone	Dilute KOH	8 ns	0.532	70	1.8	1.6	[439]
chloroaluminium phthalocyanine	Methanol	8 ns	0.532	70	1.8	1.1	[439]
$[N(C_2H_5)_4]^+(FeCo_3(CO)_{12})^-$	Methylene chloride	8 ns	0.532	70	3.8	5.5	[439]
$WS_4Cu_4I_2(py)_6$	DMF	40 ps	0.532	64	1.5	0.59	[442]

5.2. Spherically Shaped Carbon Nanostructures

Tutt and Kost [439] were the first to observe the optical limiting performance in C_{60} solution. As listed in Table 13, C_{60} exhibits optical limiting behavior with a saturation threshold lower than those reported for C_{70} and other materials currently in use. RSA is the dominant mechanism: there exist allowed broad-band transitions from the first excited singlet and triplet to the higher excited states, and this results in absorption in the visible and near-infrared range that is much stronger than the absorption from the ground state, leading to RSA. Because the higher ground-state absorption cross section for C_{70} leads to a smaller ratio of excited-state to ground-state cross sections, a higher threshold for optical limiting in C_{70} is observed [439, 440]. In comparison with the C_{60}-toluene system, a reduction of the saturated threshold by a factor of 3 to 5 is obtained by Belousov and colleagues [441] in a C_{60}-CCl_4 system (see Table 13). Results similar to those of the C_{60}-CCl_4 system are obtained in a C_{60}-decalin solution [441]. Solutions of C_{60} in methylene chloride give results similar to those of the C_{60}-toluene system [439]. Very recently, Song et al. [442] have shown that the picosecond optical limiting threshold of C_{60} was higher than that of the nanosecond one.

Further experiments [443] show that the optical limiting performance of C_{60} in a room-temperature solution toward nanosecond laser pulses at 532 nm strongly depends on the fullerene solution concentrations (see Table 13), which give a significant optical limiting contribution. The optical limiting performance of the more concentrated C_{60} solution is generally stronger than that of dilute ones. For example, the C_{60} solution in toluene of 8.2×10^{-4} M concentration displays strong optical limiting, whereas in toluene of 1.6×10^{-4} M gives a higher saturated value. In summary, RSA for excitation of C_{60} solution with ps duration laser pulses is due to absorption from the first singlet excited state because the population in the first triplet excited state is negligible, whereas for excitation with laser pulses having a duration of a few nanoseconds, the dominant contribution to RSA comes from absorption from the first triplet excited state. RSA is stronger for nanosecond laser pulses than for ps laser pulses for wavelengths between 620 nm ands 810 nm.

Sun et al. [444] found that the saturated threshold values of the mixture C_{60}/C_{70} solution (see Table 13) are higher than that of C_{60} solution but lower than that of C_{70} solution, which is actually caused by the lower triplet state quantum yield and stronger absorption in the visible wavelength for C_{70} than C_{60}. Because the triplet state of C_{60} [445] with lifetime 330 ± 25 ns and C_{70} [446] with lifetime 730 ± 50 ns can be quenched by $^3O_2$, the existence of $^3O_2$ in the solution will influence the RSA process and the optical limiting behavior. Sun and colleagues [444] found that no argon-degased solution of the mixture C_{60}/C_{70} has a higher ILF threshold and a higher saturated threshold than the argon-degased solution (see Table 13).

Optical limiting performance was also observed in a mixture of higher fullerenes C_{76} and C_{84} at $\lambda = 1064$ nm for $\tau = 10$ ns [441]. Their saturated threshold value is, however, higher than that of C_{60}-toluene in the visible region. Again, this is determined by the low quantum yield into the triplet state for the higher fullerenes.

All nonradiative relaxation processes lead to heating of the solvent. Justus and colleagues [425, 447] pointed out that RSA increases absorption and enhances thermal defocusing, which shows additional contribution to the observed optical limiting performances in C_{60} solution. Negative thermal lensing in C_{60} is clearly reflected in the z-scan studies of Mishra et al. [448]. The thermal origin of negative lensing was also established on the optical limiting for different solvents [448]. C_{60} in carbon disulfide, which has the largest thermal figure of merit, exhibits stronger optical limiting behavior [448]. A different kind of solvent dependence was also observed by Sun and Riggs [419]: the optical limiting response of C_{60} solutions in electron-donating solvents (for example, diethylamine and dimethylamine) was much weaker than that in toluene solution. This can be understood: (1) upon photoexcitation of C_{60}, the dominant photogenerated transient species are C_{60} anion and diethylamine or dimethylamine cation; (2) because absorption from the C_{60} anion is weaker than that from the triplet C_{60}, the optical limiting performance is degraded. It should be mentioned, however, that thermal refraction contributes to the optical limiting behavior of ns laser pulses but not of ps laser pulses [449].

The work of Mishra et al. [450, 451] showed the significant contribution of nonlinear scattering to the observed optical limiting performance in C_{60} solution. In their limiter geometry, the transmission could decrease at higher fluences only due to RSA or abberations which come from the nonlinear lens. Their observed reduction in the transmission at higher fluences was larger than that expected from a theoretical analysis including RSA and nonlinear refraction, where the difference is from the nonlinear scattering.

In addition, the work of Riggs and Sun [419, 443] showed us that RSA in C_{60} also has some contributions from singlet and triplet excimer-like states formed due to bimolecular processes involving C_{60} molecules in the triplet state.

5.3. Carbon Blacks and Rod-Shaped Carbon Nanostructures

The optical limiting properties of carbon blacks [452–458], MWNTs [456], and SWNTs [458–466] carbon nanotubes have been recently studied. Some of the results are listed in Table 13.

5.3.1. Carbon Blacks

Optical limiting of carbon black suspensions (CBS) and carbon black deposited on glass was studied [452, 456, 457]. The carbon blacks were shown to exhibit very strong broadband (for example, 0.532, 1.064 μm laser wavelength) optical limiting properties for nanosecond pulses. By observing scattered light intensity as a function of input irradiance, it was clearly shown that the incident light becomes very strongly scattered as the incident energy increases. Measurements of the angular distribution of scattered light show a Mie scattering pattern typical of scattering particles larger than the original carbon black particles. It was concluded that the dominant nonlinearity that leads to the optical limiting observed in carbon black materials was the thermally induced scattering and absorption by microplasmas formed after thermionic emission from the laser-heated carbon particles. Details about the optical limiting of carbon blacks can be also found in the review article by Tutt and Boggess [414].

5.3.2. Rod-Shaped Carbon Nanostructures

Compared with C_{60} and carbon black suspension (CBS), carbon nanotubes are also good candidates for optical limiting applications. Sun et al. [453] and Chen et al. [454] were the first to measure the energy-dependent transmission of MWNTs suspended in distilled water [453] and ethanol [454], or MWNTs embedded in PMMA films [453] in the visible and infrared spectral regions (example, 532 nm, 700 nm, and 1064 nm) with 7 ns laser pulses. For MWNTs in ethanol, the limiting threshold η, defined as the value of the input fluence at which the transmittance falls to half of the linear transmittance, is about 1.0 J/cm^2 at 532 nm (lower than those of C_{60} and CBS) and 6 J/cm^2 at 1064 nm (lower than that of CBS; optical limiting phenomena totally vanish for C_{60} toluene solution). For MWNTs in water, $\eta \approx 1.0, 5.5, 13$ J/cm^2 at 532 nm, 700 nm, and 1064 nm, respectively. For MWNT in PMMA, $\eta \approx 3.1, 4.1, 8.0$ J/cm^2 at 532 nm, 700 nm, and 1064 nm, respectively. For MWNT and carbon particles in PMMA, the η values at 532 nm, 700 nm, and 1064 nm are almost the same as those for MWNT in PMMA. Clearly, MWNTs, unlike C_{60}, which has no limiting response at 1064 nm, are a broadband limiter up to 1064 nm. Because carbon nanotubes with a lower work function, lower electron binding energy, and stronger plasma excitation have no ground-state absorption at 532 nm and 1064 nm, the broadband limiting response should result from another mechanism. Recently, the energy-dependent transmission of MWNT suspension measured with picosecond and nanosecond laser pulses has shown a strong optical limiting action on the nanosecond time scales but does not display any limiting behavior in the picosecond regime even with an input fluence as high as 3 J/cm^2 [455], which is different from that of C_{60} but similar to that of CBS. The picosecond time-resolved pump-probe experiment of Sun et al. [455] indicated that it took about 0.5 ns for the nonlinear transmission to appear for both MWNT and CBS. Especially, the smaller the external diameters of MWNT, the faster and larger are the changes of the transmission. The nonlinear scattering experiments of Sun et al. [455] with nanosecond laser pulses at 532 nm confirmed that nonlinear scattering exists in MWNT suspension because the observed limiting behavior of MWNT is similar to that in CBS. Thus, the optical limiting action in the MWNT suspension comes from a mechanism similar to that in CBS.

Vivien and colleagues [458–463, 465] were the first to report the NLO transmittance of SWNT in a water suspension [458–460, 463, 465] or in chloroform [458, 460–463, 465] for pulse durations ranging from 3 to 100 ns and for wavelengths from 430 nm to 1064 nm. Their results are summarized in Table 14. For SWNT in water [459, 460, 463, 465], $\eta = 3.8 \times 10^{-6}$ J/cm^2 (a bit higher than that of CBS) for 7 ns Nd:YAG laser pulse at 532 nm, and 2.6×10^{-5} J/cm^2 (lower than that of CBS) for 7 ns Nd:YAG laser pulse at 1064 nm. Obviously, the detailed studies of Vivien and colleagues [459–465] demonstrated that SWNTs are also efficient limiters over a broad wavelength range. Also, their results clearly show the wavelength effect [463] (attributed to the larger absorption cross sections of the suspensions at shorter wavelengths and to the larger scattering cross sections of the laser-induced scattering centers [460]): the shorter the wavelength, the lower the limiting threshold η and the better the limiting efficiency. Their experimental measurements show the pulse-duration

Table 14. Optical limiting threshold η of SWNT suspended in chloroform or in water using different laser source with different wavelength λ and pulse duration τ (adapted from Vivien et al. [458–463, 465] and Xie et al. [25]).

Solvent	Laser Source	τ (ns)	λ (μm)	η (kJ/m^2)
Water	Nd:YAG	7	0.532	3.8×10^{-3}
	Nd:YAG	7	1.064	2.6×10^{-1}
	Ti:sapphire	25	0.7	10
	Ti:sapphire	15	0.8	3
	Ti:sapphire	30	0.9	11
	OPO	3	0.46	9
	OPO	3	0.58	30
	OPO	3	0.68	70
Chloroform	Nd:YAG	5	1.064	2
	Nd:YAG	5	0.532	2
	Nd:YAG	2	0.532	11
	Ti:sapphire	100	0.67	3
	Ti:sapphire	25	0.7	1.8
	Ti:sapphire	12	0.73	10
	Ti:sapphire	15	0.8	1
	Ti:sapphire	30	0.9	3
	Ti:sapphire	80	1	0.2
	OPO	3	0.46	10
	OPO	3	0.58	30
	OPO	3	0.68	60
	OPO	3	0.694	60

effects reflecting two different mechanisms (vapor-bubble growth and sublimation of carbon nanotubes) that contributed to the optical limiting performance: for example, only the sublimation effect at 1064 nm for 5 ns Nd:YAG laser pulse; both mechanisms at 1000 nm for 80 ns Ti:sapphire laser pulse. The limiting performances observed with 5 ns Nd:YAG laser pulse at 532 nm are considerably better than those obtained with 2 ns Nd:YAG laser pulse at 532 nm. Such strong pulse-duration dependence of the limiting performance can be explained by the fact that the maximum size of the scattering centers in the nanosecond regime is reached only after the end of the incident pulses [458]. That is also why optical limiting is not observed at all in the picosecond range but appears in the nanosecond and longer pulses. Moreover, the solvent effects [462, 463] demonstrated that MWNT in chloroform exhibit better limiting performance than MWNT in water because of two main reasons [463]: (1) because the absorption cross sections of the particles are slightly larger in chloroform than in water, chloroform exhibits lower thermal conductivity and diffusivity; (2) with longer pulses from Ti-sapphire laser, chloroform has a very low heat of vaporization, surface tension, and viscosity, which lead to faster solvent-bubble growth. Nonlinear scattering and nonlinear refraction are the dominant mechanisms [461].

Meanwhile, Mishra et al. [466] carried out a detailed study on the optical limiting behavior of SWNT suspensions in different host liquids, for example, ethanol, water, and ethylene glycol. Their results, listed in Table 13, demonstrated that optical limiting behavior in SWNT suspension appears mainly because of absorption-induced scattering in the suspension [466]. Especially, the extent of nonlinear scattering for good limiting performance in carbon nanotubes strongly depends on the host liquid. For example, optical limiting in the ethanol suspension was the strongest among the three suspensions: water, ethylene glycol, and ethanol (see Table 13).

Recently, Riggs et al. [456] and Jin et al. [457] have investigated the size-dependent optical limiting behaviors of SWNTs and MWNTs. Their results are listed in Table 13. These experimental measurements show that carbon nanotubes of large aspect ratio (or, say, long nanotube) possess stronger optical limiting properties. Apparently the carbon nanotubes exhibit a significantly weaker optical limiting response in homogeneous solutions than in suspensions, reflecting that the optical limiting properties of the suspended versus solubilized carbon nanotubes represent dominating contributions by a nonlinear scattering mechanism versus a nonlinear absorption mechanism [456].

The preliminary experimental results of Riggs et al. [456] pointed out that the optical limiting results of carbon nanotube suspensions of the same linear transmittance, similar to those of CBS, are independent of nanotube concentrations, whereas the solubilized carbon nanotubes, similar to fullerenes in solution discussed previously, show significant solution-concentration-dependent optical limiting behaviors.

5.3.3. Carbon Nanostructure Composite Material

The use of carbon nanostructures in a practical application of optical limiting would require some form of matrix to disperse the nanostructures and allow the fabrication of films, coatings, or suspensions of the matrix and carbon nanostructure composite material. On the other hand, the previously reviewed carbon nanotubes were suspended in liquids because of their poor solubility in most solvents. Those suspensions are unstable at high concentrations. So, it is necessary to get stable solutions of carbon nanotubes. Recently, O'Flaherty et al. [467] and Jin and colleagues [468] studied the optical limiting performance of several polymer-coated and polymer-grafted MWNTs (which form stable solutions in DMF) using 532 nm 7 ns Nd:YAG laser pulses. The poly(ethylene oxide)/MWNT, poly(4-vinylpyridine)/MWNT, poly(2-vinylpyridine)/MWNT, and poly(4-vinyl-phenol)/MWNT all exhibited optical limiting behaviors similar to that of MWNT suspension in DMF: the η values for all samples are about 1 J/cm^2. Considering that there is no optical limiting effect in all polymer-DMF solutions, the observed optical limiting behaviors of polymer/MWNT composites are obviously due to the components of MWNTs. Hence, the polymer does not change the NLO properties of MWNTs, and a large variety of polymers can be used as the matrix of carbon nanotubes for the limiting applications. Because carbon nanotubes are broadband optical limiters, the polymer/MWNT composites are also broadband limiters. The important thing is that their polymer/MWNT solutions are very stable toward air and laser radiation [468].

5.3.4. Doped Carbon Nanotubes

An instructive approach—that is, tailoring the local electronic properties of carbon nanotubes—has been proposed as a good approach to understand mechanisms involved in the limiting behavior of suspensions. As discussed in Ref. [469], the substituting doping of carbon nanotubes is responsible for stimulating a number of structural and electronic properties and thus the third-order optical nonlinearities of carbon nanotubes. Recently, Xu and colleagues [470, 471] measured the optical limiting properties of B- and/or N-doped carbon nanotubes. By varying the incident energy and measuring the transmitted energy, they observed enhanced optical limiting behaviors of B- or N-doped carbon nanotubes. In comparison with the nonlinear transmittance versus incident fluence of pure and B- or N-doped carbon nanotubes at 532 nm and 1064 nm, doped carbon nanotubes have better optical limiting properties (lower threshold values) than pure carbon nanotubes. Their results on the pure and B-doped carbon nanotube under identical input fluence (0.5 J/cm^2) indicate that the transmittance drops by about 60% and 33% for B-doped and pure carbon nanotubes, respectively. The optical nonlinearity within the B-doped sample is stronger than that in the undoped one. Although Fe catalyst particles were also found in N-doped carbon nanotubes, they were shown to make no contribution to optical limiting behavior [470, 471]. A study by Jin et al. [472] found that the optical limiting behavior of carbon nanotubes in PVDF/DMF solution is size dependent. The nanotubes of large aspect ratio possess stronger limiting properties. However, the limiting is obtained by nanotube bundles, not by individual nanotubes. Using electron microscopy, Xu and colleagues [470, 471] found that for each type of doped carbon nanotube, the nanotube bundles exceed 100 μm, which is significantly longer than the wavelength of incident light. In spite of these studies, a full mechanism for the enhanced optical limiting performance of B- or N-doped carbon nanotubes is still unclear.

5.4. Functionalized Carbon Nanostructures

A category of metal-fullerene complexes [473–477] was synthesized in the laboratory. Their charge-transfer nature qualifies these complexes as a promising group of excited-state absorption optical limiting materials, such as enhanced optical limiting performance of

a molybdenum complex of fullerene (η^2-C_{60})Mo(CO)$_2$(o-phen)(DBM) (DBM = dibutyl maleate; o-phen = 1,10-phenanthroline) [476]. Recently, Liu et al. [440] also observed enhanced optical limiting performance of a novel molybdenum complex of fullerene (η^2-C_{70})Mo(CO)$_2$(o-phen)(DBM) relative to fullerene C_{70} with nanosecond laser pulses. This is attributed to the increment on the triplet-state absorption because of intramolecular charge transfer.

Copper (II) phthalocyanines (CuPc) and C_{60} have been used to construct practical optical limiters because of their good optical limiting properties. As far as the solubilities and aggregations of these optical limiting molecules are concerned, combining phthalocyanine and fullerene in one compound through the Diels-Alder reaction has many advantages, for example, combining the optical limiting properties of both molecules, improving the solubility and stability, and avoiding the phase separation problems exciting in the mixture of various materials. Recently, Zhu et al. [478] demonstrated that a higher concentration solution CuPc-C_{60} shows similar optical limiting behavior to CuPc, while in a lower concentration solution the CuPc-C_{60} shows similar optical limiting behavior to C_{60} at low input fluence (<0.2 J/cm^2) and to that of CuPc at higher input fluence (>0.2 J/cm^2). In this kind of combined molecule, there are two separated conjugation systems, both of which independently contribute to the RSA.

The fullerene-metal nanocomposites remain a relatively unexplored area [479, 480]. Recently, Sun et al. [481] have prepared a novel [60] fullerene-Ag nanocomposite (DTC_{60}-Ag) by the *in-situ* reduction of silver ions encapsulated in a new monofunctionalized methano-[60]fullerene derivative (DTC_{60}) with reverse micellelike structure. As listed in Table 13, their experimental measurements demonstrated that the optical limiting behavior of DTC_{60}-Ag is better than that of both C_{60} and DTC_{60}, even better than the results obtained with novel materials such as silver-dendrimer nanocomposite [482], the AF-380 dye [483], and single-walled carbon nanotube suspensions [484]. Similar results [485] have recently been obtained for a new C_{60} hexamalonic derivative nanocomposite ($HDTC_{60}$-Ag), which is a better optical power limiter than the parent $HDTC_{60}$ and DT-Ag. Recently, gold nanoparticles have attracted considerable attention because of their potential applications as optical devices and nanoelectronic devices, given their higher surface-to-volume ratios and small size effects [486, 487]. However, less work has been done for optical nonlinearities of these nanostructural materials, especially for those with zero-valent metal nanoparticles [488]. Very recently, Fang et al. [489] synthesized two novel C_{60}-derived nanoparticles containing bipyridyl (BPY) and tripyridyl (TPY) groups. The optical limiting responses of C_{60}BPY, C_{60}TPY, C_{60}BPY-Au, and C_{60}TPY-Au nanoparticles in chloroform were measured by using a Nd:YAG laser at both picosecond and nanosecond scales and 532 nm. Their measured results are listed in Table 13. The difference between the optical limiting properties of both C_{60}BPY and C_{60}TPY is little. However, the optical limiting performance of C_{60}TPY-Au at both picosecond and nanosecond scales was much better than that of C_{60}BPY-Au. Their further Z-scan measurements indicate that nonlinear absorption (for example, excited-state and the surface plasmon absorptions) and nonlinear scattering contribute a lot to the optical limiting performance of C_{60}TPY-Au, whereas there is no nonlinear scattering for C_{60}BPY-Au. All these are caused by the excited-state interaction between the C_{60} and silver or gold nanoparticles, showing the strong nonlinear refraction (or, say, self-focusing effect), as well as the interband transition due to the surface plasmon resonance.

Suo et al. [490] explored the influence on the optical limiting performance of C_{60} by attaching two strong pulling electron groups (—CN), namely, dicyanodihydrofullerene $C_{60}(CN)_2$. This molecule has an optical limiting performance for nanosecond laser similar to that of pristine C_{60}. Although this molecule has an extinction coefficient of the excited triplet state larger than that of C_{60}, its low quantum yield leads to a lower triplet-state population and thus optical limiting behavior similar to that of C_{60}. As mentioned before, there are generally two kinds of mechanisms for optical limiting performance because of nonlinear absorption: dual photon absorption and RSA. Compared with the pristine C_{60}, the picosecond optical limiting performance of $C_{60}(CN)_2$ is power independent but assigned to the RSA of the excited singlet state [490].

The incorporation of fullerene moieties into polymer systems such as polycarbonate (PC) is a hot recent topic because of the interesting optical and electronic properties that the combined materials exhibit. The work of Li and colleagues [491, 492] indicated that increasing the electron-accepting ability of fullerene derivatives leads to stronger intramolecular action between TCNEO-C_{60} moiety and PC unit, and thus TCNEO-C_{60}-PCs (for example, PCC1, PCC3; see Table 13 for different TCNEO-C_{60}/PC feed ratio; TCNEO = tetracyanomethanoxymethano) show much better optical limiting properties than C_{60}-PCs. It should be pointed out that the optical limiting characteristics of TCNEO-C_{60}-PC can be described by the Golovlev, Garrett, and Chen model [493].

Fullerene compounds are very hydrophobic. Therefore, most of the related physical and chemical studies about their structure, properties, and reactions have been performed in nonpolar solvents, such as toluene or benzene. However, if the fullerenes are embedded in a suitable water-soluble molecule, water solubility can be achieved. γ-Cyclodextrin has been proven to be a rather suitable candidate for producing a water-soluable complex that has a stoichiometry of one C_{60} for two γ-cyclodextrins [494]. Using 532-nm, 10-ns laser pulses, Konstantaki et al. [495] observed significant optical limiting action from a C_{60}/γ-cyclodextrin complex, the effect being slightly lower than that of the pristine C_{60}. However, aging of the C_{60}/γ-cyclodextrin–water solution leads to the formation of aggregates that can enhance the optical limiting action by a factor of almost 2 [495].

It has been shown that appropriate fullerene functionalization increases the solubility in polar solvent and allows the preparation of solid materials via a sol–gel method [30, 496, 497]. Recently, Signorini et al. [497] compared the optical limiting behavior of a pirrolydinofullerene derivative (FULP) solution with a FULP-doped sol–gel sample. Compared with the green spectral region (532 nm) for both C_{60} and FULP in solution, enhanced optical limiting efficiency is demonstrated in the red spectral region (690 nm), that is, showing better optical limiting performances than pristine fullerene. The inclusion in the sol–gel matrix does not influence the optical limiting performances of the fullerene derivative. The RSA is still the predominant mechanism for the observed optical limiting behavior.

Several groups have shown that the optical limiting performance of fullerenes in solid matrix—for example, in a poly(methyl methacrylate) (PMMA) matrix—is rather different from that in solution [498–500]. The dramatic difference should be contributed to mechanisms other than the RSA because the ground-state and triplet-triplet transient absorption spectra of C_{60} in PMMA polymer film are similar to those in room-temperature toluene solution [500]. Riggs and Sun [443] systematically investigated the optical limiting properties of *tert*-bytyl methano-C_{60} carboxylate (TBM[60]CO) in room-temperature toluene solutions of different concentrations, in PMMA polymer films, and in highly viscous solvent-polymer blends. Similar to C_{60}, optical limiting performance of TBM[60]CO strongly depends on the fullerene solution concentrations due to bimolecular excited-state processes of the fullerenes. Recently, Riggs and Sun [501] studied a series of mono- and multiple functionalized fullerene derivatives [502] at different solution concentrations, giving a generalized understanding of the concentration-dependent optical limiting in fullerene solutions. The medium viscosity is shown to contribute to effects on the optical limiting contribution that are also associated with excited triplet-state bimolecular processes (in particular self-quenching and triplet-triplet annihilation), resulting in weaker optical limiting performance in highly viscous solvent-polymer blends [443]. For example, for the highly viscous solution (the presence of large quantities of PMMA polymer in toluene) of TMB[60]CO at 71% linear transmittance, the fluence saturation threshold value is about 0.17 J/cm^2, which is about 20% higher than that for TMB[60]CO in room-temperature toulene solution of the same linear transmittance [443]. For another example, at 69% linear transmittance, the fluence saturation threshold value for TBM[60]CO in a highly viscous blend of poly(propionylethyleneimine) (PPEI) polymer chloroform solution is significantly higher than in room-temperature chloroform solution [443]. Optical limiting properties of TBM[60]CO are found to be essentially independent of film thickness. Lacking meaningful molecular diffusion in polymer films may be the reason for the poor optical limiting behavior of TBM[60]CO in both thin and thick PMMA films than in room-temperature solution. Based on the modified RSA model including both unimolecular and bimolecular excited-state processes, Riggs and Sun [443]

indicated that the fullerene solution concentration affects the efficiency of the bimolecular excited-state processes, whereas the medium viscosity influences the bimolecular rate constants through changes in the molecular diffusivity.

6. COHERENT CONTROL THEORY OF NONLINEAR OPTICS

A key experiment to demonstrate wave behavior is the well-known Young's double-slit experiment addressed in high school physics. It provides us a basis for the rapidly developing field of physical science, coherent control of quantum dynamics of atoms and molecules [503–506]. The essence of coherent control lies in the interference principle of waves. As two or more competing pathways arrive at the same final state, quantum interference would happen. The frequency and intensity properties of lasers have long been used to probe and even alter properties of matter. However, the most characteristic property of laser light—namely, its well-defined phase—is rarely regarded as a control parameter. Manykin and Alfanas'ev [507] were the first to consider the interference of quantum mechanical transition amplitudes for two pathways coupling the same initial and final state of an atom, showing us a possibility of employing laser phase. In particular, simultaneous single- and three-photon absorptions were suggested to control state populations and hence the transmission of the medium [507]. This is analogous to the interference of two beams in a Young's double-slit experiment except that now one has an effective "matter interferometer," in that it is electrons that are being controlled.

Recently, the coherent control approach has been successfully demonstrated, both theoretically and experimentally, in unimolecular breakdown reactions [508–510], reactive scattering [511], electron distribution excited in a metal [512], the energy and angular distribution of autoionized electrons [513, 514], photocurrent generation in bulk semiconductors [515], spontaneous emission near a photonic band edge [516], light absorption and terahertz radiation in semiconductor nanostructures [517, 518], optical phonon emission rates and optical gain from electronic intersubband transitions in semiconductors [519, 520], optical dynamics in semiconductor microcavities [521], the polarization of an optical field [522], the total ionization yield in a two-color ionization process [523], and others [524–529]. Among the coherent control scenarios, a successful one is to employ two laser fields to induce a transition in an atom, a molecule, or a solid, and coherent control is achieved by varying the "external parameters" (such as the relative phase and amplitude of the two fields, which can be adjusted experimentally) so that the induced transition amplitudes interfere constructively or destructively.

In this section, we review a coherent control approach we proposed [24, 530] to study the molecular linear polarizability α, first hyperpolarizability β, and second hyperpolarizability γ at a desired frequency. In our coherent control scenario, an initial state, which comprises a superposition of two eigenstates of the radiation-free molecular Hamiltonian, is prepared, and both states are excited to the same final state by using two cw laser fields. As an example, we demonstrated that the γ magnitude of a nitrogen molecule at a desired frequency can be coherently controlled, either constructively enhanced or strongly decreased, by varying the relative amplitude and phases between two laser fields and those in the initially prepared superposition state. This approach opens a door to achieving large γ magnitude of molecules at a desired frequency, which is required for photonic applications. Based on this control approach, we are able to get a large γ value of carbon nanostructures at a desired frequency.

We denote $\widehat{H}_0$, ϵ_n, and $|\Phi_n(\mathbf{r})\rangle$ as the radiation-free molecular Hamiltonian, its discrete or continuous set of energy eigenvalues, and corresponding eigenfunctions, respectively. They satisfy the time-independent Schrödinger equation

$$\widehat{H}_0|\Phi_n(\mathbf{r})\rangle = \epsilon_n|\Phi_n(\mathbf{r})\rangle \tag{74}$$

where $\epsilon_n = \hbar\omega_n$ and ω_n is the angular frequency. Here it is assumed that these solutions are chosen in such a manner that they constitute a complete, orthonormal set satisfying the condition $\langle\Phi_m(\mathbf{r})|\Phi_n(\mathbf{r})\rangle = \delta_{mn}$.

In the presence of the radiation field, the time evolution of the molecule-field coupling system is governed by the Hamiltonian

$$\hat{H}(t) = \hat{H}_0 + \hat{V}(t) \tag{75}$$

where the interaction Hamiltonian, $\hat{V}(t)$, which describes the interaction of the molecule with the radiation field, is given in the dipole approximation by $\hat{V}(\mathbf{r}, t) = -\hat{u} \bullet \mathbf{F}(t)$, where $\hat{u} = -e\hat{\mathbf{r}}(t)$ is the electric dipole moment operator, $-e$ is the charge of the electron, and $\mathbf{F}(t)$ represents the radiation field. Assuming that all of the properties of the molecule-field coupling system can be described by the wave function $|\Psi(\mathbf{r}, t)\rangle$, we have

$$i\hbar \frac{\partial |\Psi(\mathbf{r}, t)\rangle}{\partial t} = \hat{H}(t)|\Psi(\mathbf{r}, t)\rangle \tag{76}$$

In general, th prior time-dependent Schrödinger equation cannot be solved exactly. So, it is often adequate to solve this equation through the use of perturbation theory. Exactly, to solve the time-dependent Schrödinger equation systematically in terms of a perturbation expansion, we replace the Hamiltonian $\hat{H}(t)$ by

$$\hat{H}(t) = \hat{H}_0 + \lambda \hat{V}(t) \tag{77}$$

where λ is a continuously varying parameter ranging from zero to unity that characterizes the strength of the interaction, and the value $\lambda = 1$ corresponds to the actual physical situation. Moreover, we seek a solution to the time-dependent Schrödinger's equation in the form of a power series in λ:

$$|\Psi(\mathbf{r}, t)\rangle = |\Psi^{(0)}(\mathbf{r}, t)\rangle + \lambda|\Psi^{(1)}(\mathbf{r}, t)\rangle + \lambda^2|\Psi^{(2)}(\mathbf{r}, t)\rangle + \cdots + \lambda^N|\Psi^{(N)}(\mathbf{r}, t)\rangle \tag{78}$$

Thereby, introducing it into the time-dependent Schrödinger equation and requiring that all terms that are proportional to λ^N satisfy the equality separately, we obtain the set of equations

$$i\hbar \frac{\partial}{\partial t}|\Psi^{(0)}(\mathbf{r}, t)\rangle = \hat{H}_0|\Psi^{(0)}(\mathbf{r}, t)\rangle \tag{79}$$

$$i\hbar \frac{\partial}{\partial t}|\Psi^{(N)}(\mathbf{r}, t)\rangle = \hat{H}_0|\Psi^{(N)}(\mathbf{r}, t)\rangle + \hat{V}(\mathbf{r}, t)|\Psi^{(N-1)}(\mathbf{r}, t)\rangle \qquad N = 1, 2, 3, \ldots \tag{80}$$

These equations are solved by making use of the fact that the energy eigenfunctions for the free molecule constitute a complete set of basis functions in terms of which any function can be expanded. Hence, the full time-dependent wavefunction $|\Psi^{(N)}(\mathbf{r}, t)\rangle$ can be expanded in the term of the eigenfunction $|\Phi_n(\mathbf{r})\rangle$ of the radiation-free molecular Hamiltonian $\hat{H}_0$

$$|\Psi^{(N)}(\mathbf{r}, t)\rangle = \sum_l C_l^{(N)}(t) e^{-i\epsilon_l t/\hbar}|\Phi_l(\mathbf{r})\rangle \tag{81}$$

where $C_l^{(N)}$ gives the probability amplitude that, to the Nth order in the perturbation, the molecule is in the energy eigenstate $|\Phi_l(\mathbf{r})\rangle$ at time t. Furthermore, we find that the probability amplitudes are given by

$$C_m^{(N)}(t) = (i\hbar)^{-1} \sum_l \int_{-\infty}^{t} V_{ml}(t') C_l^{(N-1)}(t') e^{i\omega_{ml}t'} \, dt' \tag{82}$$

where $\omega_{ml} = (\epsilon_m - \epsilon_l)/\hbar$ is the transition frequency between eigenstates $|\Phi_m(\mathbf{r})\rangle$ and $|\Phi_l(\mathbf{r})\rangle$ of $\hat{H}_0$, and we have introduced the matrix elements of the perturbing Hamiltonian, which are defined by $V_{ml}(t') = \langle \Phi_m(\mathbf{r})|\hat{V}(t')|\Phi_l(\mathbf{r})\rangle$. The form of $C_m^{(N)}(t)$ given here demonstrates the usefulness of the perturbation techniques. Once the probability amplitudes of order $(N-1)$ are determined, the amplitude of the next higher order (N) can be obtained by straightforward time integration.

Then, the development of our coherent control theory of nonlinear optics is based on the following control scenario. The molecule-field coupling system is initially prepared in a superposition of two eigenstates, $|\Phi_a(\mathbf{r})\rangle$ and $|\Phi_b(\mathbf{r})\rangle$, of the radiation-free molecular Hamiltonian $\widehat{H}_0$ [531–541]

$$|\Psi(\mathbf{r}, t=0)\rangle = C_a|\Phi_a(\mathbf{r})\rangle + C_b|\Phi_b(\mathbf{r})\rangle \tag{83}$$

where $|C_a|^2 + |C_b|^2 = 1$ and the relative phase between both states is defined by $\phi = \arg(\frac{C_a}{C_b})$. Both states are excited to the same final state (denoted by the energy ϵ) by using two cw laser fields, which are in the form [504, 541]

$$\mathbf{E}(t) = \mathbf{E}(\omega_1)e^{-i\omega_1 t} + \mathbf{E}(\omega_2)e^{-i\omega_2 t} \tag{84}$$

where $\hbar\omega_1 = \epsilon - \epsilon_a$ and $\hbar\omega_2 = \epsilon - \epsilon_b$. The relative amplitude and phase of both cw laser fields are defined by $\zeta = |\frac{\mathbf{E}(\omega_2)}{\mathbf{E}(\omega_1)}|$ and $\psi = \arg(\frac{\mathbf{E}(\omega_2)}{\mathbf{E}(\omega_1)})$, respectively.

Then, based on the perturbation theory and the quantum theory of molecular polarizability and hyperpolarizability [95], we got the probability amplitudes [24, 530]:

$$C_l^{(0)}(t) = C_l\delta_{la} + C_l\delta_{lb} \tag{85}$$

$$C_l^{(1)}(t) = C_a\hbar^{-1}\sum_{p=1}^{2}\frac{[U_{la}\bullet\mathbf{E}(\omega_p)]e^{i(\omega_{la}-\omega_p)t}}{\omega_{la}-\omega_p} + C_b\hbar^{-1}\sum_{p=1}^{2}\frac{[U_{lb}\bullet\mathbf{E}(\omega_p)]e^{i(\omega_{lb}-\omega_p)t}}{\omega_{lb}-\omega_p} \tag{86}$$

$$\begin{aligned} C_l^{(2)}(t) = {} & C_a\hbar^{-2}\sum_{p=1}^{2}\sum_{q=1}^{2}\sum_{m}\frac{[U_{lm}\bullet\mathbf{E}(\omega_q)][U_{ma}\bullet\mathbf{E}(\omega_p)]e^{i(\omega_{la}-\omega_p-\omega_q)t}}{(\omega_{la}-\omega_p-\omega_q)(\omega_{ma}-\omega_p)} \\ & + C_b\hbar^{-2}\sum_{p=1}^{2}\sum_{q=1}^{2}\sum_{m}\frac{[U_{lm}\bullet\mathbf{E}(\omega_q)][U_{mb}\bullet\mathbf{E}(\omega_p)]e^{i(\omega_{lb}-\omega_p-\omega_q)t}}{(\omega_{lb}-\omega_p-\omega_q)(\omega_{mb}-\omega_p)} \end{aligned} \tag{87}$$

$$\begin{aligned} C_l^{(3)}(t) = {} & C_a\hbar^{-3}\sum_{p=1}^{2}\sum_{q=1}^{2}\sum_{r=1}^{2}\sum_{mn}\frac{[U_{ln}\bullet\mathbf{E}(\omega_r)][U_{nm}\bullet\mathbf{E}(\omega_q)][U_{ma}\bullet\mathbf{E}(\omega_p)]e^{i(\omega_{la}-\omega_p-\omega_q-\omega_r)t}}{(\omega_{la}-\omega_p-\omega_q-\omega_r)(\omega_{ma}-\omega_p-\omega_q)(\omega_{na}-\omega_p)} \\ & + C_b\hbar^{-3}\sum_{p=1}^{2}\sum_{q=1}^{2}\sum_{r=1}^{2}\sum_{mn}\frac{[U_{ln}\bullet\mathbf{E}(\omega_r)][U_{nm}\bullet\mathbf{E}(\omega_q)][U_{mb}\bullet\mathbf{E}(\omega_p)]e^{i(\omega_{lb}-\omega_p-\omega_q-\omega_r)t}}{(\omega_{lb}-\omega_p-\omega_q-\omega_r)(\omega_{mb}-\omega_p-\omega_q)(\omega_{nb}-\omega_p)} \end{aligned} \tag{88}$$

where $U_{lj} = \langle\Phi_l(\mathbf{r})|\hat{u}|\Phi_j(\mathbf{r})\rangle$ is the transition-dipole matrix element between $|\Phi_l(\mathbf{r})\rangle$ and $|\Phi_j(\mathbf{r})\rangle$ of $\widehat{H}_0$.

First, using above results, we got the linear dipole moment $p^{(1)}(t)$ of molecule [24, 530]:

$$\begin{aligned} p^{(1)}(t) = {} & \hbar^{-1}\sum_l\Bigg\{|C_a|^2\left(\frac{[U_{la}\bullet\mathbf{E}(\omega_1)]U_{al}}{\omega_{la}-\omega_1} + \frac{[U_{al}\bullet\mathbf{E}(\omega_1)]U_{la}}{\omega_{la}^*+\omega_1}\right) \\ & + |C_b|^2\left(\frac{[U_{lb}\bullet\mathbf{E}(\omega_1)]U_{bl}}{\omega_{lb}-\omega_1} + \frac{[U_{bl}\bullet\mathbf{E}(\omega_1)]U_{lb}}{\omega_{lb}^*+\omega_1}\right) \\ & + C_a^*C_b\left(\frac{[U_{al}\bullet\mathbf{E}(\omega_2)]U_{lb}}{\omega_{lb}^*+\omega_1} + \frac{[U_{lb}\bullet\mathbf{E}(\omega_2)]U_{al}}{\omega_{la}-\omega_1}\right)\Bigg\}e^{-i\omega_1 t} \\ & + \hbar^{-1}\sum_l\Bigg\{|C_a|^2\left(\frac{[U_{la}\bullet\mathbf{E}(\omega_2)]U_{al}}{\omega_{la}-\omega_2} + \frac{[U_{al}\bullet\mathbf{E}(\omega_2)]U_{la}}{\omega_{la}^*+\omega_2}\right) \\ & + |C_b|^2\left(\frac{[U_{lb}\bullet\mathbf{E}(\omega_2)]U_{bl}}{\omega_{lb}-\omega_2} + \frac{[U_{bl}\bullet\mathbf{E}(\omega_2)]U_{lb}}{\omega_{lb}^*+\omega_2}\right) \\ & + C_aC_b^*\left(\frac{[U_{la}\bullet\mathbf{E}(\omega_1)]U_{bl}}{\omega_{lb}-\omega_2} + \frac{[U_{bl}\bullet\mathbf{E}(\omega_1)]U_{la}}{\omega_{la}^*+\omega_2}\right)\Bigg\}e^{-i\omega_2 t} \end{aligned}$$

$$+\hbar^{-1}\sum_{l}\left\{C_aC_b^*\left(\frac{[U_{la}\bullet\mathbf{E}(\omega_2)]U_{bl}}{\omega_{la}-\omega_2}+\frac{[U_{bl}\bullet\mathbf{E}(\omega_2)]U_{la}}{\omega_{lb}^*+\omega_2}\right)e^{-i(2\omega_2-\omega_1)t}\right.$$
$$\left.+C_a^*C_b\left(\frac{[U_{al}\bullet\mathbf{E}(\omega_1)]U_{lb}}{\omega_{la}^*+\omega_1}+\frac{[U_{lb}\bullet\mathbf{E}(\omega_1)]U_{al}}{\omega_{lb}-\omega_1}\right)e^{-i(2\omega_1-\omega_2)t}\right\} \quad (89)$$

where l is summed over all the dipole allowed transitions. Of course, if we wanted to describe the dephasing processes, which are not accompanied by the transfer of population, it would be appropriate to pursue this work with the density matrix formalism [95]. In the previous equation, the last sum term is a "satellite term," which does not affect the observable dipole moment at the desired angular frequency ω_1, or ω_2. Separating the average value of the induced dipole moment into its desired angular frequency (ω_1 and ω_2) components [95, 541] and based on the quantum theory of molecular polarizability introduced before, finally we got the linear polarizability $\alpha(\omega)$ of molecules at the desired angular frequencies ω_1 and ω_2, respectively [24, 530]:

$$\alpha_{ij}(\omega_1)=\hbar^{-1}\sum_{l}\left\{\kappa_{aa}\left(\frac{U_{al}^iU_{la}^j}{\omega_{la}-\omega_1}+\frac{U_{al}^jU_{la}^i}{\omega_{la}^*+\omega_1}\right)+\kappa_{bb}\left(\frac{U_{bl}^iU_{lb}^j}{\omega_{lb}-\omega_1}+\frac{U_{bl}^jU_{lb}^i}{\omega_{lb}^*+\omega_1}\right)\right.$$
$$\left.+\kappa_{ab}\left(\frac{U_{al}^jU_{lb}^i}{\omega_{lb}^*+\omega_1}+\frac{U_{al}^iU_{lb}^j}{\omega_{la}-\omega_1}\right)\right\} \quad (90)$$

$$\alpha_{ij}(\omega_2)=\hbar^{-1}\sum_{l}\left\{\kappa_{aa}\left(\frac{U_{al}^iU_{la}^j}{\omega_{la}-\omega_2}+\frac{U_{al}^jU_{la}^i}{\omega_{la}^*+\omega_2}\right)+\kappa_{bb}\left(\frac{U_{bl}^iU_{lb}^j}{\omega_{lb}-\omega_2}+\frac{U_{bl}^jU_{lb}^i}{\omega_{lb}^*+\omega_2}\right)\right.$$
$$\left.+\kappa_{ba}\left(\frac{U_{bl}^iU_{la}^j}{\omega_{lb}-\omega_2}+\frac{U_{bl}^jU_{la}^i}{\omega_{la}^*+\omega_2}\right)\right\} \quad (91)$$

where the five parameters in these equations are defined as follows [24, 530]:

$$\kappa_{aa}=|C_a|^2 \quad (92)$$
$$\kappa_{bb}=1-\kappa_{aa} \quad (93)$$
$$\kappa_{ab}=\sqrt{\kappa_{aa}\kappa_{bb}}\zeta e^{i\delta} \quad (94)$$
$$\kappa_{ba}=\sqrt{\kappa_{aa}\kappa_{bb}}\zeta^{-1}e^{-i\delta} \quad (95)$$
$$\delta=\psi-\phi \quad (96)$$

For convenience, these equations can be rewritten in a succinct form [24, 530]

$$\alpha(\omega_1)=\kappa_{aa}\alpha_{aa}^N(\omega_1)+\kappa_{bb}\alpha_{bb}^N(\omega_1)+\kappa_{ab}\alpha_{ab}^I(\omega_1) \quad (97)$$
$$\alpha(\omega_2)=\kappa_{aa}\alpha_{aa}^N(\omega_2)+\kappa_{bb}\alpha_{bb}^N(\omega_2)+\kappa_{ba}\alpha_{ba}^I(\omega_2) \quad (98)$$

Similarly, we are able to get the second- and third-order nonlinear optical polarizabilities of molecules at their desired angular frequencies in succinct forms. In detail, the second-order optical polarizability β of molecules at the desired frequency $\omega_s=2\omega_1, 2\omega_2, \omega_1+\omega_2$ are given by [24, 530]

$$\beta(\omega_s=2\omega_1)=\kappa_{aa}\beta_{aa}^N(\omega_s)+\kappa_{bb}\beta_{bb}^N(\omega_s)+\kappa_{ab}\beta_{ab}^I(\omega_s) \quad (99)$$
$$\beta(\omega_s=2\omega_2)=\kappa_{aa}\beta_{aa}^N(\omega_s)+\kappa_{bb}\beta_{bb}^N(\omega_s)+\kappa_{ba}\beta_{ba}^I(\omega_s) \quad (100)$$
$$\beta(\omega_s=\omega_1+\omega_2)=\kappa_{aa}\beta_{aa}^N(\omega_s)+\kappa_{bb}\beta_{bb}^N(\omega_s)+\kappa_{ab}\beta_{ab}^I(\omega_s)+\kappa_{ba}\beta_{ba}^I(\omega_s) \quad (101)$$

and the third-order nonlinear optical polarizabilities γ of molecule at the desired angular frequencies $\omega_s=3\omega_1, 3\omega_2, 2\omega_1+\omega_2, \omega_1+2\omega_2$ are given by [24, 530]

$$\gamma(\omega_s=3\omega_1)=\kappa_{aa}\gamma_{aa}^N(\omega_s)+\kappa_{bb}\gamma_{bb}^N(\omega_s)+\kappa_{ab}\gamma_{ab}^I(\omega_s) \quad (102)$$
$$\gamma(\omega_s=3\omega_2)=\kappa_{aa}\gamma_{aa}^N(\omega_s)+\kappa_{bb}\gamma_{bb}^N(\omega_s)+\kappa_{ba}\gamma_{ba}^I(\omega_s) \quad (103)$$

$$\gamma(\omega_s = 2\omega_1 + \omega_2) = \kappa_{aa}\gamma^{\Lambda}_{aa}(\omega_s) + \kappa_{bb}\gamma^{\Lambda}_{bb}(\omega_s) + \kappa_{ab}\gamma^{I}_{ab}(\omega_s) + \kappa_{ba}\gamma^{I}_{ba}(\omega_s) \quad (104)$$

$$\gamma(\omega_s = \omega_1 + 2\omega_2) = \kappa_{aa}\gamma^{\Lambda}_{aa}(\omega_s) + \kappa_{bb}\gamma^{\Lambda}_{bb}(\omega_s) + \kappa_{ab}\gamma^{I}_{ab}(\omega_s) + \kappa_{ba}\gamma^{I}_{ba}(\omega_s) \quad (105)$$

In these succinct equations, α^{Λ}_{ii}, β^{Λ}_{ii}, and γ^{Λ}_{ii} represent the normal polarizability with all kinds of transition processes both starting and ending in the state $|\Phi_i(\mathbf{r})\rangle$; α^{I}_{ij}, β^{I}_{ij}, and γ^{I}_{ij} are the new types of interference terms with all kinds of transition processes starting at the state $|\Phi_i(\mathbf{r})\rangle$ and ending at the state $|\Phi_j(\mathbf{r})\rangle$ (details about these terms are given in Appendix A). The structure of these equations is of the type we expect in the coherent control theory: each has one $\kappa_{aa}(\ldots)$ term associated with the excitation of the $|\Phi_a(\mathbf{r})\rangle$ state, and one $\kappa_{bb}(\ldots)$ term associated with the excitation of the $|\Phi_b(\mathbf{r})\rangle$ state, and one κ_{ab} term or/and one $\kappa_{ba}(\ldots)$ term corresponding to the interference between the two excitation routes. The interference term, which can be either constructive or destructive, is in general different for each case. What makes these equations so important in practice is that the interference term has coefficients whose magnitude and sign depend on experimentally controllable parameters. Thus, by varying the magnitude and the relative phase ϕ of the coefficients C_a and C_b, the relative phase ψ and the amplitude ratio ζ of two cw laser fields, we are able to alter directly the interference term and hence control the α, β, and γ of molecules at a desired frequency. The desired superposition state could be carried out by a number of different schemes [503–506, 531–539], for example, the scheme involving a two-photon absorption from a pump laser. Of course, because of collisions, spontaneous emission, or other dephasing effects [524–529], the loss of the coherence between $|\Phi_a(\mathbf{r})\rangle$ and $|\Phi_b(\mathbf{r})\rangle$ would cause the reduction in magnitude or the disappearance of the interference term.

For simplicity, we have applied this approach to the nitrogen molecule [542] and, as an example, report the results of the coherent control over the second-order hyperpolarizability γ at the desired frequency $3\omega_1$ and $3\omega_2$. We considered the nonresonant case and thus do not include the population decay rate in the numerical calculation. Because of quantum interference, we demonstrated that the nonresonant γ magnitude of molecule at a desired frequency can be coherently controlled, either constructively enhanced or strongly decreased, by varying the relative amplitude and phase between two laser fields and those in the initially prepared superposition state [24, 530]. This approach has opened a door to getting a large γ magnitude of molecules at a desired frequency, which is required for photonic applications [24, 530]. Surely, based on this theory, we are also able to get a large γ value of carbon nanostructures at a desired frequency.

7. SUMMARY, REMARKS, AND OUTLOOKS

In summary, we have introduced carbon nanostructures, nonlinear optical interactions, and theoretical tools for electronic structure and (hyper)polarizability calculations. Theoretical and experimental studies on the second- and third-order optical nonlinearities of carbon nanostructures are reviewed. In this chapter, we have also presented in detail the ESSH model where the Coulomb interaction has been included to describe the carbon nanostructures, including higher fullerenes and fullerene-related nanotubes, and to investigate the dynamical and static third-order optical nonlinearities of armchair, zigzag, and chiral carbon nanotubes. The cap, symmetry, size, and chiral effects are discussed, and two scaling laws of the static second-order hyperpolarizability of armchair and zigzag carbon nanotubes are empirically arrived at, respectively. By including the effect of the dopant ions into the ESSH model, a further ESSH model is developed to study the doping effect on the third-order optical nonlinearities of carbon nanostructures. Finally, we have reviewed the coherent control theory of nonlinear optics we developed to study molecular hyperpolarizability. Our numerical experiment has demonstrated that the second-order hyperpolarizability of molecule at a desired frequency can be coherently controlled, either greatly enhanced or heavily decreased, which opens a door to achieving the large nonlinear optical response of a molecule needed for photonic application.

As discussed before, C_{60} can act as electron donors or acceptors. This feature, plus the intrinsic SHG activity, has led researchers to explore the possibility of fabrication of charge-transfer complexes with creation of a permanent dipole moment at ground state and an

enhanced quadratic NLO response with a good temporal stability. Indeed, the fabrication of stable and efficient noncentrosymmetric molecules for quadratic NLO is an interesting challenge because of the large class of possible applications of these materials in tunable light sources, frequency converters, electro-optic modulation, ultrashort electric pulse generation, and so on. Because of the problems and especially the costs that were encountered with fabrication of noncentrosymmetric structures, artificial structures such as poled polymers have attracted increased interest. Controlled charge-transfer multilayered structures may offer an alternative solution, well argued by the controlled fabrication and expected enhanced stability. In this aspect, further studies on such structures are expected in the near future.

In previous theoretical studies, most works focused on carbon nanotubes of small size. To study typical carbon nanotubes, which are generically seen in experiment [for example, (10,10) carbon nanotube], it would be very interesting and significant to develop the theoretical models because of the computational complexity that arises from the big size of carbon nanotubes. On the other hand, the large NLO response in carbon nanostructures is mainly produced by delocalized π electrons, as in conjugated polymer chains. However, the three-dimensional character of spherically shaped carbon nanostructures may cause severe limitations on their NLO properties and thus make their γ magnitude become smaller than those of linear polymers containing a similar number of carbon atoms. Thus, it is possible that large NLO responses would be observed as carbon nanotubes are paralleled in a plane. This is interesting and can be expected to be done in the near future.

Bond-length alternation has been a good structural parameter to describe the electrical properties of molecules. When the electron–lattice interaction is introduced in fullerenes and carbon nanotubes, bond-length alternation is expected. So, what are the calculated hyperpolarizability values? Is it essential to take account of the electron–lattice interaction to obtain a large second-order hyperpolarizability γ? On the other hand, the bond-length alternation should depend on odd or even numbers of carbon rings to be added to the spherically shaped carbon nanostructures. Then, what is the effect of an odd or even number of rings on the second-order hyperpolarizability γ of fullerenes and carbon nanotubes? This is an interesting question because the γ magnitude of carbon nanotubes may be considered as the limit.

In the case of graphite, the valence orbital is a $\pi(2p_z)$ orbital, and there is no interaction between the π and $\sigma(2s \text{ and } 2p_{x,y})$ orbitals because of their different symmetries. In the case of a tubule, the curvature of the tubule gives rise to some mixing of σ and π bands; that is, the bottom of antibonding σ^* bands exits in the energy band of antibonding π bands. The study of the electronic structure of carbon nanotubes and the mixing effect of σ and π bands has shown that this effect is small at the Fermi energy within the tight-binding approximation for σ and π bands. Based on this, we have neglected the mixing of σ and π orbitals and have not considered the transition from π to σ^* in previous theoretical models. Surely, it would be interesting to further study the effect on the second-order hyperpolarizability γ of carbon nanotubes because of the transition from π to σ^*.

Also, it would be interesting to explore structure property relationships for optical nonlinearities of carbon nanotubes from two complementary angles: (1) to explore the possibility of obtaining structural information from nonlinear optical experiments and (2) to assess the potential of fullerenes, fullerites, and their derivatives for the practical nonlinear optical devices. Recently, it has been shown that screening the optical fields due to electron–electron interaction may reduce the optical susceptibilities of these molecules by a large factor. Experimental work is expected to clarify this issue.

In this chapter, the coherent control theory to study the molecular hyperpolarizability at a desired frequency is presented. Of course, it is possible for us to get larger γ values of carbon nanostructures at the desired frequency based on this theory. How is the superposition state established? How can we control and enhance their second-order hyperpolarizability by changing the magnitude or relative phase of the coefficient of the superposition state and those of two incident coherent lasers? As well as these questions, a number of other factors should be considered in the coherent control theory, for example, degrading processes due to Doppler broadening and collisions, as well as the decoherence due to collisions.

The advantage of the type of chromophore proposed by Asselberghs et al. [400] is that the switchability can be addressed in two different ways: proton transfer and redox switchabilities. As we know, electro-optical (EO) modulation is the main application for second-order NLO materials. In EO modulators, an applied electric field changes the refractive index of the NLO material through the EO coefficient. Based on the work of Asselberghs et al. [400], embedding the chromophore with the first-order hyperpolarizability, which can be switched by proton transfer in a matrix with such a photoaddressable proton donor, can make an EO modulator that can be controlled by light, that is, an "on" state without light and an "off" state with light. On the other hand, the combination of a second-order NLO chromophore with a photoaddressable proton donor would constitute a composite third-order optical material because the second-order NLO polarizability of a molecule would be influenced by light. This may allow for all-optical data manipulation, and the composite materials may provide a new driving force in the field of third-order NLO materials.

As shown before, the smaller the π-π overlap in the three-dimensional conjugated spheres (for example, the isomeric fullerenes), the higher the barrier for the π-conjugated electron flow and thus the lowering of the second hyperpolarizability. Such concepts can be extended to explain the γ values of one-dimensional and two-dimensional π-conjugated systems, for example, *trans*-1,3-butadiene, *trans*-1,3,5-hexatriene, styrene, biphenyl, and 2,2-bithiophene.

As discussed previously, the enhancement of the third-order optical nonlinearity of carbon nanostructures is proportional to the increase of charge-transfer strength. This is a very important rule for us to design and synthesize new molecules for third-order optical nonlinear applications. This is also the essence of the donor–acceptor NLO model. Recently, Xie et al. [543] showed that $C_{60-m}N_m$ and $C_{60-n}B_n$ heterofullerenes are good donors and acceptors, respectively [543]. Then, new NLO nanostructures could be designed by linking them with a spacer made of one or several units of conjugated molecules, such as a polyene or an aromatic chain. In such NLO nanostructures, charge would migrate from $C_{60-m}N_m$ to $C_{60-n}B_n$ upon electronic excitation, giving rise to a large dipole moment along the direction connecting the donor-acceptor pair. Thus, large NLO response is expected in these donor- and acceptor-based nanostructures. In particular, they can be engineered by controlling the dopant number n and m.

Recently, Bockrath et al. [544] reported a controlled chemical doping of individual semiconducting SWNT ropes with potassium. Based on their scheme, they are able to control the level of doping by reversibly intercalating and deintercalating potassium, and potassium doping could change the carriers in the ropes from holes to electrons [544]. Their experiments open the way toward other experiments that require controlled doping, such as making nanoscale p-n junctions. Xie et al. [543] demonstrated that incorporating $C_{60-m}N_m$ and $C_{60-n}B_n$ into a semiconducting SWNT would result in n- and p-type SWNT-based nanostructures. What are the NLO properties of this kind of nanostructure?

The small fullerene C_{36} is much more reactive than C_{60} and C_{70}. It has been shown that the ground state of C_{36} is of diradical nature and the crystal C_{36} is formed through covalent bonds among C_{36} units. Obviously, this is contrasted with C_{60} and C_{70}. On the other hand, bounding the electron donor groups to C_{36} would remove the strong diradical character and stabilize the entire system, showing the possibility of synthesizing these molecules. Hence C_{36} chemistry will be further explored to study the possibility of using C_{36} derivatives for building NLO materials.

The improvement of the optical limiting performance observed in the molybdenum complex of C_{60} or C_{70}, which is attributed to the increment on the triplet-state absorption by intramolecular charge transfer, shows that the charge-transfer nature would make metal-fullerene complexes good candidates for optical limiting applications and leads to a new means for significantly improving the performance of fullerenes.

The solubility of C_{60} can be improved by combining it with an alkyl group of substituted CuPc through Diels-Alder reaction. The combined molecules exhibit good optical limiting performance. Hence, based on the combination of various optical limiting molecules through covalent bonds, broadband optical limiting materials may be designed, and phase separation problems, which exist in the mixture of various materials, can also be avoided.

The large optical limiting performance observed from the novel nanocomposites DTC_{60}-Ag or $HDTC_{60}$-Ag at 532 nm is due to the combination effort from the C_{60} derivative and silver nanoparticles. Thus, these kinds of novel nanocomposites would lead to novel optoelectronics as well as catalytic properties assembled by fullerene and metals. Surely, much work needs to be done to understand the interesting photophysical phenomena observed in these systems.

The attachment of two cyano groups to C_{60} reduces its quantum yield, increases the absorption of the excited triplet state, and leads to $C_{60}(CN)_2$, which has a similar optical limiting behavior to C_{60} for nanosecond laser at 532 nm. Because the differential absorption between the excited triplet state and the ground state in $C_{60}(CN)_2$ is larger in the near-infrared region, we expect that this molecule may also have better optical limiting performance in the near-infrared region than at 532 nm.

Different from adsorption or noncovalent functionalization, the molecule can attach to the sidewall of a carbon nanotube via a tube-molecule covalent bond [469]. Such covalent functionalization might lead to new opportunities in nanotube-based materials and devices [469]. Experiments on nanotube covalent functionalization start with the fluorination of SWNTs and the substitution reaction of fluorinated SWNTs in solutions. On the other hand, direct functionalization to the sidewall of SWNTs by various chemical groups, such as atomic hydrogen, aryl groups, nitrenes, carbenes, and radicals, COOH and NH_2, *N* alkylidene amino groups, alkyl groups, and aniline have been reported [469]. The covalent bond formed between functional groups and a carbon nanotube sidewall is expected to disturb the perfect tube π bonds via the local sp^3 rehybridization. Thus, the electronic and optical properties of carbon nanotube should be modified by the functionalization. It was found that the band-to-band transition features of π electrons in the UV-visible spectra of pristine SWNTs disappear upon covalent functionalization. (See the detailed review by Zhao and Xie [469].) Hence, it would be interesting to investigate the effect of covalent sidewall functionalization on the NLO properties of pristine carbon nanotubes.

Most theoretical and experimental works on the NLO of carbon nanostructures were focused on the spherically and rod-shaped carbon nanostructures. There are no studies on the NLO properties of several interesting carbon nanostructures, for example, carbon onions, carbon nanoparticles, carbon nanoparallelepipeds, rings of single-walled carbon nanotubes, graphitic nanocones, carbon nanohorns, carbon-beaded cone microstructures, and carbon microtrees. This is an open and interesting important area for the near future.

TD-DFT is a relatively expensive numerical method that cannot be employed to calculate the properties of very large systems. The recent development of a linear-scaling TD-DFT by Yam, Yokojima, and Chen [145] is exciting news in the computational community because it scales as $O(N)$, where N is the number of atoms involved. This method has successfully evaluated the linear optical response of large molecular systems. It would be interesting to develop a NLO formalism within this theory for studying the NLO properties of carbon and other nanostructures.

As mentioned in Section 2, the tight-binding approach to TD-DFT (TD-DFT-TB) developed by Niehaus et al. [146] is able to study the linear optical response of large nanosystems with hundreds of atoms. As of now, there is no report about the application of this approach to studying NLO responses of large nanosystems. Hence, the development of the NLO formalism within TD-DFT-TB would be also very interesting and important for studying the NLO properties of carbon and other nanostructures.

ACKNOWLEDGMENTS

One of us (RHX) would like to thank two of his colleagues, Lasse Jensen and Qin Rao, for helpful discussions and contributions to Section 4; Qin Rao for the design of Fig. 3; Garnett W. Bryant, Vedene H. Smith Jr., Jijun Zhao, Zhongfang Chen, and Jiangbin Gong for valuable comments; and the Alexander von Humboldt Foundation, the Natural Science and Engineering Research Council of Canada, the Chinese Center of Advanced Science and Technology (World Laboratory), National Laboratory of Solid State Microstructures in China, and Tian Ma Microelectronics Co. Ltd. in Shenzhen, China, for support.

APPENDIX: DETAILS OF INTERFERENCE TERMS

In this appendix, the details of the Cartesian components of interference terms, α^I_{ab}, α^I_{ba}, β^I_{ab}, β^I_{ba}, γ^I_{ab}, and γ^I_{ba} at the desired frequency ω_s, presented in the chapter, are summarized. Here the index $ijkh$ denotes the Cartesian components.

I. Interference terms of molecular polarizability α:

$$\alpha^I_{ab}(\omega_s = \omega_1)|_{ij} = \hbar^{-1} \sum_l \left\{ \frac{U^j_{al} U^i_{lb}}{\omega^*_{lb} + \omega_1} + \frac{U^i_{al} U^j_{lb}}{\omega_{la} - \omega_1} \right\} \tag{A1}$$

$$\alpha^I_{ba}(\omega_s = \omega_2)|_{ij} = \hbar^{-1} \sum_l \left\{ \frac{U^i_{bl} U^j_{la}}{\omega_{lb} - \omega_2} + \frac{U^j_{bl} U^i_{la}}{\omega^*_{la} + \omega_2} \right\} \tag{A2}$$

II. Interference terms of first hyperpolarizability β:

$$\begin{aligned}\beta^I_{ab}(\omega_s = 2\omega_1)|_{ijk} = \hbar^{-2} \sum_{lm} \Bigg\{ & \frac{U^i_{al} U^j_{lm} U^k_{mb}}{(\omega_{la} - 2\omega_1)(\omega_{mb} - \omega_1)} + \frac{U^j_{am} U^k_{ml} U^i_{lb}}{(\omega^*_{lb} + 2\omega_1)(\omega^*_{ma} + \omega_1)} \\ & + \frac{U^j_{am} U^i_{ml} U^k_{lb}}{(\omega^*_{mb} + \omega_1)(\omega_{lb} - \omega_1)} + \frac{U^i_{al} U^j_{lm} U^k_{mb}}{(\omega_{la} - 2\omega_1)(\omega_{ma} - \omega_1)} \\ & + \frac{U^j_{am} U^k_{ml} U^i_{lb}}{(\omega^*_{lb} + 2\omega_1)(\omega^*_{mb} + \omega_1)} + \frac{U^j_{am} U^i_{ml} U^k_{lb}}{(\omega^*_{ma} + \omega_1)(\omega_{la} - \omega_1)} \Bigg\}\end{aligned} \tag{A3}$$

$$\begin{aligned}\beta^I_{ba}(\omega_s = 2\omega_2)|_{ijk} = \hbar^{-2} \sum_{lm} \Bigg\{ & \frac{U^i_{bl} U^j_{lm} U^k_{ma}}{(\omega_{lb} - 2\omega_2)(\omega_{mb} - \omega_2)} + \frac{U^j_{bm} U^k_{ml} U^i_{la}}{(\omega^*_{la} + 2\omega_2)(\omega^*_{ma} + \omega_2)} \\ & + \frac{U^j_{bm} U^i_{ml} U^k_{la}}{(\omega^*_{mb} + \omega_2)(\omega_{lb} - \omega_2)} + \frac{U^i_{bl} U^j_{lm} U^k_{ma}}{(\omega_{lb} - 2\omega_2)(\omega_{ma} - \omega_2)} \\ & + \frac{U^j_{bm} U^k_{ml} U^i_{la}}{(\omega^*_{la} + 2\omega_2)(\omega^*_{mb} + \omega_2)} + \frac{U^j_{bm} U^i_{nl} U^k_{la}}{(\omega^*_{ma} + \omega_2)(\omega_{la} - \omega_2)} \Bigg\}\end{aligned} \tag{A4}$$

$$\begin{aligned}\beta^I_{ba}(\omega_s = \omega_1 + \omega_2)|_{ijk} = \hbar^{-2} \sum_{lm} \Bigg\{ & \frac{U^i_{bl} U^j_{lm} U^k_{ma}}{(\omega_{lb} - \omega_1 - \omega_2)(\omega_{ma} - 2\omega_1)} + \frac{U^k_{bm} U^j_{ml} U^i_{la}}{(\omega^*_{la} + \omega_1 + \omega_2)(\omega^*_{ma} + 2\omega_2)} \\ & + \frac{U^j_{bm} U^i_{ml} U^k_{la}}{(\omega^*_{ma} + 2\omega_2)(\omega_{la} - 2\omega_1)} \Bigg\}\end{aligned} \tag{A5}$$

$$\begin{aligned}\beta^{II}_{ab}(\omega_s = \omega_1 + \omega_2)|_{ijk} = \hbar^{-2} \sum_{lm} \Bigg\{ & \frac{U^i_{al} U^j_{lm} U^k_{mb}}{(\omega_{la} - \omega_1 - \omega_2)(\omega_{ma} - 2\omega_1)} + \frac{U^j_{am} U^k_{ml} U^i_{lb}}{(\omega^*_{lb} + \omega_1 + \omega_2)(\omega^*_{ma} + 2\omega_2)} \\ & + \frac{U^j_{am} U^i_{ml} U^k_{lb}}{(\omega^*_{ma} + 2\omega_2)(\omega_{la} - 2\omega_1)} \Bigg\}\end{aligned} \tag{A6}$$

III. Interference terms of the second hyperpolarizability γ:

$$\begin{aligned}\gamma^I_{ab}(\omega_s = 3\omega_1)|_{ijkh} = \hbar^{-3} \sum_{lmn} \Bigg\{ & \frac{U^i_{al} U^j_{ln} U^k_{nm} U^h_{mb}}{(\omega_{la} - 3\omega_1)(\omega_{nb} - 2\omega_1)(\omega_{mb} - \omega_1)} + \frac{U^h_{am} U^k_{mn} U^j_{nl} U^i_{lb}}{(\omega^*_{lb} + 3\omega_1)(\omega^*_{na} + 2\omega_1)(\omega^*_{ma} + \omega_1)} \\ & + \frac{U^h_{an} U^i_{nl} U^j_{lm} U^k_{mb}}{(\omega_{lb} - 2\omega_1)(\omega_{mb} - \omega_1)(\omega^*_{nb} + \omega_1)} + \frac{U^k_{am} U^j_{ml} U^i_{ln} U^h_{nb}}{(\omega^*_{la} + 2\omega_1)(\omega^*_{ma} + \omega_1)(\omega_{na} - \omega_1)} \\ & + \frac{U^i_{al} U^j_{ln} U^k_{nm} U^h_{mb}}{(\omega_{la} - 3\omega_1)(\omega_{na} - 2\omega_1)(\omega_{mb} - \omega_1)} + \frac{U^h_{am} U^k_{mn} U^j_{nl} U^i_{lb}}{(\omega^*_{lb} + 3\omega_1)(\omega^*_{nb} + 2\omega_1)(\omega^*_{ma} + \omega_1)}\end{aligned}$$

$$+\frac{U^h_{an}U^i_{nl}U^j_{lm}U^k_{mb}}{(\omega_{la}-2\omega_1)(\omega_{mb}-\omega_1)(\omega^*_{na}+\omega_1)}+\frac{U^k_{am}U^j_{ml}U^i_{ln}U^h_{nb}}{(\omega^*_{lb}+2\omega_1)(\omega^*_{ma}+\omega_1)(\omega_{nb}+\omega_1)}$$
$$+\frac{U^i_{al}U^j_{ln}U^k_{nm}U^h_{mb}}{(\omega_{la}-3\omega_1)(\omega_{na}-2\omega_1)(\omega_{ma}-\omega_1)}+\frac{U^h_{am}U^k_{mn}U^j_{nl}U^i_{lb}}{(\omega^*_{lb}+3\omega_1)(\omega^*_{nb}+2\omega_1)(\omega^*_{mb}+\omega_1)}$$
$$+\frac{U^h_{an}U^i_{nl}U^j_{lm}U^k_{mb}}{(\omega_{la}-2\omega_1)(\omega_{ma}-\omega_1)(\omega^*_{na}+\omega_1)}+\frac{U^k_{am}U^j_{ml}U^i_{ln}U^h_{nb}}{(\omega^*_{lb}+2\omega_1)(\omega^*_{mb}+\omega_1)(\omega_{nb}+\omega_1)}\Bigg\} \quad \text{(A7)}$$

$$\gamma^I_{ba}(\omega_s=3\omega_2)|_{ijkh}$$
$$=\hbar^{-3}\sum_{lmn}\Bigg\{\frac{U^i_{bl}U^j_{ln}U^k_{nm}U^h_{ma}}{(\omega_{lb}-3\omega_2)(\omega_{nb}-2\omega_2)(\omega_{mb}-\omega_2)}+\frac{U^h_{bm}U^k_{mn}U^j_{nl}U^i_{la}}{(\omega^*_{la}+3\omega_2)(\omega^*_{na}+2\omega_2)(\omega^*_{ma}+\omega_2)}$$
$$+\frac{U^h_{bn}U^i_{nl}U^j_{lm}U^k_{ma}}{(\omega_{lb}-2\omega_2)(\omega_{mb}-\omega_2)(\omega^*_{nb}+\omega_2)}+\frac{U^k_{bm}U^j_{ml}U^i_{ln}U^h_{na}}{(\omega^*_{la}+2\omega_2)(\omega^*_{ma}+\omega_2)(\omega_{na}+\omega_2)}$$
$$+\frac{U^i_{bl}U^j_{ln}U^k_{nm}U^h_{ma}}{(\omega_{lb}-3\omega_2)(\omega_{nb}-2\omega_2)(\omega_{ma}-\omega_2)}+\frac{U^h_{bm}U^k_{mn}U^j_{nl}U^i_{la}}{(\omega^*_{la}+3\omega_2)(\omega^*_{na}+2\omega_2)(\omega^*_{mb}+\omega_2)}$$
$$+\frac{U^h_{bn}U^i_{nl}U^j_{lm}U^k_{ma}}{(\omega_{lb}-2\omega_2)(\omega_{ma}-\omega_2)(\omega^*_{nb}+\omega_2)}+\frac{U^k_{bm}U^j_{ml}U^i_{ln}U^h_{na}}{(\omega^*_{la}+2\omega_2)(\omega^*_{mb}+\omega_2)(\omega_{na}+\omega_2)}$$
$$+\frac{U^i_{bl}U^j_{ln}U^k_{nm}U^h_{ma}}{(\omega_{lb}-3\omega_2)(\omega_{na}-2\omega_2)(\omega_{ma}-\omega_2)}+\frac{U^h_{bm}U^k_{mn}U^j_{nl}U^i_{la}}{(\omega^*_{la}+3\omega_2)(\omega^*_{nb}+2\omega_2)(\omega^*_{mb}+\omega_2)}$$
$$+\frac{U^h_{bn}U^i_{nl}U^j_{lm}U^k_{ma}}{(\omega_{la}-2\omega_2)(\omega_{ma}-\omega_2)(\omega^*_{na}+\omega_2)}+\frac{U^k_{bm}U^j_{ml}U^i_{ln}U^h_{na}}{(\omega^*_{lb}+2\omega_2)(\omega^*_{mb}+\omega_2)(\omega_{nb}-\omega_2)}\Bigg\} \quad \text{(A8)}$$

$$\gamma^I_{ab}(\omega_s=2\omega_1+\omega_2)|_{ijkh}$$
$$=\hbar^{-3}\sum_{lmn}\Bigg\{\frac{U^i_{al}U^j_{ln}U^k_{nm}U^h_{mb}}{(\omega_{lb}-\omega_1-2\omega_2)(\omega_{na}-2\omega_1)(\omega_{mb}-\omega_1)}+\frac{U^h_{am}U^k_{mn}U^j_{nl}U^i_{lb}}{(\omega^*_{la}+\omega_1+2\omega_2)(\omega^*_{nb}+2\omega_1)(\omega^*_{ma}+\omega_1)}$$
$$+\frac{U^k_{an}U^i_{nl}U^j_{lm}U^h_{mb}}{(\omega_{la}-2\omega_1)(\omega_{mb}-\omega_1)(\omega^*_{na}+\omega_2)}+\frac{U^h_{am}U^j_{ml}U^i_{ln}U^k_{nb}}{(\omega^*_{lb}+2\omega_1)(\omega^*_{ma}+\omega_1)(\omega_{nb}+\omega_2)}$$
$$+\frac{U^i_{al}U^j_{ln}U^h_{nm}U^k_{mb}}{(\omega_{lb}-2\omega_2-\omega_1)(\omega_{na}-2\omega_1)(\omega_{mb}-\omega_2)}+\frac{U^k_{am}U^h_{mn}U^j_{nl}U^i_{lb}}{(\omega^*_{la}+2\omega_2+\omega_1)(\omega^*_{nb}+2\omega_1)(\omega^*_{ma}+\omega_2)}$$
$$+\frac{U^k_{an}U^i_{nl}U^h_{lm}U^j_{b}}{(\omega_{la}-2\omega_1)(\omega_{mb}-\omega_2)(\omega^*_{na}+\omega_2)}+\frac{U^j_{am}U^h_{ml}U^i_{ln}U^k_{nb}}{(\omega^*_{lb}+2\omega_1)(\omega^*_{ma}+\omega_2)(\omega_{nb}+\omega_2)}$$
$$+\frac{U^i_{al}U^h_{ln}U^j_{nm}U^k_{mb}}{(\omega_{lb}-2\omega_2-\omega_1)(\omega_{nb}-2\omega_2)(\omega_{mb}-\omega_2)}+\frac{U^k_{am}U^j_{mn}U^h_{nl}U^i_{lb}}{(\omega^*_{la}+2\omega_2+\omega_1)(\omega^*_{na}+2\omega_2)(\omega^*_{ma}+\omega_2)}$$
$$+\frac{U^h_{an}U^i_{nl}U^j_{lm}U^k_{mb}}{(\omega_{lb}-2\omega_2)(\omega_{mb}-\omega_2)(\omega^*_{na}+\omega_1)}+\frac{U^k_{am}U^j_{ml}U^i_{ln}U^h_{nb}}{(\omega^*_{la}+2\omega_2)(\omega^*_{ma}+\omega_2)(\omega_{nb}+\omega_1)}\Bigg\} \quad \text{(A9)}$$

$$\gamma^I_{ba}(\omega_s=2\omega_1+\omega_2)|_{ijkh}$$
$$=\hbar^{-3}\sum_{lmn}\Bigg\{\frac{U^i_{bl}U^j_{ln}U^k_{nm}U^h_{ma}}{(\omega_{la}-3\omega_1)(\omega_{na}-2\omega_1)(\omega_{ma}-\omega_1)}+\frac{U^h_{bm}U^k_{mn}U^j_{nl}U^i_{la}}{(\omega^*_{lb}+3\omega_1)(\omega^*_{nb}+2\omega_1)(\omega^*_{mb}+\omega_1)}$$
$$+\frac{U^h_{bn}U^i_{nl}U^j_{lm}U^k_{ma}}{(\omega_{la}-2\omega_1)(\omega_{ma}-\omega_1)(\omega^*_{nb}+\omega_1)}+\frac{U^k_{bm}U^j_{ml}U^i_{ln}U^h_{na}}{(\omega^*_{lb}+2\omega_1)(\omega^*_{mb}+\omega_1)(\omega_{na}+\omega_1)}\Bigg\} \quad \text{(A10)}$$

$$\gamma^{ijkh}_{lmn}(\omega_4=\omega_1+2\omega_2)|_{ijkh}$$

$$=\hbar^{-3}\sum_{lmn}\left\{-\frac{U^i_{bl}U^j_{ln}U^k_{mn}U^h_{ma}}{(\omega_{la}-2\omega_1-\omega_2)(\omega_{na}-2\omega_1)(\omega_{ma}-\omega_1)}+\frac{U^h_{bn}U^k_{nm}U^j_{ml}U^i_{la}}{(\omega^*_{lb}+2\omega_1+\omega_2)(\omega^*_{nb}+2\omega_1)(\omega^*_{mb}+\omega_1)}\right.$$

$$+\frac{U^h_{bn}U^j_{nl}U^i_{lm}U^k_{ma}}{(\omega_{la}-2\omega_1)(\omega_{ma}-\omega_1)(\omega^*_{nb}+\omega_2)}+\frac{U^k_{bm}U^j_{ml}U^i_{ln}U^h_{na}}{(\omega^*_{lb}+2\omega_1)(\omega^*_{mb}+\omega_1)(\omega_{na}+\omega_2)}$$

$$+\frac{U^i_{bl}U^j_{ln}U^k_{nm}U^h_{ma}}{(\omega_{la}-2\omega_1-\omega_2)(\omega_{nb}-2\omega_2)(\omega_{ma}-\omega_1)}+\frac{U^h_{bm}U^k_{mn}U^j_{nl}U^i_{la}}{(\omega^*_{lb}+2\omega_1+\omega_2)(\omega^*_{na}+2\omega_2)(\omega^*_{mb}+\omega_1)}$$

$$+\frac{U^h_{bn}U^j_{nl}U^i_{lm}U^k_{ma}}{(\omega_{lb}-2\omega_2)(\omega_{ma}-\omega_1)(\omega^*_{nb}+\omega_1)}+\frac{U^k_{bm}U^j_{ml}U^i_{ln}U^h_{na}}{(\omega^*_{la}+2\omega_2)(\omega^*_{mb}+\omega_1)(\omega_{na}+\omega_1)}$$

$$+\frac{U^i_{bl}U^j_{ln}U^k_{nm}U^h_{ma}}{(\omega_{la}-\omega_2-2\omega_1)(\omega_{nb}-2\omega_2)(\omega_{ma}-\omega_2)}+\frac{U^h_{bm}U^k_{mn}U^j_{nl}U^i_{la}}{(\omega^*_{lb}+\omega_2+2\omega_1)(\omega^*_{na}+2\omega_2)(\omega^*_{mb}+\omega_2)}$$

$$\left.+\frac{U^h_{bn}U^j_{nl}U^i_{lm}U^k_{ma}}{(\omega_{lb}-2\omega_2)(\omega_{ma}-\omega_2)(\omega^*_{nl}+\omega_1)}+\frac{U^k_{bm}U^j_{ml}U^i_{ln}U^h_{na}}{(\omega^*_{la}+2\omega_2)(\omega^*_{mb}+\omega_2)(\omega_{na}+\omega_1)}\right\} \qquad \text{(A11)}$$

$$\gamma^{ijkh}_{abn}(\omega_4=\omega_1+2\omega_2)|_{ijkh}$$

$$=\hbar^{-3}\sum_{lmn}\left\{\frac{U^i_{al}U^j_{ln}U^k_{nm}U^h_{mb}}{(\omega_{lb}-3\omega_2)(\omega_{nb}-2\omega_2)(\omega_{mb}-\omega_2)}+\frac{U^h_{am}U^k_{mn}U^j_{nl}U^i_{lb}}{(\omega^*_{la}+3\omega_2)(\omega^*_{na}+2\omega_2)(\omega^*_{ma}+\omega_2)}\right.$$

$$\left.+\frac{U^h_{an}U^j_{nl}U^i_{lm}U^k_{mb}}{(\omega_{lb}-2\omega_2)(\omega_{mb}-\omega_2)(\omega^*_{na}+\omega_2)}+\frac{U^k_{am}U^j_{ml}U^i_{ln}U^h_{nb}}{(\omega^*_{la}+2\omega_2)(\omega^*_{ma}+\omega_2)(\omega_{nb}+\omega_2)}\right\} \qquad \text{(A12)}$$

REFERENCES

1. M. Ratner and D. Ratner, "Nanotechnology: A Gentle Introduction to the Next Big Idea." Prentice Hall, Englewood Cliffs, NJ, 2002.
2. C. P. Poole Jr. and F. J. Owens, "Introduction to Nanotechnology." Wiley, Hoboken, NJ, 2003.
3. H. S. Nalwa, Ed., "Encyclopedia of Nanoscience and Nanotechnology," Vol. 1–10. American Scientific Publishers, Stevenson Ranch, CA, 2004.
4. G. Benedek, P. Milani, and V. G. Ralchenko, Eds., "Nanostructured Carbon for Advanced Applications." Kluwer, Dordrecht, 2001.
5. R. Setton, P. Bernier, and S. Lefrant, "Carbon Molecules and Materials." Taylor & Francis, London, 2002.
6. A. A. Lucas, F. Moreau, and P. Lambin, *Rev. Mod. Phys.* 74, 1 (2002).
7. L. A. Rohlfing, D. M. Cox, and A. Kaldor, *J. Chem. Phys.* 81, 3322 (1984).
8. K. Raghavachari and J. S. Binkley, *J. Chem. Phys.* 87, 2191 (1987).
9. H. W. Kroto, J. R. Heath, S. C. Obrien, R. E. Curl, and R. E. Smalley, *Nature (London)* 318, 162 (1985).
10. E. Osawa, *Kagaku (Kyoto)*, 25, 854 (1970).
11. H. W. Kroto, J. E. Fischer, and D. E. Cox, "The Fullerenes." Pergamon, Oxford, 1993.
12. A. Hirsch, "The Chemistry of the Fullerenes." Thieme, New York, 1994.
13. R. E. Smalley, *Rev. Mod. Phys.* 69, 723 (1997).
14. R. F. Curl, *Rev. Mod. Phys.* 69, 691 (1997).
15. H. W. Kroto, *Rev. Mod. Phys.* 69, 703 (1997).
16. R. F. Curl and R. E. Smalley, *Sci. Am.* 264, 54 (1991).
17. W. Krätschmer, L. D. Lamb, K. Fostiropoulos, and D. R. Huffman, *Nature (London)* 347, 354 (1990).
18. D. R. Huffman, *Phys. Today* 44, 22 (1991).
19. W. E. Billups and M. A. Ciufolini, Ed., "Buckminsterfullerenes." VCH, New York, 1993.
20. P. W. Fowler and D. E. Manlopoulos, "An Atlas of Fullerenes." Clarendon, Oxford, 1995.
21. J. Cioslowski, "Electronic Structure Calculations on Fullerenes and Their Derivatives." Oxford University Press, New York, 1995.
22. M. S. Dresselhaus, G. Dresselhaus, and P. C. Eklund, "Science of Fullerenes and Carbon Nanotubes." Academic, New York, 1996.
23. W. Andreoni, Ed., "The Physics of Fullerene-Based and Fullerene-Related Materials." Kluwer, New York, 2000.

24. R. H. Xie, in "Handbook of Advanced Electronic and Photonic Materials and Devices," (H. S. Nalwa Ed.), Vol. 9, pp. 267–307, Nonlinear Optical Materials, Academic, New York, 2000.
25. R. H. Xie, Q. Rao, and L. Jensen, in "Encyclopedia of Nanoscience and Nanotechnology" (H. S. Nalwa, Ed.), American Scientific Publisher, CA, 2004.
26. R. H. Xie, J. Zhao, and Q. Rao, in "Encyclopedia of Nanoscience and Nanotechnology" (H. S. Nalwa, Ed.), Vol. 1–10, American Scientific Publisher, CA, 2004.
27. K. M. Kadish and R. S. Ruoff, Eds., "Fullerenes: Chemistry, Physics, and Technology," Wiley, New York, 2000.
28. E. Osawa, Ed., "Perspectives of Fullerene Nanotechnology," Kluwer Academic, New York, 2002.
29. T. Akasaka and S. Nagase, Eds., "Endofullerenes: A New Family of Carbon Clusters," Kluwer, Boston, 2002.
30. E. Nakamura and H. Isobe, *Acc. Chem. Res.* 36, 807 (2003).
31. J. B. Donnet, R. C. Bansal, and M. J. Wang, "Carbon Black," Marcel Dekker, New York, 1993.
32. Y. Saito, T. Yoshikawa, M. Okuda, M. Ohkohchi, Y. Ando, A. Kasuya, and Y. Nishina, *Chem. Phys. Lett.* 209, 72 (1994).
33. D. Ugarte, *Nature (London)* 359, 707 (1992).
34. A. W. P. Fung, Z. H. Wang, K. Lu, M. S. Dresselhaus, and R. W. Pekala, *J. Mat. Res.* 8, 1875 (1993).
35. S. Iijima, *Nature (London)* 354, 56 (1991).
36. T. W. Ebbesen and P. M. Ajayan, *Nature (London)* 358, 220 (1992).
37. S. Iijima and T. Ichihashi, *Nature (London)* 363, 603 (1993).
38. V. P. Dravid, X. Lin, Y. Wang, X. K. Wang, A. Yee, J. B. Ketterson, and R. P. H. Chang, *Science* 259, 1601 (1993).
39. T. W. Ebbesen, H. Hiura, J. Fujita, Y. Ochiai, S. Matsui, and K. Tanigaki, *Chem. Phys. Lett.* 209, 83 (1993).
40. T. W. Ebbesen, *Phys. Today* 49, 26 (1996).
41. T. W. Ebbesen, "Carbon Nanotubes," CRC, Boca Raton, Florida, 1997.
42. M. Endo, S. Iijima, and M. S. Dresselhaus, "Carbon Nanotubes," Pergamon, Oxford, 1996.
43. P. M. Ajayan and T. W. Ebbesen, *Rep. Prog. Phys.* 60, 1025 (1997).
44. 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. Tomanek, J. E. Fischer, and R. E. Smalley, *Science* 373, 483 (1996).
45. C. Journet, W. K. Maser, P. Bernier, A. Loiseau, M. L. Delachapelle, S. Lefrant, P. Deniard, R. Lee, and J. E. Fischer, *Nature (London)* 388, 756 (1997).
46. Y. Zhang and S. Iijima, *Philos. Mag. Lett.* 78, 139 (1998).
47. M. Yudasaka, T. Ichihashi, T. Komatsu, and S. Iijima, *Chem. Phys. Lett.* 299, 91 (1999).
48. F. Kokai, K. Takahashi, M. Yudasaka, R. Yamada, T. Ichihashi, and S. Iijima, *J. Phys. Chem.* 103, 4346 (1999).
49. B. C. Satiskumar, A. Govindaraj, R. Sen, and C. N. R. Rao, *Chem. Phys. Lett.* 293, 47 (1998).
50. J. M. Mao, L. F. Sun, L. X. Qian, Z. W. Pan, B. H. Chang, W. Y. Zhou, G. Wang, and S. Xie, *Appl. Phys. Lett.* 72, 3297 (1998).
51. S. Bandow, S. Asaka, Y. Saito, A. M. Rao, L. Grigorian, E. Richter, and P. C. Eklund, *Phys. Rev. Lett.* 80, 3779 (1998).
52. J. Kong, A. M. Cassell, and H. Dai, *Chem. Phys. Lett.* 292, 567 (1998).
53. H. Dai, A. G. Rinzler, P. Nikolaev, A. Thess, D. T. Colbert, and R. E. Smalley, *Chem. Phys. Lett.* 260, 471 (1996).
54. M. Yydasaka, T. Komatsu, I. Ichihashi, and S. Iijima, *Chem. Phys. Lett.* 278, 102 (1998).
55. M. Yydasaka, T. Komatsu, I. Ichihashi, Y. Achiba, and S. Iijima, *J. Phys. Chem. B* 102, 4892 (1998).
56. 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).
57. M. S. Dresselhaus, G. Dresselhaus, and R. Saito, *Phys. Rev. B* 45, 6234 (1992).
58. M. S. Dresselhaus, G. Dresselhaus, and R. Saito, *Carbon* 33, 883 (1995).
59. R. Saito, M. Fujita, G. Dresselhaus, and M. S. Dresselhaus, *Appl. Phys. Lett.* 60, 2204 (1992).
60. R. A. Jishi, M. S. Dresselhaus, and G. Dresselhaus, *Phys. Rev. B* 47, 16671 (1993).
61. R. A. Jishi, L. Venkataraman, M. S. Dresselhaus, and G. Dresselhaus, *Phys. Rev. B* 51, 11176 (1995).
62. J. Li, C. Papadopoulos, and J. Xu, *Nature (London)* 402, 253 (1999).
63. T. W. Ebbesen, "Carbon Nanotubes: Preparation and Properties," CRC, New York, 1996.
64. R. Saito, G. Dresslehaus, and M. S. Dresselhaus, "Physical Properties of Carbon Nanotubes," Imperial College, London, 1998.
65. K. Tanaka, T. Yamabee, K. Fukui, T. Yamabe, and K. Fukui, "The Science and Technologies of Carbon Nanotubes," Elsevier Science, New York, 1999.
66. K. C. Venema, "Electronic Structure of Carbon Nanotubes," Delft University Press, Delft, 2000.
67. L. P. Biro, C. A. Bernardo, and G. G. Tibbetts, "Carbon Filaments and Nanotubes: Common Origins, Differing Applications," Kluwer Academic, New York, 2001.
68. M. S. Dresselhaus, G. Dresselhaus, and P. Avouris, Eds., "Carbon Nanotubes: Synthesis, Structures, Properties, and Applications," Springer-Verlag, New York, 2001.
69. P. J. F. Harris, "Carbon Nanotubes and Related Structures," Cambridge University Press, New York, 2001.
70. S. Reich, J. Janina, and C. Thomsen, "Carbon Nanotubes," Wiley, New York, 2002.
71. J. P. Lu and J. Han, *Inter. J. High Electronics and System* 9, 101 (1998).
72. L. Dai, *Polym. Adv. Technol.* 10, 357 (1999).
73. P. M. Ajayan, *Chem. Rev.* 99, 1787 (1999).
74. C. N. R. Rao, B. C. Satishkumar, A. Govindaraj, and M. Nath, *Chem. Phys. Chem* 2, 78 (2001).
75. S. B. Sinnott and R. Andrews, *Critical Reviews in Solid State and Materials Science* 26, 145 (2001).

76. P. Avouris, *Chem. Phys.* 281, 429 (2002).
77. R. H. Baughman, A. A. Zakhidov, and W. A. de Heer, *Science* 297, 787 (2002).
78. Y. Saito and T. Matsumoto, *Nature (London)* 392, 237 (1998).
79. R. Martel, H. R. Shea, and P. Avouris, *J. Phys. B* 103, 7551 (1999).
80. M. Ahlskog, E. Seynaeve, R. J. M. Vallers, C. Van Haesendonck, A. Fonseca, K. Hernadi, and J. B. Nagy, *Chem. Phys. Lett.* 300, 202 (1999).
81. A. Krishnan, E. Dujardin, M. M. J. Treacy, J. Hugdahl, S. Lynum, and T. W. Ebbesen, *Nature (London)* 388, 451 (1997).
82. S. Iijima, M. Yudasaka, R. Yamada, S. Bandow, K. Suenaga, F. Kokai, and K. Takahashi, *Chem. Phys. Lett.* 309, 165 (1999).
83. R. L. Jocobsen and M. Monthioux, *Nature (London)* 385, 211 (1997).
84. P. M. Ajayan, J. M. Nugent, R. W. Siegel, B. Wei, and P. Kohler-Redlich, *Nature (London)* 404, 243 (2000).
85. M. Bianchetti, P. F. Buonsante, F. Ginelli, H. E. Roman, R. A. Broglia, and F. Alasia, *Phys. Rep.* 357, 459 (2002).
86. P. A. Franken, A. E. Hill, C. W. Peters, and G. Weinreich, *Phys. Rev. Lett.* 7, 118 (1961).
87. G. C. Baldwin, "An Introduction to Nonlinear Optics," Plenum, New York, 1969.
88. F. Zernike and J. E. Midwinter, "Applied Nonlinear Optics," Wiley, New York, 1973.
89. Y. R. Shen, "The Principles of Nonlinear Optics," Wiley, New York, 1984.
90. M. Schubert and B. Wilhelmi, "Nonlinear Optics and Quantum Electronics," Wiley, New York, 1986.
91. A. Yariv, "Quantum Electronics," Wiley, New York 1989.
92. J. Messier, F. Kajzar, P. N. Prasad, and D. Ulrich, "Nonlinear Optical Effects in Organic Polymers," Kluwer Academic, Dordrecht, 1991.
93. J. Messier, F. Kajzar, and P. N. Prasad, "Organic Molecules for Nonlinear Optics and Photonics," Kluwer Academic, Dordrecht, 1991.
94. P. N. Prasad and D. J. Williams, "Introduction to Nonlinear Optical Effects in Molecules and Polymers," Wiley, New York, 1991.
95. R. W. Boyd, "Nonlinear Optics," Academic, New York, 1992.
96. J. I. Sakai, "Phase Conjugate Optics," McGraw-Hill, Singapore, 1992.
97. H. S. Nalwa, *Adv. Mater.* 5, 341 (1993).
98. J. Zyss, "Molecular Nonlinear Optics," Academic, New York, 1994.
99. H. S. Nalwa, Nonlinear Optical Properties of π-conjugated Materials, in "Handbook of Organic Conductive Molecules and Polymers" (H. S. Nalwa, Ed.), Chap. 6, Vol.4, pp. 261–363, Wiley, Chichester, 1997.
100. H. S. Nalwa, Organic Materials for Third-order Nonlinear Optics, in "Nonlinear Optics of Organic Molecules and Polymers" (H. S. Nalwa and S. Miyata, Eds.), Chap. 11, pp. 611–797, CRC, Boca Raton, FL, 1997.
101. U. Woggon, "Optical Properties of Semiconductor Quantum Dots," Springer-Verlag, Berlin, 1997.
102. J. F. Nye, "The Physical Properties of Crystals," Clarendon, Oxford, 1985.
103. A. Willetts, J. E. Rice, and D. M. Burland, *J. Chem. Phys.* 97, 7590 (1992).
104. D. A. McQuarrie, "Quantum Chemistry," University Science Books, Mill Valley, CA, 1983.
105. D. B. Cook, "Handbook of Computational Quantum Chemistry," Oxford University Press, New York, 1998.
106. G. E. Scuseria, *Science* 271, 942 (1996).
107. P. R. Taylor, E. Bylaska, J. H. Weare, and R. Kawai, *Chem. Phys. Lett.* 235, 558 (1995).
108. G. Scuseria, *Chem. Phys. Lett.* 176, 423 (1991).
109. M. Häser, J. Almlöf, and G. E. Scuseria, *Chem. Phys. Lett.* 181, 497 (1991).
110. D. Bakowies, M. Bühl, and W. Thiel, *Chem. Phys. Lett.* 247, 491 (1995).
111. J. Cioslowski, N. Rao, and D. Moncrieff, *J. Am. Chem. Soc.* 124, 8485 (2002).
112. G. E. Scuseria, *Chem. Phys. Lett.* 243, 193 (1995).
113. D. L. Strout and G. E. Scuseria, *J. Chem. Phys.* 102, 8448 (1995).
114. R. G. Parr and W. T. Yang, "Density Functional Theory of Atoms and Molecules," Oxford University Press, New York, 1989.
115. P. Hohenberg and W. Kohn, *Phys. Rev.* 136, B864 (1964).
116. W. Kohn and L. J. Sham, *Phys. Rev.* 140, A1133 (1965).
117. A. D. Becke, *Phys. Rev. A* 38, 3098 (1988).
118. J. P. Perdew and Y. Wang, *Phys. Rev. B* 45, 13244 (1992).
119. S. H. Vosko, L. Wilk, and M. Nusair, *Can. J. Phys.* 58, 1200 (1980).
120. C. Lee, W. Yang, and R. G. Parr, *Phys. Rev. B* 37, 785 (1988).
121. A. D. Becke, *J. Chem. Phys.* 98, 5648 (1993).
122. J. C. Slater, *Phys. Rev.* 81, 385 (1951).
123. R. Bauernschmitt, R. Ahlrichs, F. H. Hennrich, and M. M. Kappes, *J. Am. Chem. Soc.* 120, 5052 (1998).
124. I. Vasiliev, S. Ogüt, and J. R. Chelikowsky, *Phys. Rev. B* 65, 115416 (2002).
125. M. Benz, M. Fanti, P. Fowler, D. Fuchs, M. Kappes, C. Lehner, R. Michel, G. Orlandi, and F. Zerbetto, *J. Phys. Chem.* 100, 13399 (1996).
126. R. Bendale and M. Zerner, *J. Phys. Chem.* 99, 13830 (1995).
127. E. Runge and E. K. U. Gross, *Phys. Rev. Lett.* 52, 997 (1984).
128. R. Bauernschmitt and R. Ahlrichs, *Chem. Phys. Lett.* 256, 454 (1996).
129. M. E. Casida, C. Jamorski, K. C. Casida, and D. R. Salahub, *J. Chem. Phys.* 108, 4439 (1998).
130. R. E. Stratmann, G. E. Scuseria, and M. J. Frisch, *J. Chem. Phys.* 109, 8218 (1998).
131. S. J. A. van Gisbergen, J. G. Snijders, and E. J. Baerends, *J. Chem. Phys.* 103, 9347 (1995).

132. C. Jamorski, M. E. Casida, and D. R. Salahub, *J. Chem. Phys.* 104, 5134 (1996).
133. M. Petersilka, U. J. Gossmann, and E. K. U. Gross, *Phys. Rev. Lett.* 76, 1212 (1996).
134. H. Appel, E. K. U. Gross, and K. Burke, *Phys. Rev. Lett.* 90, 043005 (2003).
135. M. E. Casida, in "Recent Advances in Density Functional Methods" (D. P. Chong, Ed.), Vol. 1, World Scientific, Singapore, 1995.
136. M. E. Casida, in "Recent Developments and Applications of Modern Density Functional Theory, Theoretical and Computational Chemistry" (J. M. Seminario, Ed.), Vol. 4, Elsevier, Amsterdam, 1996.
137. T. L. Beck, *Rev. Mod. Phys.* 72, 1041 (2000).
138. F. Calvayrac, P. G. Reinhard, E. Suraud, and C. A. Ullrich, *Phys. Rep.* 337, 493 (2000).
139. G. T. Velde, F. M. Bickelhaupt, E. J. Baerends, C. F. Guerra, S. J. A. Van Gisbergen, J. G. Snijders, and T. Ziegler, *J. Comput. Chem.* 22, 931 (2001).
140. G. Onida, L. Reining, and A. Rubio, *Rev. Mod. Phys.* 74, 601 (2002).
141. F. Furche, *J. Chem. Phys.* 114, 5982 (2001).
142. P. Salek, O. Vahtras, T. Helgaker, and H. Agren, *J. Chem. Phys.* 117, 9630 (2002).
143. S. J. A. van Gisbergen, J. G. Snijders, and E. J. Baerends, *J. Chem. Phys.* 109, 10644 (1998).
144. S. Tretiak and V. Chernyak, *J. Chem. Phys.* 119, 8809 (2003).
145. C. Y. Yam, S. Yokojima, and G. Chen, *J. Chem. Phys.* 119, 8794 (2003).
146. T. A. Niehaus, S. Suhai, F. D. Sala, P. Lugli, M. Elstner, G. Seifert, and T. Frauenheim, *Phys. Rev. B* 63, 085108 (2001).
147. M. Elstner, D. Porezag, G. Jungnickel, J. Elsner, M. Haugk, T. Frauenheim, S. Suhai, and G. Seifert, *Phys. Rev. B* 58, 7260 (1998).
148. M. Elstner, D. Porezag, G. Jungnickel, T. Frauenheim, S. Suhai, and G. Seifert, in "Tight Binding Approach to Computational Materials Science" (P. Turchi, A. Gonis, and A. Colombo, Eds.), MRS Symp. Proc. No. 491, p. 131. Materials Research Society, Pittsburgh, 1998.
149. R. H. Xie, G. W. Bryant, G. Sun, M. C. Nicklaus, D. Heringer, T. Frauenheim, M. R. Manaa, V. H. Smith Jr., Y. Araki, and O. Ito, *J. Chem. Phys.*, submitted.
150. R. H. Xie, G. W. Bryant, G. Sun, M. C. Nicklaus, D. Heringer, T. Frauenheim, M. R. Manaa, V. H. Smith Jr., Y. Araki, N. Tagmatrachis, H. Shinohara, and O. Ito, *J. Chem. Phys.*, submitted.
151. M. W. Feyereisen, J. Nichols, J. Oddershede, and J. Simons, *J. Chem. Phys.* 96, 2978 (1992).
152. H. Weiss, R. Ahlrichs, and M. Häser, *J. Chem. Phys.* 99, 1262 (1993).
153. J. C. Slater, *Phys. Rev.* 36, 57 (1930).
154. S. F. Boys, *Proc. R. Soc. London, Series A* 200, 542 (1950).
155. S. F. Boys, G. B. Cook, C. M. Reeves, and I. Shavitt, *Nature (London)* 178, 1207 (1956).
156. W. J. Hehre, R. F. Stewart, and J. A. Pople, *J. Chem. Phys.* 51, 2657 (1969).
157. J. S. Binkley and J. A. Pople, *J. Am. Chem. Soc.* 102, 939 (1980).
158. W. J. Hehre, R. Ditchfield, and J. A. Pople, *J. Chem. Phys.* 56, 2257 (1972).
159. R. Krishnan, J. S. Binkley, R. Seeger, and J. A. Pople, *J. Chem. Phys.* 72, 650 (1980).
160. M. J. Frisch, J. A. Pople, and J. S. Binkley, *J. Chem. Phys.* 80, 3265 (1984).
161. R. Pariser and R. G. Parr, *J. Chem. Phys.* 21, 466 (1953).
162. R. Pariser and R. G. Parr, *J. Chem. Phys.* 21, 767 (1953).
163. J. A. Pople, *Trans. Faraday Soc.* 49, 1375 (1953).
164. J. Hubbard, *Proc. Roy. Soc. London A* 276, 238 (1963).
165. R. Hoffmann, *J. Chem. Phys.* 39, 1397 (1963).
166. M. Wolfsberg and L. Helmoholz, *J. Chem. Phys.* 20, 837 (1952).
167. R. H. Xie, G. W. Bryant, S. Lee, and W. Jaskolski, *Phys. Rev. B* 65, 235306 (2002).
168. C. H. Xu, C. Z. Wang, C. T. Chan, and K. M. Ho, *J. Phys. Condens. Matter* 4, 6047 (1992).
169. A. Warshel and M. Karplus, *J. Am. Chem. Soc.* 94, 5612 (1972).
170. J. F. Ward, *Rev. Mod. Phys.* 37, 1 (1965).
171. H. A. Kurtz, J. J. P. Stewart, and K. M. Dieter, *J. Comp. Chem.* 11, 82 (1990).
172. R. J. Bartlett and G. D. Purvis, *Phys. Rev. A* 20, 1313 (1979).
173. G. R. J. Williams, *J. Mol. Struct. (THEOCHEM)* 151, 215 (1987).
174. T. Hara, S. Narita, S. Kumei, and T. Shibuya, *Int. J. Quantum Chem.* 85, 136 (2001).
175. T. Hara, S. Narita, and T. Shibuya, *Fullerene Sci. Technol.* 3, 459 (1995).
176. W. J. Blau, H. J. Bryne, D. J. Cardin, T. J. Dennis, J. P. Hare, H. W. Kroto, R. Taylor, and D. R. M. Walton, *Phys. Rev. Lett.* 67, 1423 (1991).
177. G. Senatore and K. R. Subbaswamy, *Phys. Rev. A* 35, 4532 (1987).
178. Z. H. Levine, *Phys. Rev. B* 49, 4532 (1994).
179. A. D. Corso, F. Mauri, and A. Rubio, *Phys. Rev. B* 53, 15638 (1996).
180. S. J. A. van Gisbergen, J. G. Snijders, and E. J. Baerends, *J. Chem. Phys.* 109, 10644 (1998); 109, 10657 (1998); (erratum) 111, 6652 (1999).
181. G. Nicolas and P. Durand, *J. Chem. Phys.* 70, 2020 (1979).
182. J. M. Andre, L. A. Burke, J. Delhalle, G. Nicolas, and P. Durand, *Int. J. Quantum Chem. Symp.* 13, 283 (1979).
183. J. L. Bredas, in "Handbook of Conducting Polymers" (T. A. Skotheim, Ed.), Vol. 2, p. 859. Marcel-Dekker, New York, 1986.
184. J. L. Bredas, R. R. Chance, R. Silbey, G. Nicolas, and P. Durant, *J. Chem. Phys.* 75, 255 (1981).
185. J. L. Bredas, B. Themans, and J. M. Andre, *J. Chem. Phys.* 78, 6137 (1983).
186. L. Silberstein, *Philos. Mag.* 33, 92 (1917).

187. L. Silberstein, *Philos. Mag.* 33, 521 (1917).
188. J. Applequist, *Acc. Chem. Res.* 10, 79 (1977).
189. L. Jensen, P. O. Astrand, K. O. Sylvester-Hvid, and K. V. Mikkelsen, *J. Phys. Chem. A* 104, 1563 (2000).
190. L. Jensen, K. O. Sylvester-Hvid, P. O. Astrand, and K. V. Mikkelsen, *J. Phys. Chem. A* 107, 2270 (2003).
191. M. L. Olson and K. R. Sundberg, *J. Chem. Phys.* 69, 5400 (1978).
192. L. Jensen, P. O. Astrand, and K. V. Mikkelsen, *Int. J. Quantum Chem.* 84, 513 (2001).
193. B. Shanker and J. Applequist, *J. Phys. Chem.* 98, 6486 (1994).
194. G. R. Meredith, L. T. Cheng, H. Hsiung, H. A. Vanherzeele, and F. C. Zumsteg, in "Materials for Nonlinear and Electro-Optics" (M. H. Lyons, Ed.), pp. 139–150, IOP, New York, 1989.
195. H. Hoshi, N. Nakamura, Y. Maruyama, T. Nakagawa, S. Suzuki, H. Shiromaru, and Y. Achiba, *Jpn. J. Appl. Phys.* 30, L1397 (1991).
196. X. K. Wang, T. G. Zhang, W. P. Lin, S. Z. Liu, G. K. Wong, M. M. Kappes, R. P. H. Chang, and J. B. Ketterson, *Appl. Phys. Lett.* 60, 810 (1992).
197. B. Koopmans, A. Anema, H. T. Jonkman, G. A. Sawatzky, and F. van der Woude, *Phys. Rev. B* 48, 2759 (1993).
198. F. Kajzar, C. Taliani, R. Zamboni, S. Rossini, and R. Danieli, *Synthetic Metals* 77, 257 (1996).
199. K. Kuhnke, M. Epple, and K. Kern, *Chem. Phys. Lett.* 294, 241 (1998).
200. Y. R. Shen, *Ann. Rev. Phys. Chem.* 40, 327 (1989).
201. D. Wilk, D. Johannsmann, C. Stanners, and Y. R. Shen, *Phys. Rev. B* 51, 10057 (1995).
202. A. M. Janner, Ph.D. Thesis, University of Groningen, Groningen, 1998.
203. G. Wang, J. Wen, Q. Houng, S. Qian, and X. Lu, *J. Phys. D: Appl. Phys.* 32, 84 (1999).
204. T. Rasing, G. Berkovic, Y. R. Shen, S. G. Grubb, and M. W. Kim, *Chem. Phys. Lett.* 130, 1 (1986).
205. P. A. Heiney, J. E. Fischer, A. R. McGhie, W. J. Romanow, A. M. Denenstein, J. P. McCauley Jr., A. B. Smith III, and D. E. Cox, *Phys. Rev. Lett.* 66, 2911 (1991).
206. W. I. F. David, R. M. Ibberson, T. J. S. Dennis, J. P. Hare, and K. Prassides, *Europhys. Lett.* 18, 219 (1992).
207. B. Koopmans, A. M. Janner, H. T. Jonkman, G. A. Sawatzky, and F. van der Woude, *Phys. Rev. Lett.* 71, 3569 (1993).
208. R. W. Munn, Z. Shuai, and J.-L. Bredas, *J. Chem. Phys.* 108, 5975 (1998).
209. Y. Liu, H. Jiang, W. Wang, Y. Li, and J. Zheng, *Phys. Rev. B* 50, 4940 (1994).
210. A. Sassara, G. Zerza, and M. Chergui, *J. Phys. Chem. A* 102, 3072 (1998).
211. S. Qin, W. M. You, and Z. B. Su, *Phys. Rev. B* 48, 17562 (1993).
212. A. Lucas, G. Gensterblum, J. J. Pireaux, P. A. Thiry, R. Caudano, J. P. Vigneron, and W. Krätschmer, *Phys. Rev. B* 45, 13694 (1992).
213. Z. Shuai and J. L. Bredas, *Mol. Cryst. Liq. Cryst.* 256, 801 (1994).
214. F. Kajzar, C. Taliani, R. Danieli, S. Rossini, and R. Zamboni, *Phys. Rev. Lett.* 73, 1617 (1994).
215. B. F. Levine and C. G. Bethea, *J. Chem. Phys.* 66, 1070 (1977).
216. J. L. Oudar, *J. Chem. Phys.* 67, 446 (1977).
217. R. E. Haufler, J. Conceicao, 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, L. J. Wilson, R. Curl, and R. E. Smalley, *J. Phys. Chem.* 94, 8634 (1990).
218. Y. Wang and L. T. Cheng, *J. Phys. Chem.* 96, 1530 (1992).
219. A. Weller, in "Exciplex" (M. Gordon and W. R. Ware, Eds.), Academic, New York, 1975.
220. G. R. Meredith, L. T. Cheng, H. Hsiung, H. A. Vanherzeele, and F. C. Zumsteg, in "Materials for Nonlinear and Electro-optics" (M. H. Lyons, Ed.), pp. 139–150, IOP, New York, 1989.
221. L. T. Cheng, W. Tam, S. H. Stevenson, G. R. Meredith, G. Rikken, and S. R. Marder, *J. Phys. Chem.* 95, 10631 (1991).
222. M. Lamrani, R. Hamasaki, M. Mitsuishi, T. Miyashita, and Y. Yamamoto, *Chem. Comm.* 1595 (2000).
223. K. D. Kampe, N. Egger, and M. Vogel, *Angew. Chem., Int. Ed. Engl.* 32, 1174 (1993).
224. L. B. Gan, D. J. Zhou, C. P. Luo, C. H. Huang, T. K. Li, J. Bai, X. S. Zhao, and X. H. Xia, *J. Phys. Chem.* 98, 12459 (1994).
225. F. Kajzar, Y. Okada-Shudo, C. Meritt, and Z. Kafafi, *Synthetic Metals* 94, 91 (1998).
226. N. Matsuzawa, D. A. Dixon, and T. Fukunaga, *J. Phys. Chem.* 96, 7584 (1992).
227. J. Jiang, J. Dong, Q. Xu, and D. Y. Xing, *Z. Phys. D* 37, 341 (1996).
228. P. S. Baran, R. R. Monaco, A. U. Khan, D. I. Schuster, and S. R. Wilson, *J. Am. Chem. Soc.* 119, 8363 (1997).
229. S. Wang, R. M. Leblanc, F. Arias, and L. Echegoyen, *Langmuir* 13, 1672 (1997).
230. Z. X. Guo, Y. L. Li, F. L. Bai, J. M. Yan, Z. X. Ge, D. B. Zhu, J. H. Si, P. X. Ye, L. X. Wang, and T. K. Li, *J. Phys. Chem. Solids* 61, 1089 (2000).
231. Z. X. Guo, Y. L. Li, Z. X. Ge, J. Yan, D. B. Zhu, J. H. Shi, and P. X. Ye, *Appl. Phys. B* 71, 545 (2000).
232. E. E. B. Campbell, M. Fanti, I. V. Hertel, R. Mitzner, and F. Zerbetto, *Chem. Phys. Lett.* 288, 131 (1998).
233. I. Fuks, I. V. Kityk, J. Kasperczyk, J. Berdowski, and I. Schirmer, *Chem. Phys. Lett.* 353, 7 (2002).
234. F. Lin, S. Mao, Z. Meng, H. Zeng, J. Qiu, Y. Yue, and T. Guo, *Appl. Phys. Lett.* 65, 2522 (1994).
235. H. Zeng, Z. Sun, Y. Segawa, F. Lin, S. Mao, Z. Xu, and S. H. Tang, *J. Phys. D: Appl. Phys.* 33, L93 (2000).
236. J. Ridley and M. C. Zerner, *Theor. Chim. Acta (Berlin)* 32, 111 (1973).
237. S. D. Bella, I. L. Fragala, M. A. Ratner, and T. J. Marks, *J. Am. Chem. Soc.* 115, 682 (1993).
238. A. Ulman, C. S. Willand, W. Kohler, D. R. Robello, D. J. Williams, and L. Handley, *J. Am. Chem. Soc.* 112, 7083 (1990).
239. J. Li, J. Feng, and C. Sun, *J. Phys. Chem.* 98, 8636 (1994).
240. Y. Liu, D. Zhang, H. Hu, and C. Liu, *J. Mol. Struct. (Theochem)* 545, 97 (2001).

241. W. Fu, J.-K. Feng, G.-B. Pan, and X. Zhang, *Theor. Chem. Acc.* 106, 241 (2001).
242. K. C. Rustagi, L. Ramaniah, and S. V. Nair, *Int. J. Mod. Phys. B.* 6, 3941 (1992).
243. I. M. L. Billas, C. Massobrio, M. Boero, M. Parrinello, W. Branz, F. Tast, N. Malinowski, M. Heinebrodt, and T. P. Martiin, *J. Chem. Phys.* 111, 15 (1999).
244. L. Jensen, P. T. van Duijnen, J. G. Snijders, and D. P. Chong, *Chem. Phys. Lett.* 359, 524 (2002).
245. M. Fanti, G. Orlandi, and F. Zerbetto, *J. Phys. Chem. A.* 101, 3015 (1997).
246. J. Zyss and I. Ledoux, *Chem. Rev.* 94, 77 (1994).
247. J. Zyss, S. Brasselet, V. R. Thalladi, and G. R. Desiraju, *J. Chem. Phys.* 109, 658 (1998).
248. A. G. H. Barbosa and M. A. C. Nascimento, *Chem. Phys. Lett.* 343, 15 (2001).
249. C. Piskoti, J. Yarger, and A. Zettl, *Nature (London)* 393, 771 (1998).
250. J. R. Heath, *Nature (London)* 393, 730 (1998).
251. M. Cote, J. C. Grossman, M. L. Cohen, and S. G. Louie, *Phys. Rev. Lett.* 81, 697 (1998).
252. H. Kietzmann, R. Rochow, G. Ganteför, W. Eberhardt, K. Vietze, G. Seifert, and P. W. Fowler, *Phys. Rev. Lett.* 81, 5378 (1998).
253. P. W. Fowler, D. Mitchell, and F. Zerbetto, *J. Am. Chem. Soc.* 121, 3218 (1999).
254. P. G. Collins, J. C. Grossman, M. Cote, M. Ishigami, C. Piskoti, S. G. Louie, M. L. Cohen, and A. Zettl, *Phys. Rev. Lett.* 82, 165 (1999).
255. J. C. Grossman, M. Cote, S. G. Louie, and M. L. Cohen, *Chem. Phys. Lett.* 284, 344 (1998).
256. Z. Slanina, X. Zhao, and E. Osawa, *Chem. Phys. Lett.* 290, 311 (1998).
257. E. Halac, E. Burgos, and H. Bonadeo, *Chem Phys. Lett.* 299, 64 (1999).
258. Z. Slanina, F. Uhlik, X. Zhao, and E. Osawa, *J. Chem. Phys.* 113, 4933 (2000).
259. L. Yuan, J. Yang, K. Deng, and Q. Zhu, *J. Phys. Chem. A* 104, 6666 (2000).
260. M. N. Jagadeesh and J. Chandrasekhar, *Chem. Phys. Lett.* 305, 298 (1999).
261. P. W. Fowler, T. Heine, K. M. Rogers, J. P. B. Sandall, G. Seifert, and F. Zerbetto, *Chem. Phys. Lett.* 300, 369 (1999).
262. S. P. Karna, P. N. Prasad, and M. Dupuis, *J. Chem. Phys.* 94, 1171 (1991).
263. M. Sheik-Bahae, A. A. Said, T. H. Wei, D. J. Hagan, and E. W. van Stryland, *IEEE J. Quantum Electron.* 26, 760 (1990).
264. M. O. Levenson and N. Bloembergen, *J. Chem. Phys.* 60, 1323 (1974).
265. A. F. Garito, J. R. Heflin, K. Y. Yong, and O. Zamani-Khamiri, *Proc. Soc. Photon. Opt. Instrum. Eng.* 971, 2 (1988).
266. S. R. Marder, W. E. Torruellas, M. Blanchard-Desce, V. Ricci, G. I. Stegeman, S. Gilmour, J. L. Bredas, J. Li, G. U. Bublitz, and S. G. Boxer, *Science* 276, 1233 (1997).
267. H. Thinpont, G. L. A. Rikken, E. W. Meijer, W. ten Hoeve, and H. Hynberg, *Phys. Rev. Lett.* 65, 2141 (1990).
268. F. C. Spano and Z. G. Soos, *J. Chem. Phys.* 99, 9265 (1993).
269. I. D. W. Samuel, I. Ledoux, C. Dhenaut, J. Zyss, H. H. Fox, R. R. Schrock, and R. J. Silbey, *Science* 265, 1070 (1994).
270. G. Chen, D. Lu, and W. A. Goddard III, *J. Chem. Phys.* 101, 5860 (1994).
271. D. Lu, G. Chen, and W. A. Goddard III, *J. Chem. Phys.* 101, 4920 (1994).
272. D. Lu, G. J. W. Perry, and W. A. Goddard III, *J. Am. Chem. Soc.* 116, 10679 (1994).
273. S. Mukamel, A. Tahahashi, H. X. Wang, and G. Chen, *Science* 266, 250 (1994).
274. D. Lu, B. Marten, M. Ringnalda, R. A. Friesner, and W. A. Goddard III, *Chem. Phys. Lett.* 257, 224 (1996).
275. T. Hamada, *J. Chem. Soc. Faraday Trans.* 92, 3165 (1996).
276. M. Hasan, S. J. Kim, J. L. Toto, and B. Kirtman, *J. Chem. Phys.* 105, 186 (1996).
277. E. Sohmen, J. Fink, and W. Kratschmer, *Z. Phys. B* 86, 86 (1992).
278. W. J. Blau, H. J. Byrne, D. J. Cardin, T. J. Dennis, J. P. Hare, H. W. Kroto, R. Taylor, and D. R. M. Walton, *Phys. Rev. Lett.* 67, 1423 (1991).
279. R. J. Knize and J. P. Partanen, *Phys. Rev. Lett.* 68, 2704 (1992).
280. Z. H. Kafafi, F. J. Bartoli, J. R. Lindle, and R. G. S. Pong, *Phys. Rev. Lett.* 68, 2705 (1992).
281. F. Kajzar, C. Taliani, R. Zamboni, S. Rossini, and R. Danieli, in "Fullerenes: Status and Perspectives: First Italian Workshop on Fullerenes" (C. Taliani, G. Ruani, and R. Zamboni, Eds.). World Scientific, Singapore, 1992.
282. F. Kajzar, C. Taliani, R. Zamboni, S. Rossini, and R. Danieli, *Synth. Met.* 54, 21 (1993).
283. F. Kajzar, C. Taliani, R. Zamboni, S. Rossini, and R. Zamboni, *Chem. Phys. Lett.* 217, 418 (1994).
284. Q. Gong, Y. Sun, Z. Xia, Y. H. Zou, Z. Gu, X. Zhou, and D. Qiang, *J. Appl. Phys.* 71, 3025 (1992).
285. Z. H. Kafafi, J. R. Lindle, R. G. S. Pong, F. J. Bartoli, L. J. Lingg, and J. Milliken, *Chem. Phys. Lett.* 188, 492 (1992).
286. G. B. Talapatra, N. Manickam, M. Samoc, M. E. Orczyk, S. P. Karna, and P. N. Prasad, *J. Phys. Chem.* 96, 5206 (1992).
287. J. S. Meth, H. Vanherzeele, and Y. Wang, *Chem. Phys. Lett.* 197, 26 (1992).
288. Z. Zhang, D. Wang, P. Ye, Y. Li, P. Wu, and D. Zhu, *Opt. Lett.* 17, 973 (1992).
289. D. Neher, G. I. Stegeman, F. A. Tinker, and N. Peyghambarian, *Opt. Lett.* 17, 1491 (1992).
290. S. C. Yang, Q. Gong, Z. Xia, Y. H. Zou, Y. Q. Wu, D. Qiang, Y. L. Sun, and Z. N. Gu, *Appl. Phys. B* 52, 51 (1992).
291. S. R. Flom, R. G. S. Pong, F. J. Bartoli, and Z. H. Kafafi, *Phys. Rev. B* 46, 15598 (1992).
292. J. Li, J. Feng, and J. Sun, *Chem. Phys. Lett.* 203, 560 (1993).
293. J. Li, J. Feng, and C. Sun, *Int. J. Quantum. Chem.* 52, 673 (1994).

294. N. Matsuzawa and D. A. Dixon, *J. Phys. Chem.* 96, 6241 (1992).
295. M. J. Rosker, H. O. Marcy, T. Y. Chang, J. T. Khoury, K. Hansen, and R. L. Whetten, *Chem. Phys. Lett.* 196, 427 (1992).
296. D. Jonsson, P. Norman, K. Ruud, H. Ågren, and T. Helgaker, *J. Chem. Phys.* 109, 572 (1998).
297. J. R. Lindle, R. G. S. Pong, F. J. Bartoli, and Z. H. Kafafi, *Phys. Rev. B* 48, 9447 (1993).
298. N. Tang, J. P. Partanen, R. W. Hellwarth, and R. J. Knize, *Phys. Rev. B* 48, 8404 (1993).
299. W. Ji, S. H. Tang, G. Q. Xu, H. S. O. Chan, S. H. Ng, and W. W. Ng, *J. Appl. Phys.* 74, 3669 (1993).
300. F. Sun, S. Zhang, Z. Zia, Y. H. Zou, X. Chen, D. Qiang, X. Zhou, and Y. Wu, *Phys. Rev. B* 51, 4614 (1995).
301. P. Norman, Y. Lu, D. Jonsson, and H. Ågren, *J. Chem. Phys.* 106, 8788 (1997).
302. L. Geng and J. C. Wright, *Chem. Phys. Lett.* 249, 105 (1996).
303. J. Li, S. Wang, H. Yang, Q. Gong, X. An, H. Chen, and D. Qiang, *Chem. Phys. Lett.* 288, 175 (1998).
304. H. Huang, G. Gu, S. Yang, J. Fu, P. Yu, G. K. L. Wong, and Y. Du, *Chem. Phys. Lett.* 272, 427 (1997).
305. H. Huang, G. Gu, S. Yang, J. Fu, P. Yu, G. K. L. Wong, and Y. Du, *J. Phys. Chem.* 102, 61 (1998).
306. R. J. Knize, *Opt. Commun.* 106, 95 (1994).
307. R. H. Xie and J. Jiang, *Appl. Phys. Lett.* 71, 1029 (1997).
308. R. H. Xie and J. Jiang, *Chem. Phys. Lett.* 280, 66 (1997).
309. R. H. Xie, *Nuovo Cimento D* 19, 1867 (1997).
310. Y. Luo, *Chem. Phys. Lett.* 289, 350 (1998).
311. C. E. Moore, B. H. Cardelino, and X. Q. Wang, *J. Phys. Chem.* 100, 4685 (1996).
312. Z. Shuai and J. L. Bredas, *Phys. Rev. B* 46, 16135 (1992).
313. Z. Shuai and J. L. Bredas, *Phys. Rev. B* 48, 11520 (1993).
314. M. Fanti, G. Orlandi, and F. Zerbetto, *J. Am. Chem. Soc.* 117, 6101 (1995).
315. M. Fanti, P. W. Fowler, G. Orlandi, and F. Zerbetto, *J. Chem. Phys.* 107, 5072 (1997).
316. R. H. Xie and Q. Rao, *Appl. Phys. Lett.* 72, 2358 (1998).
317. R. H. Xie, *J. Chem. Phys.* 108, 3626 (1998).
318. K. Harigaya, *J. Phys.: Condens. Matter.* 10, 6845 (1998).
319. K. Harigaya, *J. Lumin.* 76, 652 (1998).
320. K. Harigaya, *Synthetic Metals* 91, 379 (1997).
321. K. Harigaya, *Japan J. Appl. Phys.* 36, L485 (1997).
322. C. E. Moore, B. H. Cardelino, D. O. Frazier, J. Niles, and X. Q. Wang, *Theochem J. Mol. Struct.* 454, 135 (1998).
323. Z. Shuai and J. L. Bredas, *Phys. Rev. B* 46, 4395 (1992).
324. X. G. Wan, J. M. Dong, and D. Y. Xing, *J. Phys. B* 30, 1323 (1997).
325. R. H. Xie, *Z. Naturforsch. A* 54, 348 (1999).
326. Y. T. Lin, J. H. Sheu, and S. L. Lee, *Chem. Phys. Lett.* 345, 228 (2001).
327. Y.-T. Lin, R. K. Mishra, and S.-L. Lee, *Mol. Phys.* 97, 987 (1999).
328. K. C. Rustagi, S. V. Nair, and L. M. Ramaniah, *Progress in Crystal Growth & Characterization of Materials* 34, 81 (1997).
329. V. P. Belousov, I. M. Belousova, V. P. Budtov, V. V. Danilov, O. B. Danilov, A. G. Kalintsev, and A. A. Mak, *J. Opt. Technol.* 64, 1081 (1997).
330. S. J. A. van Gisbergen, J. G. Snijders, and E. J. Baerends, *Phys. Rev. Lett.* 78, 3097 (1997).
331. J.-I. Iwata, K. Yabana, and G. F. Bertsch, *J. Chem. Phys.* 115, 8773 (2001).
332. Y. Luo, P. Norman, P. Macak, and H. Ågren, *Phys. Rev. B* 61, 3060 (2000).
333. E. A. Perpète, B. Champagne, and B. Kirtman, *Phys. Rev. B* 61, 13137 (2000).
334. N. Matsuzawa and D. A. Dixon, *J. Phys. Chem.* 96, 6872 (1992).
335. A. A. Quang and M. R. Pederson, *Phys. Rev. B* 46, 12906 (1992).
336. K. Harigaya and S. Abe, *Jpn. J. Appl. Phys.* 31, L887 (1992).
337. A. Takahashi, H. X. Wang, and S. Mukamel, *Chem. Phys. Lett.* 216, 394 (1993).
338. M. M. Mestechkin and G. E. Whyman, *Chem. Phys. Lett.* 214, 144 (1993).
339. G. E. Whyman and M. M. Mestechkin, *Opt. Commun.* 109, 410 (1994).
340. E. Westin and A. Rosén, *Int. J. Mod. Phys. B* 6, 3893 (1992).
341. P. W. Fowler, H. M. Kelly, M. Malagoli, and R. Zanasi, *Int. J. Mod. Phys. B* 6, 3903 (1992).
342. Y. Wang, G. F. Bertsch, and D. Tomanek, *Z. Physik D* 25, 181 (1993).
343. F. Willaime and L. M. Falicov, *J. Chem. Phys.* 98, 6369 (1993).
344. T. Hara, Y. Nomura, S. Narita, and T.-I. Shibuya, *Chem. Phys. Lett.* 240, 610 (1995).
345. E. Westin and A. Rosén, *Appl. Phys. A* 60, 49 (1995).
346. B. H. Cardelino, C. E. Moore, and D. O. Frazier, *J. Phys. Chem. A* 101, 2207 (1997).
347. Y.-T. Lin and S.-L. Lee, *Int. J. Quant. Chem.* 75, 457 (1999).
348. A. Göller and U.-W. Grummt, *Int. J. Quant. Chem.* 77, 727 (2000).
349. Y. Nomura, T. Miyamoto, T. Hara, S. Narita, and T. I. Shibuya, *J. Chem. Phys.* 112, 6603 (2000).
350. W.-D. Cheng, D.-S. Wu, H. Zhang, and H.-X. Wang, *Phys. Rev. B* 66, 113401 (2002).
351. R. Lascola and J. C. Wright, *Chem. Phys. Lett.* 269, 79 (1997).
352. R. Lascola and J. C. Wright, *Chem. Phys. Lett.* 290, 117 (1998).
353. R. H. Xie and J. Jiang, *J. Appl. Phys.* 83, 3001 (1998).
354. R. H. Xie and Q. Rao, *Chem. Phys. Lett.* 313, 211 (1999).
355. X. Liu, J. Si, B. Chang, G. Xu, Q. Yang, Z. Pan, and S. Xie, *Appl. Phys. Lett.* 74, 164 (1999).
356. S. Xie, W. Li, Z. Pan, B. Chang, and L. Sun, *J. Phys. Chem. Solids* 61, 1153 (2000).

357. Z. Shi, Y. Lian, X. Zhou, Z. Gu, Y. Zhang, S. Iijima, Q. Gong, H. Li, and S. L. Zhang, *Chem. Commun.* 461 (2000).
358. S. Wang, W. Huang, H. Yang, Q. Gong, Z. Shi, X. Zhou, D. Qiang, and Z. Gu, *Chem. Phys. Lett.* 320, 411 (2000).
359. Y. C. Chen, N. R. Raravikar, L. S. Schadler, P. M. Ajayan, Y. P. Zhao, T. M. Lu, G. C. Wang, and X. C. Zhang, *Appl. Phys. Lett.* 81, 975 (2002).
360. X. Wan, J. Dong, and D. Y. Xing, *Solid State Commun.* 107, 791 (1998).
361. V. A. Margulis and T. A. Sizikova, *Physica B* 245, 173 (1998).
362. X. Wan, J. Dong, and D. Y. Xing, *Phys. Rev. B* 58, 6756 (1998).
363. J. Jiang, J. Dong, X. Wan, and D. Y. Xing, *J. Phys. B: At. Mol. Opt. Phys.* 31, 3079 (1998).
364. J. Jiang, J. Dong, and D. Y. Xing, *Phys. Rev. B* 59, 9838 (1999).
365. V. A. Margulis, *J. Phys.: Condens. Matter* 11, 3065 (1999).
366. V. A. Margulis, E. A. Gaiduk, and E. N. Zhidkin, *Opt. Commun.* 183, 317 (2000).
367. J. Orr and J. F. Ward, *Mol. Phys.* 20, 513 (1971).
368. Z. Shuai and J. L. Bredas, *Phys. Rev. B* 44, 5962 (1991).
369. J. Yu, Ph.D. dissertation, University of Houston, Houston, 1991.
370. D. S. Wu, W. D. Cheng, H. Zhang, X. D. Li, Y. Z. Lan, D. G. Chen, Y. J. Gong, and Y. C. Zhang, *Phys. Rev. B* 68, 125402 (2003).
371. L. Jensen, P. O. Astrand, and K. V. Mikkelsen, *Nano Lett.* 3, 661 (2003).
372. S. Botti, R. Ciardi, L. De Dominicis, L. S. Asilyan, R. Fantoni, and T. Marolo, *Chem. Phys. Lett.* 378, 117 (2003).
373. C. Stanciu, R. Ehlich, V. Petrov, O. Steinkellner, J. Herrmann, I. V. Hertel, G. Y. Slepyan, A. A. Khrutchinski, S. A. Maksimenko, F. Rotermund, E. E. B. Campbell, and F. Rohmund, *Appl. Phys. Lett.* 81, 4064 (2002).
374. Y. Nishikawa, A. Tackeuchi, S. Nakwamura, S. Muto, and N. Yokoyama, *Appl. Phys. Lett.* 66, 839 (1995).
375. S. Nakamura, K. Tajima, and Y. Sugimoto, *Appl. Phys. Lett.* 65, 283 (1994).
376. H. Han, S. Vijayalakshmi, A. Lan, Z. Iqbal, H. Grebel, E. Lalanne, and A. M. Johnson, *Appl. Phys. Lett.* 82, 1458 (2003).
377. J. R. Heflin, D. Marciu, C. Figura, S. Wang, P. Burbank, S. Stevenson, H. C. Dorn, and J. C. Withers, in "Fullerene and Photonics III" (Z. H. Kafafi, Ed.), Proc. SPIE: Soc. Photo-Opt. Instrum. Eng., Vol. 2854, p. 162. SPIE-International Society for Optical Engineering, Bellingham, WA, 1996.
378. Q. Gong, P. Yuan, Z. Xia, Y. H. Zou, J. Li, H. Yang, H. Hong, F. Yu, S. Chen, Z. Wu, X. Zhou, and F. Li, *Solid State Commun.* 103, 403 (1997).
379. G. Gu, H. Huang, S. Yang, P. Yu, J. Fu, G. K. Wong, X. Wan, J. Dong, and Y. Du, *Chem. Phys. Lett.* 289, 167 (1998).
380. F. Kajzar, Y. Okada-Shudo, C. Meritt, and Z. Kafafi, *Synth. Met.* 94, 91 (1998).
381. X. Wei, C. Hu, Z. Suo, P. Wang, W. Zhang, Z. Xu, and E. C. Alyea, *Chem. Phys. Lett.* 300, 385 (1999).
382. E. E. B. Campbell, S. Couris, M. Fanti, E. Koudoumas, N. Krawez, and F. Zerbetto, *Adv. Mater.* 11, 405 (1999).
383. T. Zhang, F. Wang, H. Yang, Q. Gong, X. An, H. Chen, and D. Qiang, *Chem. Phys. Lett.* 301, 343 (1999).
384. M. Konstantaki, E. Koudoumas, S. Couris, J. M. Janot, H. Eddaoudi, A. Deratani, P. Seta, and S. Leach, *Chem. Phys. Lett.* 318, 488 (2000).
385. W. Huang, S. Wang, R. Liang, Q. Gong, W. Qiu, Y. Liu, and D. Zhu, *Chem. Phys. Lett.* 324, 354 (2000).
386. Y. Song, L. Jiang, D. Zhu, C. Wang, K. Zhao, Z. Xia, and Y. Zou, *J. Phys. Chem. Solids* 61, 1141 (2000).
387. W. Qian, L. Lin, Z. Xia, Y. Zou, S. Qian, G. Ma, Y. Lin, R. Cai, Y. Chen, and Z. E. Huang, *Chem. Phys. Lett.* 319, 89 (2000).
388. R. Liang, S. Wang, W. Huang, Q. Gong, H. Li, H. Chen, and D. Qiang, *J. Phys. D: Appl. Phys.* 33, 2249 (2000).
389. E. Koudoumas, M. Konstantaki, A. Mavromanolakis, S. Couris, Y. Ederle, C. Mathis, P. Seta, and S. Leach, *Chem. Phys. Lett.* 335, 533 (2001).
390. S. Wang, W. Huang, R. Liang, Q. Gong, H. Li, H. Chen, and D. Qiang, *Phys. Rev. B* 63, 153408 (2001).
391. S. Wang, W. Huang, R. Liang, Q. Gong, H. Li, H. Chen, and D. Qiang, *J. Phys. Chem. B* 105, 10784 (2001).
392. I. Fuks-Janczarek, S. Dabos-Seignan, B. Sahraoui, I. V. Kityk, J. Berdowski, E. Allard, and J. Cousseau, *Opt. Commun.* 211, 303 (2002).
393. W. Zhao, H. Li, R. West, and J. C. Wright, *Chem. Phys. Lett.* 281, 105 (1997).
394. O. B. Mavritsky, A. N. Egorov, A. N. Petrovsky, K. V. Yakubovsky, W. J. Blau, D. N. Weldon, and F. Z. Henary, in "Fullerenes and Photonics III" (Z. H. Kafafi, Ed.), Proc. SPIE, Soc. Photo-Opt. Instrum. Eng., Vol. 2854, p. 254. SPIE-International Society for Optical Engineering, Bellingham, WA, 1996.
395. B. Kippelen, S. R. Marder, E. Hendrickx, J. L. Maldonado, G. Guillemet, B. L. Volodin, D. D. Steele, Y. Enami, Y. J. Yao, J. F. Wang, H. Röckel, L. Erskine, and N. Peyghambarian, *Science* 279, 54 (1998).
396. S. R. Marder, B. Kippelen, A. K. Y. Jen, and N. Peyghambarian, *Nature (London)* 388, 845 (1997).
397. L. R. Dalton, W. H. Steier, B. H. Robinson, C. Zhang, A. Ren, S. Garner, A. Chen, T. Londergan, L. Irwin, B. Carlson, L. Fifield, G. Phelan, C. Kincaid, J. Amend, and A. Jen, *J. Mater. Chem.* 9, 1905 (1999).
398. M. Malaun, R. Kowallick, A. M. McDonagh, M. Marcaccio, R. L. Paul, I. Asselberghs, K. Clays, A. Persoons, B. Bildstein, C. Fiorini, J. M. Nunzi, M. D. Ward, and J. A. McCleverty, *J. Chem. Soc., Dalton Trans.* 2001, 3025 (2001).
399. M. Malaun, Z. R. Reeves, R. L. Paul, J. C. Jeffery, J. A. McCleverty, M. D. Ward, I. Asselberghs, K. Clays, and A. Persoons, *Chem. Commun.* 2001, 49 (2001).
400. I. Asselberghs, Y. Zhao, K. Clays, A. Persoons, A. Comito, and Y. Rubin, *Chem. Phys. Lett.* 364, 279 (2002).
401. J. C. Hummelen, F. Wudl, and C. Bellavia-Lund, *Top. Curr. Chem.* 199, 93 (1999).

402. B. L. Yu, H. P. Xia, C. S. Zhu, and F. X. Gan, *Appl. Phys. Lett.* 81, 2701 (2002).
403. J. Dong, J. Jiang, J. Yu, Z. D. Wang, and D. Y. Xing, *Phys. Rev. B* 52, 9066 (1995).
404. Q. Xu, J. Dong, J. Jiang, and D. Y. Xing, *J. Phys. B: At. Mol. Opt. Phys.* 29, 1563 (1996).
405. J. Jiang, J. Dong, and D. Y. Xing, *Solid State Commun.* 101, 537 (1997).
406. R. H. Xie, *Phys. Lett. A* 258, 51 (1999).
407. R. H. Xie, L. Jensen, G. W. Bryant, J. Zhao, and V. H. Smith Jr., *Chem. Phys. Lett.* 375, 445 (2003).
408. R. H. Xie, G. W. Bryant, L. Jensen, J. Zhao, and V. H. Smith Jr., *J. Chem. Phys.* 118, 8621 (2003).
409. W.-D. Cheng, D.-S. Wu, H. Zhang, D.-G. Chen, and H.-X. Wang, *Phys. Rev. B* 66, 85422 (2002).
410. J. Harris, *Phys. Rev. B* 31, 1770 (1985).
411. N. Kurita, K. Kobayashi, H. Kumahora, K. Tago, and K. Ozawa, *Chem. Phys. Lett.* 198, 95 (1992).
412. T. Guo, C. Jin, and R. E. Smalley, *J. Phys. Chem.* 95, 4948 (1991).
413. Y. Chai, T. Guo, C. Jin, R. E. Haufler, L. P. F. Chibante, J. Fure, L. Wang, J. M. Alford, and R. E. Smalley, *J. Phys. Chem.* 95, 7564 (1991).
414. L. Tutt and T. F. Boggess, *Prog. Quantum Electron.* 17, 299 (1993).
415. E. W. van Stryland, D. J. Hagan, T. Xia, and A. A. Said, in "Nonlinear Optics of Organic Molecules and Polymers" (H. S. Nalwa and S. Miyata, Eds.), p. 841. Chemical Rubber, New York, 1997.
416. T. Xia, D. J. Hagan, A. Dogariu, A. A. Said, and E. W. van Stryland, *Appl. Opt.* 36, 4110 (1997).
417. D. J. Hagan, in "Handbook of Optics, Vol. 4: Fiber Optics and Nonlinear Optics" (M. Bass, J. M. Enoch, E. W. van Stryland, and W. L. Wolfe, Eds.). McGraw-Hill, New York, 2000.
418. J. W. Perry, in "Nonlinear Optics of Organic Molecules and Polymers" (H. S. Nalwa and S. Miyata, Eds.), p. 813. Chemical Rubber, New York, 1997.
419. Y. P. Sun and J. E. Riggs, *Int. Rev. Phys. Chem.* 18, 43 (1999).
420. A. A. Said, M. Sheik-Bahae, D. J. Hagan, T. H. Wei, J. Wang, J. Young, and E. W. van Stryland, *J. Opt. Soc. Am. B* 9, 405 (1992).
421. B. L. Justus, A. L. Huston, and A. J. Campillo, *Appl. Phys. Lett.* 63, 1483 (1993).
422. K. J. McEwan, P. K. Milsom, and D. B. James, *SPIE* 3472, 42 (1998).
423. K. M. Nashold and W. D. Powell, *J. Opt. Soc. Am. B* 12, 1228 (1995).
424. K. Mansour, M. J. Soileau, and E. W. van Stryland, *J. Opt. Soc. Am. B* 9, 1100 (1992).
425. V. Joudrier, P. Bourdon, F. Hache, and C. Flytzanis, *Appl. Phys. B* 70, 105 (2000).
426. R. Crane, K. Lewis, E. W. van Stryland, and M. Khoshnevisa, Eds., "Materials for Optical Limiting I." Materials Research Society, Warrendale, PA, 1994.
427. P. Hood, R. Pachter, K. Lewis, J. W. Perry, D. Hagan, and R. Sutherland, Eds., "Materials for Optical Limiting II," Vol. 374, Materials Research Society, Warrendale, PA, 1997.
428. V. V. Golovlev, W. R. Garrett, and C. H. Chen, *J. Opt. Soc. Am. B* 13, 2801 (1996).
429. D. G. McLean, R. L. Sutherland, M. C. Brant, D. M. Brandelik, P. A. Fleitz, and T. Pottenger, *Opt. Lett.* 18, 858 (1993).
430. S. Guha, W. T. Roberts, and B. H. Ahn, *Appl. Phys. Lett.* 68, 3686 (1996).
431. S. R. Mishra, H. S. Rawat, and S. C. Mehendale, *Appl. Phys. Lett.* 71, 46 (1997).
432. J. S. Shirk, R. G. S. Pong, S. R. Flom, F. J. Bartoli, M. E. Boyle, and A. W. Snow, *Pure Appl. Opt.* 5, 701 (1996).
433. J. W. Perry, K. Mansour, I. Y. S. Lee, X. L. Wu, P. V. Bedworth, C. T. Chen, D. Ng, S. Marder, and P. Miles, *Science* 273, 1533 (1996).
434. G. R. Allan, S. J. Rychnovsky, C. H. Vezke, T. F. Boggess, and L. Tutt, *J. Phys. Chem.* 98, 3570 (1994).
435. S. Shi, W. Ji, and S. H. Tang, *J. Am. Chem. Soc.* 116, 3615 (1994).
436. T. Xia, A. Dogariu, K. Mansour, D. J. Hagan, A. A. Said, E. W. van Stryland, and S. Shi, *J. Opt. Soc. Am. B* 15, 1497 (1998).
437. M. K. M. Low, H. W. Hou, H. G. Zheng, W. T. Wong, G. X. Jin, Q. Xin, and W. Ji, *Chem. Comm.* 505 (1998).
438. D. C. Hutchings, M. Sheik-Bahae, D. J. Hagan, and E. W. van Stryland, *Opt. Quantum Electronics* 24, 1 (1992).
439. L. W. Tutt and A. Kost, *Nature (London)* 356, 225 (1991).
440. C. Liu, G. Zhao, Q. Gong, K. Tang, X. Jin, P. Cui, and L. Lei, *Opt. Commun.* 184, 309 (2000).
441. V. P. Belousov, I. M. Belousova, E. A. Gavronskaya, V. A. Grigor'ev, A. G. Kalintsev, A. V. Kris'ko, D. A. Kozlovskii, N. G. Mironova, A. G. Skobelev, and M. S. Yur'ev, *J. Opt. Technol.* 68, 876 (2001).
442. Y. Song, C. Zhang, Y. Wang, G. Fang, G. Jin, C. Chang, S. Liu, X. Xin, and H. Ye, *Chem. Phys. Lett.* 326, 341 (2000).
443. J. E. Riggs and Y. P. Sun, *J. Phys. Chem. A* 103, 485 (1999).
444. Y. Sun, Q. H. Gong, S. C. Yang, Y. H. Zhou, L. Fei, Z. Zhou, and D. Qiang, *Opt. Commun.* 102, 205 (1993).
445. J. W. Arbogast, A. P. Darmanyan, C. S. Foote, Y. Rubin, F. N. Diederich, M. Alvarez, S. J. Anz, and R. L. Whetten, *J. Phys. Chem.* 95, 11 (1991).
446. J. W. Arbogast and C. S. Foote, *J. Am. Chem. Soc.* 113, 8886 (1991).
447. B. L. Justus, Z. H. Kafafi, and A. L. Huston, *Opt. Lett.* 18, 1603 (1993).
448. S. R. Mishra, H. S. Rawat, M. P. Joshi, S. C. Mehendale, and K. C. Rustagi, *Proc. SPIE* 2284, 220 (1994).
449. S. R. Mishra and S. C. Mehendale, in "Handbook of Advanced Electronic and Photonic Materials and Devices" (H. S. Nalwa, Ed.), p. 353. Academic, New York, 2000.
450. S. R. Mishra, H. S. Rawat, M. P. Joshi, and S. C. Mehendale, *J. Phys. B* 63, L 2578 (1993).
451. S. R. Mishra, H. S. Rawat, M. P. Joshi, and S. C. Mehendale, *Appl. Phys. A* 63, 223 (1996).
452. K. Mansour, M. J. Soileau, and E. W. van Stryland, *J. Opt. Soc. Am. B* 9, 1100 (1992).
453. X. Sun, R. Q. Yu, G. Q. Xu, T. S. A. Hor, and W. Ji, *Appl. Phys. Lett.* 73, 3632 (1998).

454. P. Chen, X. Wu, X. Sun, J. Lin, W. Ji, and K. L. Tan, *Phys. Rev. Lett.* 82, 2548 (1999).
455. X. Sun, Y. Xiong, P. Chen, J. Lin, J. J. Lim, S. S. Yang, D. J. Hagan, and E. W. van Stryland, *Appl. Opt.* 39, 1998 (2000).
456. J. E. Riggs, D. B. Walker, D. L. Carroll, and Y. P. Sun, *J. Phys. Chem.* 104, 7071 (2000).
457. Z. Jin, L. Huang, S. H. Goh, G. Xu, and W. Ji, *Chem. Phys. Lett.* 352, 328 (2002).
458. L. Vivien, D. Riehl, J. F. Delouis, J. A. Delaire, F. Hache, and E. Anglaret, *J. Opt. Soc. Am. B* 19, 208 (2002).
459. L. Vivien, E. Anglaret, D. Riehl, F. Bacou, C. Journet, C. Goze, M. Andrieux, M. Brunet, F. Lafonta, P. Bernier, and F. Hache, *Chem. Phys. Lett.* 307, 317 (1999).
460. L. Vivien, D. Riehl, E. Anglaret, and F. Hache, *IEEE J. Quantum. Electron.* 36, 680 (2000).
461. L. Vivien, E. Anglaret, D. Riehl, F. Hache, F. Bacou, M. Andrieux, F. Lafonta, C. Journet, C. Goze, M. Brunet, and P. Bernier, *Opt. Commun.* 174, 271 (2000).
462. L. Vivien, D. Riehl, F. Hache, and E. Anglaret, *J. Nonlinear Opt. Phys. Mater.* 9, 297 (2000).
463. L. Vivien, D. Riehl, P. Lancon, F. Hache, and E. Anglaret, *Opt. Lett.* 26, 223 (2001).
464. S. Lefrant, E. Anglaret, L. Vivien, and D. Riehl, *VIDE-Science Tech. Appl.* 56, 2888 (2001).
465. L. Vivien, P. Lancon, D. Riehl, F. Hache, and E. Anglaret, *Carbon* 40, 1789 (2002).
466. S. R. Mishra, H. S. Rawat, S. C. Mehendale, K. C. Rustagi, A. K. Sood, R. Bandyopadhyay, A. Govindaraj, and C. N. R. Rao, *Chem. Phys. Lett.* 317, 510 (2000).
467. S. M. O'Flaherty, R. Murphy, S. V. Hold, M. Cadek, J. N. Coleman, and W. J. Blau, *J. Phys. Chem. B* 107, 958 (2003).
468. Z. Jin, X. Sun, G. Xu, S. H. Goh, and W. Ji, *Chem. Phys. Lett.* 318, 505 (2000).
469. J. Zhao and R. H. Xie, *J. Nanosci. Nanotech.* 3, 459 (2003).
470. J. F. Xu, M. Terrones, N. Grobert, R. Czerw, M. Reyes-Reyes, H. Terrones, P. M. Ajayan, and D. L. Carroll, unpublished observations.
471. J. F. Xu, R. Czerw, and D. L. Carroll, unpublished observations.
472. Z. Jin, L. Huang, S. H. Goh, G. Xu, and W. Ji, *Chem. Phys. Lett.* 352, 328 (2002).
473. L. Smilowitz, D. McBranch, V. Klimov, J. M. Robinson, A. Koskelo, M. Grigorova, and B. R. Mattes, *Opt. Lett.* 21, 922 (1996).
474. M. Iglesias and A. Santos, *Inorg. Chem. Acta* 248, 67 (1996).
475. K. L. Tang, S. J. Zheng, X. L. Jin, H. Zeng, Z. N. Gu, X. H. Zhou, and Y. Q. Tang, *J. Chem. Soc. Dalton Trans.* 19, 3585 (1997).
476. T. Q. Zhang, J. L. Li, P. Gao, Q. H. Gong, K. L. Tang, X. L. Jin, S. J. Zheng, and L. Li, *Opt. Commun.* 150, 201 (1998).
477. K. Dou, J. Y. Du, and E. T. Knobbe, *J. Luminescence* 83, 241 (1999).
478. P. Zhu, P. Wang, W. Qiu, Y. Liu, C. Ye, G. Fang, and Y. Song, *Appl. Phys. Lett.* 78, 1319 (2001).
479. M. Brust, C. J. Kiely, D. Bethell, and D. J. Schiffrin, *J. Am. Chem. Soc.* 120, 12367 (1998).
480. K. Fu, A. Kitaygorodsky, and Y. P. Sun, *Chem. Mater.* 12, 2073 (2000).
481. N. Sun, Y. Wang, Y. Song, Z. Guo, L. Dai, and D. Zhu, *Chem. Phys. Lett.* 344, 277 (2001).
482. R. G. Ispasiou, L. Balogh, O. P. Varnavski, D. A. Tomalia, and T. G. Goodson III, *J. Am. Chem. Soc.* 122, 11005 (2000).
483. M. P. Joshi, J. Swiatkiewicz, F. Xu, P. Prasad, B. A. Reinhardt, and R. Kannan, *Opt. Lett.* 23, 1742 (1998).
484. S. R. Mishra, H. S. Rawat, S. C. Mehendale, K. C. Rustagi, A. K. Sood, R. Bandyopadhyay, A. Govindaraj, and C. N. R. Rao, *Chem. Phys. Lett.* 317, 510 (2000).
485. N. Sun, Z. X. Guo, L. Dai, D. Zhu, Y. Wang, and Y. Song, *Chem. Phys. Lett.* 356, 175 (2002).
486. K. B. Andrew and M. R. Vincent, *J. Am. Chem. Soc.* 121, 4914 (1999).
487. G. R. Uspasiou, L. Balogh, O. P. Varnavski, A. T. Donald, and T. Goodson III, *J. Am. Chem. Soc.* 122, 11005 (2000).
488. S. Link and M. A. El-Sayed, *J. Phys. Chem. B* 103, 8410 (1999).
489. H. Fang, C. Du, S. Qu, Y. Li, Y. Song, H. Li, H. Liu, and D. Zhu, *Chem. Phys. Lett.* 364, 290 (2002).
490. Z. Suo, Y. Song, S. Yao, Y. Wang, and Z. Xu, *Chem. Phys. Lett.* 342, 497 (2001).
491. Y. Li, S. Wang, F. Li, C. Du, Z. Shi, D. Zhu, and Y. Song, *Chem. Phys. Lett.* 337, 403 (2001).
492. F. Li, Y. Li, Z. Ge, D. Zhu, Y. Song, and G. Fang, *J. Phys. Chem. Solids* 61, 1101 (2000).
493. V. V. Golovlev, W. R. Garrett, and C. H. Chen, *J. Opt. Soc. Am. B* 13, 2801 (1996).
494. K. I. Priyadarshi, H. Mohan, A. K. Tyagi, and J. P. Mittal, *J. Phys. Chem.* 98, 4756 (1994).
495. M. Konstantaki, E. Koudoumas, S. Couris, J. M. Janot, H. Eddaoudi, A. Deratani, P. Seta, and S. Leach, *Chem. Phys. Lett.* 318, 488 (2000).
496. R. Signorini, M. Meneghetti, R. Bozio, G. Brusatin, P. Innocenzi, M. Guglielmi, and F. D. Negra, *J. Sol-Gel Sci. Tech.* 22, 245 (2001).
497. R. Signorini, M. Meneghetti, R. Bozio, M. Maggini, G. Scorrano, M. Prato, G. Brusatin, P. Innocenzi, and M. Guglielmi, *Carbon* 38, 1653 (2000).
498. R. Gvishi, U. Narang, G. Ruland, D. N. Kumar, and P. N. Prasad, *Appl. Organomet. Chem.* 11, 107 (1997).
499. M. Maggini, G. Scorrano, M. Prato, G. Brusatin, P. Innocenzi, M. Guglielmi, A. Renier, R. Signorini, M. Meneghetti, and R. Bozio, *Adv. Mater.* 7, 404 (1995).
500. A. Kost, L. Tutt, M. B. Klein, T. K. Dougherty, and W. E. Elias, *Opt. Lett.* 18, 334 (1993).
501. J. E. Riggs and Y. P. Sun, *J. Chem. Phys.* 112, 4221 (2000).
502. Y. P. Sun and J. E. Riggs, in "Fullerenes, Recent Advances in the Chemistry and Physics of Fullerenes and Related Materials" (K. M. Kadish and P. V. Kamat, Eds.), p. 393. Electrochemical Society, Pennington, NJ, 1999.

503. A. H. Kawashima, M. M. Wefers, and K. A. Nelson, *Annu. Rev. Phys. Chem.* 46, 627 (1995).
504. M. Shapiro and P. Brumer, *J. Chem. Soc. Faraday Trans.* 93, 1263 (1997).
505. R. J. Gordon, *Annu. Rev. Phys. Chem.* 48, 595 (1997).
506. W. Potz and W. A. Schroeder, "Coherent Control in Atoms, Molecules, and Semiconductors." Kluwer, Dordrecht 1999.
507. E. A. Manykin and A. M. Afanas'ev, *Sov. Phys. JETP*, 25, 828 (1967).
508. C. Asaro, P. Brumer, and M. Shapiro, *Phys. Rev. Lett.* 60, 1634 (1988).
509. C. Chen, Y. Y. Yin, and D. S. Elliott, *Phys. Rev. Lett.* 64, 507 (1990).
510. E. Dupont, P. B. Corkum, H. C. Liu, M. Buchanan, and Z. R. Wasilewski, *Phys. Rev. Lett.* 74, 3596 (1995).
511. M. Shapiro and P. Brumer, *Phys. Rev. Lett.* 77, 2574 (1996).
512. H. Petek, A. P. Heberle, W. Nessler, H. Nagano, S. Kubota, S. Matsunami, N. Moriya, and S. Ogawa, *Phys. Rev. Lett.* 79, 4649 (1997).
513. R. van Leeuwen, M. L. Bajema, and R. R. Jones, *Phys. Rev. Lett.* 82, 2852 (1999).
514. V. B. Deyirmenjian and J. E. Sipe, *Phys. Rev. Lett.* 82, 4942 (1999).
515. A. Hache, Y. Kostoulas, R. Atanasov, J. L. P. Hughes, J. E. Sipe, and H. M. van Driel, *Phys. Rev. Lett.* 78, 306 1997).
516. T. Quang, M. Woldeyohannes, S. John, and G. S. Agarwal, *Phys. Rev. Lett.* 79, 5238 (1997).
517. W. Pötz, *Phys. Rev. Lett.* 79, 3262 (1997).
518. W. Pötz, *Appl. Phys. Lett.* 72, 3002 (1998).
519. X. Hu and W. Pötz, *Appl. Phys. Lett.* 73, 876 (1998).
520. X. Hu and W. Pötz, *Phys. Rev. Lett.* 82, 3116 (1999).
521. D. S. Citrin, *Phys. Rev. Lett.* 82, 3172 (1999).
522. S. Wielandy and A. L. Gaeta, *Phys. Rev. Lett.* 81, 3359 (1998).
523. N. L. Manakov, V. D. Ovsiannikov, and A. F. Starace, *Phys. Rev. Lett.* 82, 4791 (1999).
524. J. Gong, H. J. Woerner, and P. Brumer, *Phys. Rev. E* 68, 056202 (2003).
525. J. Gong, H. J. Woerner, and P. Brumer, *Phys. Rev. E* 68, 026209 (2003).
526. J. Gong, M. Shapiro, and P. Brumer, *J. Chem. Phys.* 118, 2626 (2003).
527. J. Gong and P. Brumer, *Phys. Rev. Lett.* 88, 203001 (2002).
528. J. Gong and P. Brumer, *J. Chem. Phys.* 115, 3590 (2001).
529. J. Gong and P. Brumer, *Phys. Rev. Lett.* 86, 1741 (2001).
530. R. H. Xie, *Phys. Lett. A* 264, 373 (2000).
531. R. H. Xie, R. B. Tao, and G. O. Xu, *Physica A* 246, 529 (1997).
532. R. H. Xie, R. B. Tao, and G. O. Xu, *Physica A* 240, 503 (1997).
533. R. H. Xie, D. H. Liu, and G. O. Xu, *Z. Phys. B* 99, 253 (1996).
534. R. H. Xie, *Phys. Rev. A* 53, 2897 (1996).
535. R. H. Xie and G. O. Xu, *Phys. Rev. E* 54, 2132 (1996).
536. R. H. Xie and G. O. Xu, *Phys. Rev. E* 54, 1402 (1996).
537. R. H. Xie and G. O. Xu, *Opt. Commun.* 128, 48 (1996).
538. R. H. Xie, X. H. Wu, D. H. Liu, and G. O. Xu, *Z. Phys. B* 101, 247 (1996).
539. R. H. Xie, D. H. Liu, and G. O. Xu, *J. Phys. B* 28, 4697 (1995).
540. R. H. Xie and P. Brumer, *Z. Naturforsch. A* 54, 171 (1999).
541. E. Mccullough, Coherently Enhanced Optical KERR effect: Case Study in a Medium of N_2, Master Thesis (M.Sc), University of Toronto, Toronto, 1997.
542. D. Stahel, M. Leoni, and K. Dressler, *J. Chem. Phys.* 79, 2541 (1983).
543. R. H. Xie, G. W. Bryant, J. Zhao, V. H. Smith Jr., A. Di Carlo, and A. Pecchia, *Phys. Rev. Lett.* 90, 206602 (2003).
544. M. Bockrath, J. Hone, A. Zettl, P. L. McEuen, A. G. Rinzler, and R. E. Smalley, *Phys. Rev. B* 61, 10606 (2000).

CHAPTER 4

Models for Optical Properties of Clusters and Nanostructures

Julio A. Alonso

Departamento de Física Teórica, Universidad de Valladolid, 47011 Valladolid, Spain, and Donostia International Physics Center, San Sebastián, Spain

Angel Rubio

Departamento de Física de Materiales, Facultad de Química, Universidad del País Vasco, Centro Mixto CSIC-UPV, and Donostia International Physics Center, San Sebastián, Spain

CONTENTS

ISBN: 1-58883-049-7

Handbook of Theoretical and Computational Nanotechnology
Edited by Michael Rieth and Wolfram Schommers
Volume 7: Pages (209–262)

1. INTRODUCTION

The electronic levels are quantized in atomic clusters and small nanostructures, essentially reflecting the behavior of electrons in a potential well of finite size. The electronic properties are sensitive to the ordering of those electronic levels and to their evolution as the number of atoms in the cluster increases. The optical spectrum provides information on the electronic structure. In particular, the optical response of the clusters depends on their size and also on the cluster structure. This is an important feature, as the determination of the structure is, in general, a hard task, using either experimental techniques or sophisticated total energy calculations, and knowledge of the geometrical structure of a cluster is required as a basis for understanding many of its properties.

In this work we review some of the theoretical methods and models that are currently in use to calculate the response of clusters to general time-dependent external fields. One is often interested in the response to an external field that is not strong, and in such a case it is enough to consider linear response. For the simple case of a system with only one particle, the excited states can be calculated by solving the time-independent Schrödinger equation. Similarly, the probability $P_{if}(t)$ for transitions from an initial state $|\varphi_i\rangle$ to an excited state $|\varphi_f\rangle$ is calculated by solving the time-dependent Schrödinger equation in the presence of the time-dependent perturbing potential and then using the expression

$$P_{if}(t) = |\langle\varphi_f|\psi(t)\rangle|^2 \tag{1}$$

where $|\psi(t)\rangle$ is the solution of the time-dependent Schrödinger equation in the presence of the external field, with initial condition $|\varphi_i\rangle$ at $t = 0$. In the usual case of a many-electron system the whole method becomes very difficult because of the complications introduced by the interactions between the electrons. In this case, the process can be conveniently formulated within the framework of density-functional theory (DFT) [1], which has become the method most often used in the study of the electronic structure of molecules, nanostructures, and solids. In particular, the spectrum of optical excitations can be efficiently calculated by the so-called time-dependent density-functional theory (TDDFT) [2, 3].

This chapter presents a review of the theoretical methods used to study the optical spectrum of clusters and nanostructures based on the TDDFT. Before discussing that general method, the main theoretical concepts required to connect with the experiments are introduced in Section 2, and a simple approach based on the study of sum rules is presented in Section 3. The general formalism of the TDDFT is given in Section 4, and its formulation in the linear response regime in Section 5. Then Sections 6–9 present several practical ways that have been introduced to apply this formalism, with a first illustrative application to atoms. Sections 10–17 then provide abundant examples of the applications of TDDFT to the study of the optical excitations in clusters; mainly, clusters of the metallic elements.

and fullerenes. A recently developed method, based on the explicit solution of the time-dependent Kohn–Sham equations, is reviewed in Sections 18 and 19. This method is also able to treat nonlinear effects, and this is one of its most interesting features. Some examples are shown in Section 20. Finally, in the last section of the chapter, Section 21, a brief account of many-body techniques is presented.

2. RELATION BETWEEN THEORY AND EXPERIMENT

When a nanostructure interacts with an applied time-dependent electric field characterized by an external potential $V_{\text{ext}}(\mathbf{r}, t)$ with Fourier components $V_{\text{ext}}(\mathbf{r}, \omega)$ (we work in the longitudinal gauge),

$$V_{\text{ext}}(\mathbf{r}, t) = \int d\omega e^{-i\omega t} V_{\text{ext}}(\mathbf{r}, \omega) \tag{2}$$

the external field induces a time-dependent perturbation of the electron density $\delta n(\mathbf{r}, t)$ (we neglect magnetic and current-induced effects)

$$\delta n(\mathbf{r}, t) = \int d\omega e^{-i\omega t} \delta n(\mathbf{r}, \omega) \tag{3}$$

The key quantity to calculate the response of the system in the linear regime is the dynamic susceptibility $\chi(\mathbf{r}, \mathbf{r}', \omega)$, which relates the individual components of the induced density to those of the external potential; that is,

$$\delta n(\mathbf{r}, \omega) = \int d^3r' \chi(\mathbf{r}, \mathbf{r}', \omega) V_{\text{ext}}(\mathbf{r}', \omega) \tag{4}$$

For the case of a dipole field, $\delta n(\mathbf{r}, \omega)$ allows us to calculate the induced dipole moment. The dynamical polarizability $\alpha(\omega)$, which is the ratio between the induced dipole moment and the intensity E_0 of the applied field, then becomes

$$\alpha(\omega) = \frac{1}{E_0} \int d^3r z \delta n(\mathbf{r}, \omega) \tag{5}$$

Dissipation results in $\delta n(\mathbf{r}, \omega)$ being a complex function, and its imaginary part represents the power absorption of the cluster that is the result of electronic excitations. By application of the Fermi's Golden Rule, one obtains the photoabsorption cross section

$$\sigma(\omega) = \frac{4\pi\omega}{c} \operatorname{Im} \alpha(\omega) \tag{6}$$

where $\operatorname{Im} \alpha(\omega)$ is the imaginarity part of the dynamical polarizability and c is the velocity of light.

Experiments have been performed to measure the photoabsorption spectrum, especially of metallic clusters. The typical experimental setup [4, 5] is shown in Fig. 1. A beam of neutral clusters is first produced in a cluster source by supersonic expension of a vapor through a small nozzle. The beam is first collimated by a rectangular aperture and then photoionized by ultraviolet light (ultraviolet lamp in the figure). The resulting cluster cations enter a quadrupole mass analyzer (QMA), and selected masses are steered into the detector. The entire cluster beam is illuminated by a collinear and counter-propagating laser beam. On absorption of light, the clusters warm up and fragment. The transverse recoil removes the daughter clusters away from the initial direction of motion of the collimated beam, and the ratio between the number of clusters of a given size arriving at the detector with and without light excitation is proportional to the absorption cross section. One can adjust the QMA to a specific cluster mass and measure the counting rate of that cluster.

In the case of metallic clusters, the process responsible for the fragmentation is the excitation of a collective mode, the so-called surface plasmon. All the valence electrons participate in this collective resonance, moving back and forth uniformly against the positive ionic background [6, 7]. For sodium clusters, for example, the excitation energy of the surface plasmon is about 3 eV. This energy is higher than the binding energy of an atom in

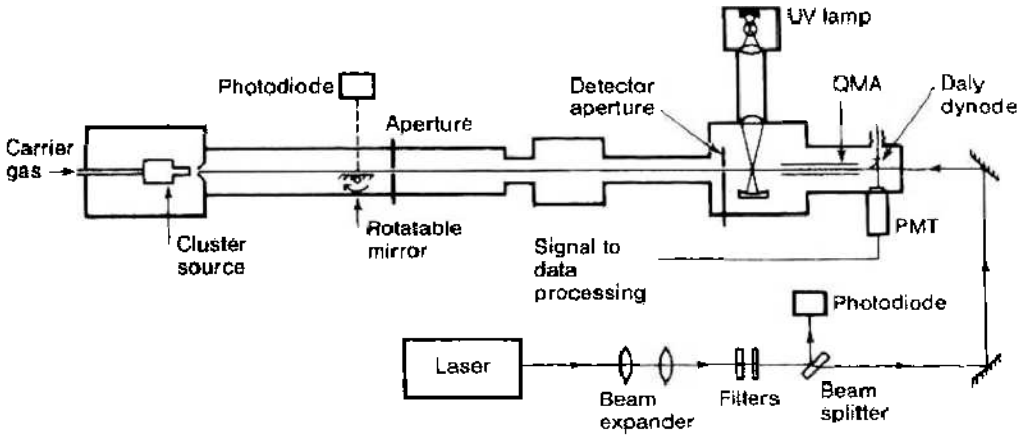


Figure 1. Experiment for measuring the photoabsorption cross section of metallic clusters. The collimated beam is ionized with ultraviolet light (UV) and the ionized clusters enter a quadrupole mass analyzer (QMA). A counter-propagating laser beam heats up the clusters. Some clusters evaporate atoms and are removed from the original beam, and the ratio between the number of clusters of a given size arriving at the detector with and without laser excitation gives the cross section. Reprinted with permission from [4], K. Selby, M. Vollmer, J. Masui, V. Kresin, W. A. de Heer, and W. D. Knight, *Phys. Rev. B* 40, 5417 (1989). © 1989, American Physical Society.

the aggregate, which is below 1 eV, and the excited cluster decays by evaporating single atoms. Using statistical models, the time required to evaporate an atom, if one assumes that the energy of the collective excitation is converted into atomic vibrations, turns out to be, for small clusters, orders of magnitude smaller than the time of flight of the molecular beam in the spectrometer. As a consequence, it can be assumed that the photoabsorption and photofragmentation cross sections are equal. When the cluster size increases, the time required to evaporate atoms also increases, and multiphoton absorption techniques have been used [6]. The surface plasmons in metallic clusters are similar to the giant dipole resonances in nuclei [8].

The integral of the photoabsorption cross section leads to the dipole sum rule (or the Thomas–Reiche–Kuhn sum rule)

$$\int_0^\infty \sigma(\omega)d\omega = 2\pi^2 \frac{e^2\hbar^2}{m_e c} Z \tag{7}$$

where e and m_e are the electron charge and mass, respectively, $\hbar$ is Planck's constant divided by 2π, and Z is the total number of electrons taking part in the collective motion. The experimental determination of $\sigma(\omega)$ then helps identify the collective nature of a resonance. The observed resonances of alkali clusters account for $\approx 60\%$ of the total dipole strength.

For spherical metal particles of a diameter $2R$, small compared to the photon wavelength, the classical theory of the dynamical polarizability [9] gives the following expression for the photoabsorption cross section:

$$\sigma(\omega) = \frac{4\pi Z e^2}{m_e c} \frac{\omega^2 \Gamma}{\left(\omega^2 - \omega_{\text{Mie}}^2\right)^2 + (\omega\Gamma)^2} \tag{8}$$

where ω_{Mie} and Γ represent the frequency and the width of the resonance, respectively. This relation assumes that all the dipole oscillator strength is exhausted by the surface plasma resonance at ω_{Mie} (plasma–pole model). The frequency of the single-dipole resonance, representing the collective oscillation of the valence electrons with respect to the positive ions, is related to the cluster radius by

$$\omega_{\text{Mie}} = \sqrt{\frac{Z\hbar^3 c}{m_e R^3}} \tag{9}$$

This gives for ω_{Mie} a value equal the bulk plasma frequency ω_{pl} divided by $\sqrt{3}$ ($\omega_{\text{pl}} = \sqrt{4\pi e^2 n^0/m_e}$, where n^0 is the average electron density in the bulk).

3. SUM RULES

The photoabsorption cross section is determined in Eq. (6) by the imaginary part of the polarizability tensor $\alpha(\omega)$. Alternatively [6], one can use the strength function $S(E)$

$$S(E) = \sum_n \delta(E - E_n)|\langle n|Q|0\rangle|^2 \tag{10}$$

connected to $\alpha(\omega)$ by

$$S(E) = -\frac{1}{\pi} \operatorname{Im} \alpha(\omega) \tag{11}$$

In Eq. (10), the ket $|0\rangle$ represents the electronic ground state of the cluster, and the sum $\sum_n$ is extended over the excited many-body states $|n\rangle$ of the system. The quantities E_n are the excitation energies, and $E = \hbar\omega$. The operator Q represents the external field, the electric dipole operator in most cases of interest. The full response is not required in some applications, and a knowledge of some moments

$$m_k = \int dE\, E^k S(E) \tag{12}$$

of the strength function is sufficient to have a correct picture of the physical processes (in this equation, k is an integer number, positive or negative). For instance, the average energy $\overline{E}$ and the variance σ^2 of $\sigma(\omega)$ can be obtained from m_0, m_1, and m_2, as $\overline{E} = m_1/m_0$ and $\sigma^2 = m_2/m_0 - (m_1/m_0)^2$. A direct evaluation of the moments m_k is difficult, because the whole excitation spectrum is needed, but the odd moments can be easily obtained with RPA precision [6, 10]. The RPA (random phase approximation) [11], can be derived as the small amplitude limit of the time-dependent Hartree–Fock theory by linearizing the equations of motion. The characteristic feature of RPA is to construct excited states as a superposition of one particle-hole excitation. The linear moment m_1 for the electric dipole operator is model independent, and its RPA value

$$m_1 = \frac{\hbar^2 e^2 Z}{2m_e} \tag{13}$$

is exact, so one obtains the sum rule of Eq. (7). By defining mean energies as $e_k = (m_k/m_{k-2})^{1/2}$, the bounds $e_1 \le \overline{E} \le e_3$ and $\sigma^2 \le (e_3^2 - e_1^2)/4$ can be proven. Also

$$m_{-1} = \alpha/2 \tag{14}$$

where α is the static polarizability [12]. As a consequence, one may estimate the centroid and variance of $S(E)$ by evaluating the three RPA moments m_{-1}, m_1 and m_3. The physical significance of the upper limit e_3 of $\overline{E}$ is that of a rapid oscillation (diabatic) of the valence electrons against the ions, whereas the lower limit e_1 is connected with a slow adiabatic motion of the electrons adjusting their density at any moment to the external field. This lower limit turns out to be a good estimation of the experimental energy of the collective excitation for metallic clusters [6, 13]. The moment m_3 represents the restoring force parameter for the collective translational oscillations of the electrons against the ionic background [6]. For a spherically symmetric electron density $n(r)$, m_3 is given by an overlap integral of the electronic and ionic densities [14]

$$m_3 = \frac{e^4}{2}\left(\frac{\hbar^2}{m}\right)\frac{4\pi}{3}\int n_+(r)n(r)d^3r \tag{15}$$

This integral is easily evaluated for a metallic cluster of radius R in the spherical jellium model (SJM). In this model, the positive charge background of the ions is smeared out over the volume of the cluster to form a distribution of positive charge with density

$$n_+(\mathbf{r}) = n_+^0 \Theta(R - r) \tag{16}$$

where $\Theta(R-r)$ is the step function with values 1 for $r < R$ and 0 for $r > R$. The radius R of the positive background is related to the number of atoms N in the cluster by the equation

$$\frac{4}{3}\pi R^3 = N\Omega \tag{17}$$

where Ω is the experimental volume per atom in the bulk metal. Also, the constant n_+^0 is related to Ω and to the valence z ($z = 1$ for alkali elements) by

$$z = n_+^0 \Omega \tag{18}$$

For the SJM, the moment m_3 becomes

$$m_3 = \frac{\hbar^4 e^4 Z}{2m_e^2 r_s^3}\left(1 - \frac{\delta Z}{Z}\right) \tag{19}$$

where r_s is related to the average electron density n^0 of the bulk metal by

$$r_s = \left(\frac{3}{4\pi n^0}\right)^{1/3} \tag{20}$$

and δZ measures the spill-out of the electronic charge beyond the radius of the positive background

$$\delta Z = 4\pi \int_R^\infty r^2 n(r)\, dr \tag{21}$$

By neglecting the spill-out charge, then $2m_{-1} = R^3$ (the classical polarizability of a metallic sphere is $\alpha = R^3$), and $e_3 = \sqrt{\hbar^2 e^2 Z/m_e \alpha} = e_1$ gives the resonance frequency of the classical Mie surface plasmon ω_{Mie}. In general, $\overline{E}$ is a reasonable estimation for the resonance energy when most of the absorption strength is concentrated in a narrow region. This is the starting point of the plasmon–pole models. In those models $e_1 = e_3$ and the knowledge of the static polarizability determines the value of the dipolar plasma resonance energy as

$$e_3 = \hbar\omega_{\text{pl}}\sqrt{\left(\frac{1}{3} - \frac{\delta Z}{3Z}\right)} \tag{22}$$

For $\delta Z = 0$, e_3 becomes equal to the classical Mie frequency ($\omega_{\text{Mie}} = \omega_{\text{pl}}/\sqrt{3}$).

Table 1 gives the calculated values of the bounds e_1 and e_3 of the dipole surface collective mode for some neutral and charged Na and K clusters [15]. Local density approximation (LDA) labels the results obtained using the LDA to exchange and correlation [1, 16] in the calculation of the density and the single-particle orbitals needed in the RPA formulas for m_k and e_k. However, nonlocal (NL) corresponds to the results obtained using a nonlocal

Table 1. Calculated random phase approximation mean energies e_1 and e_3 (in electron volts) of the dipole surface collective mode of some neutral and charged sodium and potassium clusters.

	e_1	e_3		e_1	e_3
LDA Na_8	2.83	3.14	LDA Na_9^+	3.05	3.16
NL Na_8	2.53 (2.53)	2.81	NL Na_9^+	2.79	2.93
LDA Na_{20}	2.91	3.14	LDA Na_{21}^+	3.04	3.19
NL Na_{20}	2.67 (2.46)	2.97	NL Na_{21}^+	2.77	3.03
LDA K_8	2.21	2.32	LDA K_9^+	2.33	2.39
NL K_8	1.96	2.12	NL K_9^+	2.13 (1.93)	2.22
LDA K_{20}	2.25	2.37	LDA K_{21}^+	2.32	2.40
NL K_{20}	2.06	2.24	NL K_{21}^+	2.16 (1.98)	2.28

Note: LDA and NL refer to the local density (LDA) and nonlocal (NL) approximations to exchange and correlation effects. The experimental surface plasma resonance energies are given in parentheses. Data collected from Ref. [15].

approximation, known as the weighted density approximation (WDA) [17], that goes beyond the LDA (a brief description of the WDA is presented in Section 10). The experimental energies of the surface plasma resonance are in reasonable agreement with e_1, in particular for the NL approximation, and those experimental values are given in parentheses next to the calculated e_1(NL). The effect of the cluster charge is to increase the resonance energy. The analysis of e_1 and e_3 for larger clusters shows that the resonance energy increases with size [15], a prediction in agreement with experiment [18]. The theory also allows us to obtain an upper bound of the variance of the photoabsorption cross section, but the predicted widths of the resonances are, of course, larger than the experimental widths. The variance is found to decrease by charging the cluster.

The general expressions for the odd moments corresponding to q- and l-dependent external fields $j_l(qr)Y_{l0}(\theta, \phi)$ are given in Ref. [14]. This field represents the angular decomposition of an incident photon, described as a plane wave $e^{i(qz-\omega t)}$. With those operators, one can analyze the multipolar response and also the inelastic scattering of electrons (relevant for electron energy loss spectroscopy, EELS). In small metallic clusters, and for fields of high multipolarity, there is a competition between the coulombic contribution to the response (diffusivity and collective excitations) and the kinetic energy contribution (single-particle excitations). The later dominates for large l or large momentum transfer $\mathbf{q}$, indicating the vanishing of collective effects. The response of a metallic sphere to a photon of intermediate energy is dominated by dipolar excitations and at large energies by electron-hole excitations [13, 19, 20]. As the size of the cluster increases, higher multipolar excitations start to dominate together with retardation effects (completely neglected until now).

4. TIME-DEPENDENT DENSITY-FUNCTIONAL THEORY: FORMALISM

The original formulation of the Hohenberg–Kohn–Sham DFT [1, 16, 21, 22] is an incomplete theory, so far as it is not, in general, applicable to excited states or to problems involving nonlocal potentials or time-dependent external fields, thus excluding the calculation of optical response properties, electronic spectra, quasiparticles, photochemistry, and so forth. However, theorems have now been proved for TDDFT that extend the applicability of the original theory [23]. The first applications of TDDFT were actually done before its formal development and relied on the analogies with the time-dependent Hartree–Fock (TDHF) theory [24, 25]. Techniques for avoiding summation over virtual states, typical of all perturbation methods, were developed and applied with success. This, together with the simple multiplicative form of the DFT exchange-correlation potential makes the solution of the response equations easier. These methods are now used in all fields, ranging from atomic physics to nuclear physics to condensed matter physics, including applications to the study of the optical properties of clusters [23, 26]. In this section, we introduce the foundations of TDDFT. A novel feature of this formalism, not present in ground-state DFT, is the dependence of the time-dependent density functionals on the initial many-body state. The main practical result of TDDFT is a set of time-dependent Kohn–Sham (TDKS) equations whose structure is similar to the time-dependent Hartree equations. However, the TDKS equations include, exactly in principle, all the many-body effects through an unknown time-dependent local exchange–correlation potential.

Runge and Gross [27] have developed a theory for time-dependent potentials similar to the Hohenberg–Kohn–Sham theory [21, 22]. In a first theorem, a one-to-one mapping between time-dependent potentials and time-dependent densities that are V-representable (i.e., densities that come from antisymmetric N-electron ground-state wavefunctions for some choice of external potential $V(\mathbf{r}, t)$) is proved. A second theorem proves an stationary-action principle. The proof of the first theorem is based on the evolution of a many-particle state $\Psi(t)$ under the influence of a potential $V(t)$, following the time-dependent Schrödinger equation. The initial state $\Psi(t_0) = \Psi_0$ is not required to be the ground state or some other eigenstate of the initial potential $V(\mathbf{r}, t_0) = V_0$, and this means that the case of a sudden switching of the potential is properly included in the formalism. The time-dependent potentials are required to allow for a Taylor expansion around t_0. Then the first theorem can be

formulated in the following way: The densities $n(\mathbf{r})$ and $n'(\mathbf{r})$ evolving from a common initial state $\Psi(t_0) = \Psi_0$ under the influence of two external potentials, $V(\mathbf{r}, t)$ and $V'(\mathbf{r}, t)$ which, both Taylor-expandable, are always different provided that the potentials differ by more than a purely time-dependent ($\mathbf{r}$-independent) function: $V(\mathbf{r}, t) \neq V'(\mathbf{r}, t) + c(t)$. By virtue of this theorem, the time-dependent density determines the external potential uniquely up to an additive, purely time-dependent function. The potential determines the time-dependent wavefunction, which can be considered as a functional of the time-dependent density up to a purely time-dependent phase; that is, $\Psi(t) = e^{-if(t)}\tilde{\Psi}[n](t)$. As a consequence, the expectation value of any quantum mechanical operator $\hat{Q}$ is a unique functional of the density: $Q[n](t) = \langle\tilde{\Psi}[n](t)|\hat{Q}|\tilde{\Psi}[n](t)\rangle$. In addition to their dependence on the density, these functionals depend on the initial state Ψ_0. Furthermore, the time-dependent particle and current densities can be calculated exactly from the following set of hydrodynamical equations:

$$\frac{\partial n(\mathbf{r}, t)}{\partial t} = -\nabla \cdot \mathbf{j}(\mathbf{r}, t) \tag{23}$$

$$\frac{\partial \mathbf{j}(\mathbf{r}, t)}{\partial t} = -i\langle\Psi(t)|[\hat{\mathbf{j}}, \hat{H}(t)]|\Psi(t)\rangle \tag{24}$$

The first equation stands from the continuity equation of quantum mechanics and the second from the quantum mechanical equation of motion of operators, where $\hat{\mathbf{j}}$ is the usual paramagnetic current density operator

$$\hat{\mathbf{j}}(\mathbf{r}) = \frac{1}{2i}\sum_{j=1}^{N}(\nabla_{\mathbf{r}_j}\delta(\mathbf{r} - \mathbf{r}_j) + \delta(\mathbf{r} - \mathbf{r}_j)\nabla_{\mathbf{r}_j}) \tag{25}$$

and $\hat{H}(t)$ is the Hamiltonian operator.

The second theorem deals with the variational principle of the action functional. From quantum mechanics we know that the time-dependent Schrödinger equation with the initial condition $\Psi(t_0) = \Psi_0$ corresponds to a stationary (not minimum) point of the quantum mechanical action integral

$$A = \int_{t_0}^{t_1} dt\left\langle\Psi(t)|i\frac{\partial}{\partial t} - \hat{H}(t)|\Psi(t)\right\rangle \tag{26}$$

From the previous one-to-one mapping between time-dependent potentials and densities, the action is a functional of the density $A[n]$ that must have a stationary point at the correct time-dependent density. Thus, the density can be obtained by solving the Euler equation

$$\frac{\delta A[n]}{\delta n(\mathbf{r}, t)} = 0 \tag{27}$$

with the appropriate boundary conditions. As in the static Hohenberg–Kohn formalism, one can write the action functional as

$$A[n] = B[n] - \int_{t_0}^{t_1} dt \int d\mathbf{r} n(\mathbf{r}, t)V(\mathbf{r}, t) \tag{28}$$

with a universal (Ψ_0-dependent) functional $B[n]$ formally defined as

$$B[n] = \int_{t_0}^{t_1} dt\left\langle\Psi(t)|\frac{\partial}{\partial t} - \hat{T} - \hat{U}\right|\Psi(t)\right\rangle \tag{29}$$

where $\hat{T}$ is the kinetic energy operator and $\hat{U}$ is the operator for the Coulomb interaction between the electrons. The variational Eq. (27) and the hydrodynamical Eqs. (23) and (24) are, of course, equivalent. The functionals $A[n]$ and $B[n]$ are, in general, unknown and are well defined only for V-representable densities.

As in the static case, a time-dependent Kohn–Sham scheme can be introduced by considering a noninteracting system that reproduces the exact interacting density $n(\mathbf{r}, t)$. Assuming

the V representability of the time-dependent densities, we get the following time-dependent Kohn–Sham equations:

$$\left[-\frac{\hbar^2}{2}\nabla^2 + V_H(\mathbf{r},t) + V_{xc}(\mathbf{r},t) + V(\mathbf{r},t)\right]\psi_i(\mathbf{r},t) = i\frac{\partial}{\partial t}\psi_i(\mathbf{r},t) \tag{30}$$

$$n(\mathbf{r},t) = \sum_{i=1}^{N} |\psi_i(\mathbf{r},t)|^2 \tag{31}$$

In Eq. (30), $V(\mathbf{r},t) = V_N(\mathbf{r},t) + V_{ext}(\mathbf{r},t)$ is the sum of the nuclear (or ionic) potential $V_N(\mathbf{r},t)$ and the applied perturbing field $V_{ext}(\mathbf{r},t)$. The nuclei (or the ions) occupy positions $\mathbf{R}_k(t)$. However, $V_H(\mathbf{r},t) = \int \frac{n(\mathbf{r}',t)}{|\mathbf{r}-\mathbf{r}'|}d\mathbf{r}'$ is the time-dependent Hartree potential and $V_{xc}(\mathbf{r},t)$ is the time-dependent exchange-correlation potential that is defined through the equivalence between the interacting and fictitious noninteracting systems. As in the static case, the advantage of the time-dependent Kohn–Sham-scheme lies in its computational simplicity compared to other quantum-chemical schemes as the TDHF or the configuration–interaction (CI) methods.

5. EXCITATION ENERGIES IN TIME-DEPENDENT DENSITY FUNCTIONAL THEORY: LINEAR RESPONSE

TDDFT has become the most popular method for the calculation of excitations in finite systems, both in physics (atomic, molecular, and condensed matter) [2, 3] and in quantum chemistry. As a first approximation of the excitation energies, one can simply take differences $\Delta\varepsilon = \varepsilon_l - \varepsilon_i$, between the ground-state Kohn–Sham eigenvalues. Although this procedure is not entirely justifiable, it is often employed to obtain a first approximation to the excitation spectrum. It is well known that the Kohn–Sham eigenvalues and wavefunctions do not have a precise physical interpretation, with the exception of the eigenvalue of the highest occupied state, ε_{HOMO}, which is equal to minus the ionization potential IP of the system [28, 29]. In addition, Chong et al. [30] have shown that the orbital energies of other occupied levels of atoms and molecules can be interpreted as approximate, but rather accurate, relaxed vertical removal energies. It should be stressed that the relation $\varepsilon_{HOMO} = -IP$ and that the findings of Chong et al. [30] are valid in exact DFT, but not for the approximate exchange–correlation energy functionals currently used. Those relations fail for the LDA, but better functionals, especially those improving the asymptotic behavior of V_{xc}, lead to more accurate results. An excellent agreement between the calculated Kohn–Sham energy eigenvalues and the experimental results for the removal energies of inner electrons of the noble gas atoms was obtained [31] using the nonlocal WDA approximation [17] (see Section 10). A similar good agreement was found [32] between WDA eigenvalues of occupied orbitals of sodium and potassium clusters in the spherical jellium model and quasiparticle energies calculated by Saito et al. [33] using the many-body GW formalism [3].

Another approach, called Δ-SCF (delta self-consistent field), is based on the observation that the Hohenberg–Kohn theorem and the Kohn–Sham scheme can be formulated for the lowest state of each symmetry class [34]. In fact, the single modification of the standard proofs is to restrict the variational principle to wavefunctions of a specific symmetry. The unrestricted variation will clearly yield the ground state. However, the states belonging to different symmetry classes will correspond to excited states. The excitations can then be calculated by simple total-energy differences. However, this approach suffers from two drawbacks: only the lowest lying excitation for each symmetry class is obtainable, and the exchange–correlation functional that now enters the Kohn–Sham equations depends on the particular symmetry we have selected. As specific approximations for a symmetry-dependent xc-functional are not available, one is forced to employ the usual ground-state functionals. The excitation energies calculated in this way are only of moderate quality.

TDDFT allows one to calculate the excited state energies of a many-body system based on information from an ordinary, self-consistent DFT calculation. In the time-dependent approach, one studies the behavior of the system subject to a time-dependent external perturbation. The response of the system is directly related to the N-particle excited states of

the N-particle system. We now present the formalism developed by Petersilka et al. [35] for a spin-unpolarized system. A formally exact Dyson-like representation of the linear density response of an interacting many-electron system in terms of the noninteracting Kohn–Sham response is first derived.

The linear response χ is a functional derivative

$$\chi(\mathbf{r}, t, \mathbf{r}', t') = \left.\frac{\delta n(\mathbf{r}, t)}{\delta V_{\text{ext}}(\mathbf{r}', t')}\right|_{V_{\text{ext}}=0} \tag{32}$$

that measures the degree to which the density responds to the external perturbing potential in first order. χ can also viewed as the coefficient of the linear term in a functional power series of the time-dependent density in terms of V_{ext}; that is,

$$n(\mathbf{r}, t) = n_0(\mathbf{r}, t_0) + \iint \chi(\mathbf{r}, t, \mathbf{r}', t') V_{\text{ext}}(\mathbf{r}', t') d\mathbf{r}' dt' + O(V_{\text{ext}}^2) \tag{33}$$

The linear response χ_s of the fictitious noninteracting Kohn–Sham system is defined in a similar way

$$\chi_s(\mathbf{r}, t, \mathbf{r}', t') = \left.\frac{\delta n(\mathbf{r}, t)}{\delta V_{\text{eff}}(\mathbf{r}', t')}\right|_{V_{\text{ext}}=0} \tag{34}$$

and can be computed exactly (see below). The effective potential V_{eff} corresponding to a given external potential V_{ext} is the time-dependent Kohn–Sham potential

$$V_{\text{eff}}(\mathbf{r}, t) = V_{\text{ext}}(\mathbf{r}, t) + V_N(\mathbf{r}, t) + V_H(\mathbf{r}, t) + V_{\text{xc}}(\mathbf{r}, t) \tag{35}$$

where all the components have been defined after Eq. (30). We now evaluate the following functional derivative

$$\frac{\delta V_{\text{eff}}(\mathbf{r}, t)}{\delta V_{\text{ext}}(\mathbf{r}', t')} = \delta(\mathbf{r}-\mathbf{r}')\delta(t-t') + \iint d\mathbf{r}'' dt'' \chi(\mathbf{r}'', t'', \mathbf{r}', t') \left[\frac{\delta(t-t'')}{|\mathbf{r}-\mathbf{r}''|} + K_{\text{xc}}[n_0](\mathbf{r}, t, \mathbf{r}'', t'')\right] \tag{36}$$

where

$$K_{\text{xc}}[n_0](\mathbf{r}, t, \mathbf{r}', t') = \left.\frac{\delta V_{\text{xc}}[n](\mathbf{r}, t)}{\delta n(\mathbf{r}', t')}\right|_{V_{\text{ext}}=0} \tag{37}$$

is the so-called time-dependent XC kernel (which is a functional of the initial ground-state density). Employing the chain rule for functional derivatives in Eq. (32), we have

$$\chi(\mathbf{r}, t, \mathbf{r}', t') = \iint d\mathbf{r}'' dt'' \frac{\delta n(\mathbf{r}, t)}{\delta V_{\text{eff}}(\mathbf{r}'', t'')} \frac{\delta V_{\text{eff}}(\mathbf{r}'', t'')}{\delta V_{\text{ex}}(\mathbf{r}', t')}\Big|_{V_{\text{ext}}=0} \tag{38}$$

and combining this with the previous equations, we obtain the following Dyson-like equation for the interacting linear response in terms of the Kohn–Sham noninteracting response

$$\chi(\mathbf{r}, t, \mathbf{r}', t') = \chi_s(\mathbf{r}, t, \mathbf{r}', t') + \iint d\mathbf{r}_1 d\mathbf{r}_2 \chi_s(\mathbf{r}, t, \mathbf{r}_1, t_1) \times \left[\frac{\delta(t_1 - t_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} + K_{\text{xc}}[n_0](\mathbf{r}_1, t_1, \mathbf{r}_2, t_2)\right] \chi(\mathbf{r}_2, t_2, \mathbf{r}', t') \tag{39}$$

Taking the Fourier transform with respect to time, the exact frequency-dependent linear response function becomes

$$\chi(\mathbf{r}, \mathbf{r}', \omega) = \chi_s(\mathbf{r}, \mathbf{r}', \omega) + \iint d\mathbf{r}_1 d\mathbf{r}_2 \chi_s(\mathbf{r}, \mathbf{r}_1, \omega) K(\mathbf{r}_1, \mathbf{r}_2, \omega) \chi(\mathbf{r}_2, \mathbf{r}', \omega) \tag{40}$$

where the Kernel $K(\mathbf{r}_1, \mathbf{r}_2, \omega)$ is given by

$$K(\mathbf{r}_1, \mathbf{r}_2, \omega) = \frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|} + K_{\text{xc}}[n_0](\mathbf{r}_1, \mathbf{r}_2, \omega) \tag{41}$$

Equation (40) has to be solved iteratively. Multiplying both sides of Eq. (39) with $V_{\text{ext}}(\mathbf{r}', t')$ and integrating over $\mathbf{r}'$ and t' yields

$$n_1(\mathbf{r}, t) = \int dt' \int d\mathbf{r}' \chi_s(\mathbf{r}, t, \mathbf{r}', t') V_{\text{eff}}(\mathbf{r}', t') \tag{42}$$

where the notation n_1 has been chosen to explicitly indicate linear response. Fourier transforming with respect to time we have an exact representation

$$n_1(\mathbf{r}, \omega) = \int d\mathbf{r}' \chi_s(\mathbf{r}, \mathbf{r}', \omega) V_{\text{eff}}(\mathbf{r}', \omega) \equiv \int d\mathbf{r}' \chi_s(\mathbf{r}, \mathbf{r}', \omega) \left[V_{\text{ext}}(\mathbf{r}', \omega) + \int d\mathbf{x} K(\mathbf{r}', \mathbf{x}, \omega) n_1(\mathbf{x}, \omega) \right] \tag{43}$$

showing that the exact linear density response of an interacting system, which can be directly used to compute the photoabsorption cross section [see Eqs. (5) and (6)], can be written as the linear density response on a noninteracting system to the effective perturbation V_{eff}.

6. TECHNIQUES TO STUDY EXCITATIONS IN LINEAR RESPONSE: MATRIX EIGENVALUE METHOD

Let us consider the analytical structure of the interacting linear response function for a finite system. We now show that the linear response function has poles at $\omega = \Omega_m$, where Ω_m are the N-particle excitation energies of the actual system. Using the Kubo formula for the density response in terms of a retarded density–density correlation function [36]

$$\chi(\mathbf{r}, t, \mathbf{r}', 0) = -i\Theta(t) \langle 0 | [\hat{n}(\mathbf{r}, t), \hat{n}(\mathbf{r}', 0)] | 0 \rangle \tag{44}$$

where $\Theta(t)$ is the step function, $\hat{n}(\mathbf{r}, t)$ is the density operator, and $|0\rangle$ is the exact many-particle ground state, and inserting a complete set of many-particle states $|m\rangle$, we have

$$\chi(\mathbf{r}, t, \mathbf{r}', 0) = -i\Theta(t) \sum_m \langle 0 | \hat{n}(\mathbf{r}, t) | m \rangle \langle m | \hat{n}(\mathbf{r}', 0) | 0 \rangle + i\Theta(t) \sum_m \langle 0 | \hat{n}(\mathbf{r}', 0) | m \rangle \langle m | \hat{n}(\mathbf{r}, t) | 0 \rangle \tag{45}$$

Notice that the density operator is particle-number conserving; that is, it only connects N-particle excited states with the N-particle ground state. By taking into account the time-evolution of the density operator $\hat{n}(\mathbf{r}, t) = e^{iHt} \hat{n}(\mathbf{r}, 0) e^{-iHt}$, the linear response function can be revritten as

$$\chi(\mathbf{r}, t, \mathbf{r}', 0) = -i\Theta(t) \left[\sum_m A_m(\mathbf{r}') A_m^*(\mathbf{r}) e^{i(E_0 - E_m)t} - \sum_m A_m(\mathbf{r}) A_m^*(\mathbf{r}') e^{-i(E_0 - E_m)t} \right] \tag{46}$$

where $A_m(\mathbf{r}) = \langle 0 | \hat{n}(\mathbf{r}, t) | m \rangle$, and E_m is the energy of the mth N-particle excited state. Fourier tranformation leads to the spectral representation

$$\chi(\mathbf{r}, \mathbf{r}', \omega) = \sum_m \frac{\langle 0 | \hat{n}(\mathbf{r}) | m \rangle \langle m | \hat{n}(\mathbf{r}') | 0 \rangle}{\hbar\omega - (E_m - E_0) + i\delta} - \frac{\langle 0 | \hat{n}(\mathbf{r}') | m \rangle \langle m | \hat{n}(\mathbf{r}) | 0 \rangle}{\hbar\omega + (E_m - E_0) + i\delta} \tag{47}$$

One can observe that the time-Fourier transformation of the linear response function has poles at $\hbar\omega = \pm(E_m - E_0)$; that is, at the excitation energies of the N-particle system. We have seen previously that the interpretation of the Kohn–Sham single-particle energy differences as excitation energies has no rigorous basis and that in practical implementations they may differ substantially from the experimental excitation energies.

In the work of Petersilka et al. [35], excitation energies are extracted from TDDFT by exploiting the fact that the frequency-dependent linear density response of a finite system has discrete poles at the excitation energies of the unperturbed system. The noninteracting susceptibility, which is needed to solve Eq. (40), can be computed directly from the Kubo formula and has the form

$$\chi_s(\mathbf{r}, \mathbf{r}', \omega) = \sum_{k,j} (f_k - f_j) \frac{\psi_k^*(\mathbf{r}) \psi_j(\mathbf{r}) \psi_j^*(\mathbf{r}') \psi_k(\mathbf{r}')}{\hbar(\omega - \omega_{kj}) + i\delta} \tag{48}$$

where $\hbar\omega_{kj}=\varepsilon_j-\varepsilon_k$ and f_j are the Fermi occupation numbers of the jth single-particle orbitals obtained from the Kohn–Sham calculation. As a function of ω, χ_s has poles at the Kohn–Sham orbital energy differences ω_{kj}. To calculate the shift toward the true excitation energy Ω, we rewrite Eq. (43) for the induced density as

$$\int d\mathbf{x}\left[\delta(\mathbf{r}-\mathbf{x})-\int d\mathbf{r}'\chi_s(\mathbf{r},\mathbf{r}',\omega)K(\mathbf{r}',\mathbf{x},\omega)\right]n_1(\mathbf{x},\omega)=\int d\mathbf{r}'\chi_s(\mathbf{r},\mathbf{r}',\omega)V_{\text{ext}}(\mathbf{r}',\omega) \tag{49}$$

where we see that the integral operator acting on n_1 on the left-hand side cannot be invertible for $\omega\to\Omega$ because if it were invertible we could act with the inverse operator and get on the right-hand side a finite value, in contradiction with the fact that n_1 has poles at the excitation energies. Thus, the true excitation energies Ω can be characterized as those frequencies at which the eigenvalues of this integral operator vanish. As noted previously, in the limit $\omega\to\Omega$ the linear density n_1 has a pole, whereas the right-hand side of Eq. (49) remains finite. For the equality to hold, it is therefore required that the operator acting on n_1 on the left-hand side of Eq. (49) has zero eigenvalues at the excitation energies Ω; that is, $\lambda(\omega)\to 1$, when $\omega\to\Omega$, where $\lambda(\omega)$ is the solution of the eigenvalue equation

$$\int d\mathbf{r}''\int d\mathbf{r}'\chi_s(\mathbf{r},\mathbf{r}'',\omega)\left[\frac{1}{|\mathbf{r}''-\mathbf{r}'|}+K_{\text{xc}}[n_0](\mathbf{r}'',\mathbf{r}',\omega)\right]\xi(\mathbf{r}',\omega)=\lambda(\omega)\xi(\mathbf{r},\omega) \tag{50}$$

This is a rigorous statement that allows the determination of the excitation energies from the knowledge of the independent-particle Kohn–Sham susceptibility and the kernel K_{xc}. It is possible to transform this equation into another eigenvalue equation having the true excitation energies Ω as eigenvalues [37]. We begin by defining the quantity

$$\zeta_{jk}(\omega)=\iint d\mathbf{r}'d\mathbf{r}''\psi_j^*(\mathbf{r}'')\psi_k(\mathbf{r}'')\left[\frac{1}{|\mathbf{r}''-\mathbf{r}'|}+K_{\text{xc}}[n_0](\mathbf{r}'',\mathbf{r}',\omega)\right]\xi(\mathbf{r}',\omega) \tag{51}$$

With the help of ζ_{jk}, Eq. (50) can be written in the form

$$\sum_{jk}\frac{(f_k-f_j)\psi_j(\mathbf{r})\psi_k^*(\mathbf{r})}{\omega-(\varepsilon_j-\varepsilon_k)+i\delta}\zeta_{jk}(\omega)=\lambda(\omega)\xi(\mathbf{r},\omega) \tag{52}$$

By solving this equation for $\xi(\mathbf{r},\omega)$ and inserting the result into Eq. (51) we arrive at the equation

$$\sum_{j'k'}\frac{(f_{k'}-f_{j'})M_{jk,j'k'}}{\omega-(\varepsilon_{j'}-\varepsilon_{k'})+i\delta}\zeta_{j'k'}(\omega)=\lambda(\omega)\zeta_{jk}(\omega) \tag{53}$$

where the coupling-matrix element is defined

$$M_{jk,j'k'}(\omega)=\iint d\mathbf{r}d\mathbf{r}'\psi_j^*(\mathbf{r})\psi_k(\mathbf{r})\psi_{j'}(\mathbf{r}')\psi_{k'}^*(\mathbf{r}')\left[\frac{1}{|\mathbf{r}-\mathbf{r}'|}+K_{\text{xc}}[n_0](\mathbf{r},\mathbf{r}',\omega)\right] \tag{54}$$

Introducing now the new eigenvector

$$\beta_{jk}=\frac{\zeta_{jk}(\Omega)}{\Omega-(\varepsilon_j-\varepsilon_k)} \tag{55}$$

taking the $\eta\to 0$ limit, and using the condition $\lambda(\Omega)=1$, it is straightforward to recast Eq. (53) into the eigenvalue equation

$$\sum_{j'k'}[\delta_{jj'}\delta_{kk'}(\varepsilon_{j'}-\varepsilon_{k'})+(f_{k'}-f_{j'})M_{jk,j'k'}(\Omega)]\beta_{j'k'}=\Omega\beta_{jk} \tag{56}$$

The exact solution of this equation fully incorporates the collective electronic excitations. To solve the eigenvalue Eq. (53), one can expand all quantities in an appropriate basis and solve numerically the resulting matrix-eigenvalue equation. An alternative way to calculate Ω is to expand all quantities appearing in Eq. (50) about one particular energy difference

$\Delta\varepsilon = \varepsilon_j - \varepsilon_k$ between the Kohn–Sham eigenvalues of the occupied orbital k and the unoccupied orbital j. Assuming that the true excitation energy is not far away from $\Delta\varepsilon$, it is sufficient to consider only the lowest-order terms in those expansions. This leads to

$$\Omega_{jk} = (\varepsilon_j - \varepsilon_k) + B(\Delta\varepsilon) \tag{57}$$

$B(\Delta\varepsilon)$ is a correction given by

$$B(\Delta\varepsilon) = 2R \iint d\mathbf{r} d\mathbf{r}' \psi_j(\mathbf{r}) \psi_j^*(\mathbf{r}') \psi_k(\mathbf{r}') \psi_k^*(\mathbf{r}) \left[\frac{1}{|\mathbf{r} - \mathbf{r}'|} + K_{xc}[n_0](\mathbf{r}, \mathbf{r}', \Delta\varepsilon) \right] \tag{58}$$

where R indicates the real part of the expression. Equations (57) and (58) provide a simple and fast way to calculate excitations energies, although this method is not as precise as solving Eq. (56) directly. The approximation can be viewed as an attempt to correct the Kohn–Sham excitation energies individually without including collective electronic effects. Apart from the truncation of the expansions, two main approximations are necessary: first, the static Kohn–Sham orbitals have to be calculated with an approximate static exchange-correlation potential $V_{xc}(\mathbf{r})$, and second, the frequency-dependent XC kernel $K_{xc}[n_0](\mathbf{r}, \mathbf{r}', \omega)$ has to be approximated.

It is also possible to derive an operator whose eigenvalues are the square of the excitation energies, reducing then the dimensions of the matrix Eq. (56) [38, 39]. The oscillator strengths can then be obtained from the eigenfunctions of that operator

$$\left(\omega_{jk\sigma}^2 \delta_{jj'} \delta_{kk'} \delta_{\sigma\tau} + 2\sqrt{f_{jk\sigma}} \overline{\omega_{jk\sigma}} M_{jk\sigma, j'k'\tau} \sqrt{f_{j'k'\tau} \omega_{j'k'\tau}} \right) \mathbf{F}_n = \Omega_n^2 \mathbf{F}_n \tag{59}$$

Here σ is the spin index, $\omega_{jk\sigma} = \varepsilon_{j\sigma} - \varepsilon_{k\sigma}$ are the Kohn–Sham transition energies, and $f_{jk\sigma}$ is the difference between the occupations of the j and k states. The coupling matrix $M_{jk\sigma, j'k'\tau}$ is the generalization of $M_{jk,j'k}$ of Eq. (54) to the spin-dependent case. Some simplifications can also be introduced to approximately solve this equation [39]. Let us consider the spin unpolarized case; that is, $f_{jk\uparrow} = f_{jk\downarrow} = f_{jk}$ (the results can be easily generalized to any spin configuration). Assuming that the coupling between different one-electron transitions is weak, one can neglect all matrix elements with $j \neq j'$ and $k \neq k'$. The only remaining off-diagonal elements of the coupling matrix are generated by the spin index, and Eq. (59) reduces to a series of independent 2×2 matrix equations. Solving for the transition energies,

$$\Omega_{jk} \approx \left\{ \omega_{jk} \left[\omega_{jk} + 2 f_{jk} \left(M_{jk\uparrow, jk\uparrow} \pm M_{jk\uparrow, jk\downarrow} \right) \right] \right\}^{1/2} \tag{60}$$

The plus sign describes transitions to the singlet excited state and the minus corresponds to triplet transitions. The approximation can again be viewed as an attempt to correct the Kohn–Sham excitation energies individually without including collective electronic effects. Assuming that the corrections with respect to the Kohn–Sham transition energies are small, one arrives at a further simplification by taking a linear expansion around ω_{jk}

$$\Omega_{jk} \approx (\varepsilon_j - \varepsilon_k) + f_{jk} \left(M_{jk\uparrow, jk\uparrow} \pm M_{jk\uparrow, jk\downarrow} \right) \tag{61}$$

This equation is equivalent to the result in Eq. (57).

7. THE EXCHANGE CORRELATION KERNEL

As we have seen in the previous sections, a main ingredient in linear response theory is the XC kernel $K_{xc}[n_0](\mathbf{r}, \mathbf{r}', \omega)$. This is a complex quantity that includes all nontrivial many-body effects. Approximate kernels have been proposed over the past years, and we consider here some of the most commonly used ones. The simplest and most used one is the ALDA (adiabatic local density approximation) kernel [40]

$$K_{xc}^{ALDA}[n](\mathbf{r}, t, \mathbf{r}', t') = \delta(\mathbf{r} - \mathbf{r}') \delta(t - t') K_{xc}^{LDA}[n]|_{n = n(\mathbf{r}, t)} \tag{62}$$

where

$$K_{xc}^{LDA}[n] = \frac{dV_{xc}^{LDA}[n](\mathbf{r})}{dn} \tag{63}$$

is just the derivative of the XC potential of a homogeneous electron gas with density n; that is, $V_{xc}^{LDA}[n](\mathbf{r}) = \frac{\partial E_{xc}^{hom}[n]}{\partial n(\mathbf{r})}$. The ALDA kernel is local in both the space and time coordinates.

Another XC kernel was derived by Petersilka et al. [35]. Its derivation starts from a simple analytic approximation to the exchange-only optimized effective potential (OEP) [41, 42]. The OEP kernel has the following form

$$K_{x}^{OEP}(\mathbf{r}, t, \mathbf{r}', t') = -\delta(t - t')\frac{1}{2}\frac{1}{|\mathbf{r} - \mathbf{r}'|}\frac{\left|\sum_k^{occ}\psi_k(\mathbf{r})\psi_k^*(\mathbf{r}')\right|^2}{n(\mathbf{r})n(\mathbf{r}')} \tag{64}$$

As in the ALDA case, the OEP kernel is local in time. It should be noticed that in the calculation of excitation energies through Eqs. (57) and (58) using the ALDA or the OEP kernels, only ground-state quantities are needed; that is, Kohn–Sham orbitals and energy eigenvalues.

Recently, Görling and coworkers derived and implemented the exact-exchange kernel (EXX) for solids [43]. Using the EXX, Kurth and Von Barth [44] have computed the electronic excitations for atoms obtaining good agreement with experiment, and their results are encouraging for the development of new kernels having exact-exchange as an ingredient. A simplified version of the EXX scheme has been applied to clusters and will be reviewed in Section 19.

8. APPLICATION OF LINEAR RESPONSE THEORY TO ATOMS

The lowest excitation energies of atoms of the alkaline earth and the zinc groups, calculated from Eqs. (56) and (57), are given in Table 2 [35, 39]. The LDA columns report results obtained with the LDA XC potential and the ALDA kernel. In contrast, the OEP columns correspond to the use of the OEP potential and the OEP kernel. First of all, the table shows how the difference $\Delta\varepsilon$ of energy eigenvalues is corrected by the term in Eq. (58), leading to improved excitation energies Ω^{approx}. The OEP values Ω_{OEP}^{approx} are superior to the LDA results Ω_{LDA}^{approx} and are also better than the Δ_{SCF} values. The Δ_{SCF} results are obtained by subtracting the total energies corresponding to the ground state and excited configurations. Petersilka et al. [35] have argued that the main reason for the superiority of the OEP potential is that it is self-interaction free, and therefore has the correct asymptotic $-1/r$ behavior far from the nucleus, whereas the LDA XC potential decays exponentially. From this argument, the importance of a good description of the static XC potential becomes evident. A study of the excitation energies of the CO molecule [46] again indicates a good agreement with experiment.

However, Chelikowsky and coworkers [39] have pointed out that the differences between Ω_{LDA}^{approx} and the experimental excitation energies are caused by the approximations involved

Table 2. The lowest $^1S \rightarrow {}^1P$ excitation energies for alkaline-earth atoms and atoms of the Zn group.

	Exp	Ω_{LDA}^{exact} Eq. (56)	Ω_{LDA}^{approx} Eq. (57)	Ω_{OEP}^{approx} Eq. (57)	Δ_{SCF}	$\Delta\varepsilon_{LDA}$	$\Delta\varepsilon_{OEP}$
Be	5.28	4.94	5.43	5.33	4.50	3.50	3.52
Mg	4.34	4.34	4.76	4.45	4.07	3.39	3.18
Ca	2.94	3.22	3.56	3.18	2.87	2.39	2.14
Sr	2.69	2.96	3.28	2.86	2.62	2.22	1.92
Zn	5.79	5.71	6.54	5.74	5.48	4.79	4.27
Cd	5.41	5.10	5.86	5.11	4.71	4.12	3.66

Note. The experimental values (first column) [45] are compared with results calculated from Eqs. (56) and (57) within local density approximation (LDA) and optimized effective potential (OEP) and with ordinary Δ_{SCF} values. The corresponding Kohn–Sham orbital energy differences $\Delta\varepsilon$ are shown in the last two columns. All energies are given in electron volts. SCF, self-consistent field.

Table 3. Comparison between the experimental singlet and triplet excitation energies for atoms and excitation energies calculated by the time-dependent local density approximation (TDLDA), optimized effective potential (OEP), and Δ_{SCF} methods [39].

Atom	Transition	Experiment	TDLDA Eq. (59)	OEP	Δ_{SCF}
Be	$^1S \rightarrow {}^1P$	5.28	4.94	5.33	4.50
	$^1S \rightarrow {}^3P$	2.72	2.45	1.88	2.46
Mg	$^1S \rightarrow {}^1P$	4.34	4.34	4.45	4.07
	$^1S \rightarrow {}^3P$	2.72	2.79	2.05	2.80
Ca	$^1S \rightarrow {}^1P$	2.94	3.22	3.18	2.87
	$^1S \rightarrow {}^3P$	1.89	1.93	1.22	1.96
Sr	$^1S \rightarrow {}^1P$	2.69	2.96	2.86	2.62
	$^1S \rightarrow {}^3P$	1.82	1.82	1.10	1.84
Zn	$^1S \rightarrow {}^1P$	5.79	5.71	5.74	5.48
	$^1S \rightarrow {}^3P$	4.05	4.27	3.40	4.30
Cd	$^1S \rightarrow {}^1P$	5.41	5.10	5.11	4.71
	$^1S \rightarrow {}^3P$	3.88	3.69	2.87	3.70

Note: The values of the experimental triplet transitions are the average over different spin-orbit components. All energies are given in eV.

in Eqs. (57) and (58). When the excitation energies are calculated by exactly solving Eq. (56) (the results are given in the column $\Omega_{\mathrm{LDA}}^{\mathrm{exact}}$ of Table 2), the experimental and theoretical results agree within 5–10% for all atoms. The differences between $\Omega_{\mathrm{LDA}}^{\mathrm{exact}}$ and $\Omega_{\mathrm{LDA}}^{\mathrm{approx}}$ show the important role of collective electronic effects. The electronic correlations are particularly large for Zn and Cd, being caused by d levels, which are close in energy. The encouraging conclusion is that the wrong asymptotic behavior of the LDA potential appears not to be as important for the excited state properties as previously thought [35, 47], but more work needs to be done to clarify this point. Table 3 presents the calculated singlet $^1S \rightarrow {}^1P$ and triplet $^1S \rightarrow {}^3P$ transition energies for the same atoms [39]. The TDLDA transition energies are in better agreement with experiment than the results obtained by the OEP or Δ_{SCF} methods. Although the singlet excitation energies from the OEP method are almost as accurate as the TDLDA energies, the triplet transition energies are less accurate, because the OEP method does not include coulomb correlation effects, which play a significant role for triplets.

9. SELF-CONSISTENT GREEN'S FUNCTION METHOD

There is another traditional method [24, 25] to calculate the excitations in linear response by self-consistently solving Eq. (43); this requires the evaluation of the independent-particle susceptibility χ_s given by Eq. (48). This expression involves an explicit sum over the complete energy eigenvalue spectrum of the LDA effective potential V_{eff}. That is, not only the occupied orbitals are required, but the unoccupied bound states and the continuum orbitals are required as well. This problem can be circumvented by using the Green's function associated to the LDA Schrödinger-type equation

$$\left[E + \frac{1}{2}\nabla^2 - V_{\mathrm{eff}}(\mathbf{r})\right] G(\mathbf{r}, \mathbf{r}'; E) = \delta(\mathbf{r} - \mathbf{r}') \tag{65}$$

$G(\mathbf{r}, \mathbf{r}'; E)$ has an eigenfunction expansion

$$G(\mathbf{r}, \mathbf{r}', E) = \sum_m \frac{\psi_m(\mathbf{r})\psi_m^*(\mathbf{r}')}{E - \varepsilon_m \pm i\delta} \tag{66}$$

and satisfies the symmetry condition

$$G(\mathbf{r}, \mathbf{r}'; E) = G^*(\mathbf{r}', \mathbf{r}, E) \tag{67}$$

Using the Green functions, the expression for the susceptibility χ_s becomes [24, 25]

$$\chi_s(\mathbf{r},\mathbf{r}',\omega)=\sum_{k=1}^{\text{occ}}[\psi_k^*(\mathbf{r})\psi_k(\mathbf{r}')G(\mathbf{r},\mathbf{r}';\varepsilon_k+\hbar\omega)+\psi_k(\mathbf{r})\psi_k^*(\mathbf{r}')G^*(\mathbf{r},\mathbf{r}';\varepsilon_k-\hbar\omega)] \tag{68}$$

where the sum is now restricted to occupied states. The integral Eq. (43) is then solved iteratively. If the result of the first iteration is inserted into Eq. (5), we obtain the independent particle approximation to $\alpha(\omega)$. The photoabsorption cross section is then calculated from Eq. (6).

10. APPLICATION OF LINEAR RESPONSE THEORY TO METAL CLUSTERS: SPHERICAL JELLIUM MODEL

The linear response theory can be applied to study the response of atomic clusters to a laser pulse of wavelength in the optical range. The TDLDA, in conjunction with the jellium model (see Section 3), follows the Mie result of Eq. (8) in a qualitative way, shown schematically in Fig. 2 for the case of sodium clusters [6]. The dipole absorption cross section of spherical sodium clusters usually exhibits a dominant peak, which exhausts 75–90% of the dipole sum rule and is shifted by 10–20% with respect to the Mie frequency. The centroid of the strength distribution tends toward the Mie resonance in the limit of a large metallic sphere. Its red shift in finite clusters is a quantum mechanical, finite size effect closely related to the spill-out of the electrons beyond the edge of the positive jellium background.

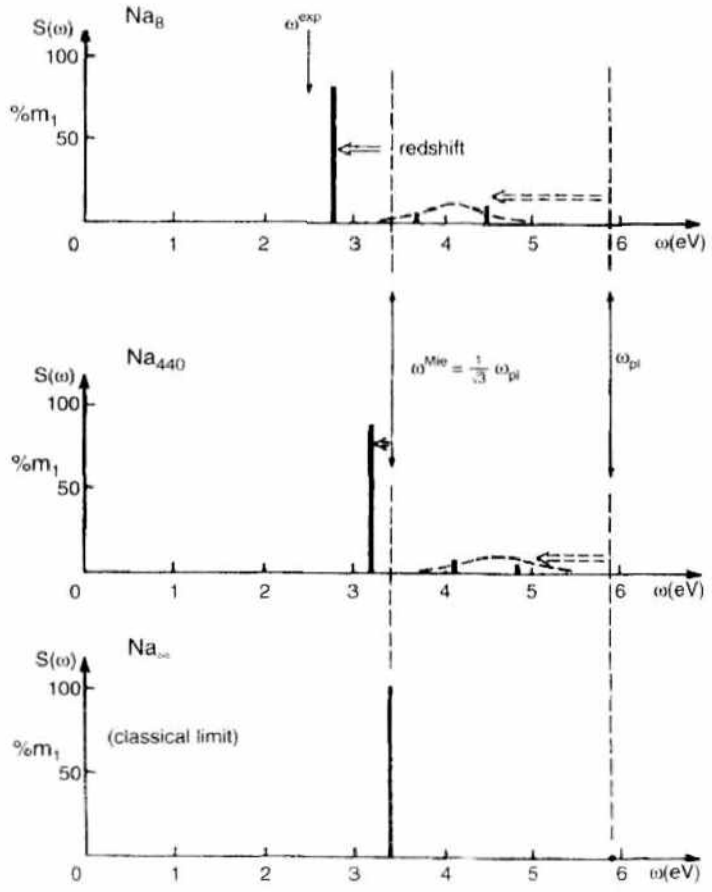


Figure 2. Schematic representation of the collective dipole spectrum of sodium clusters obtained in linear response theory. The strength function $S(\omega)$ is plotted as the percentage of the total dipole strength m_1, normalized to 100% [see Eqs. (7) and (13)]. The spectrum in the lowest panel represents the classical limit for a large metallic sphere, where all the strength is concentrated in the surface plasmon of frequency ω_{Mie} and the volume plasmon (of frequency ω_{pl}) has 0 strength. For finite clusters the surface plasmon is red-shifted and its missing strength is distributed over the reminder of the strongly fragmented volume plasmon. Reprinted with permission from [6], M. Brack, *Rev. Mod. Phys.* 65, 677 (1993). © 1993, American Physical Society.

However, about 10–25% of the dipole strength is typically found at higher energies and can be interpreted as a reminescence of a strongly fragmented volume plasmon. Often, the dominant peak is fragmented into two or more lines. For spherical clusters, this can be attributed to the interference of specific particle-hole excitations (or more complicated ones) with the predominant collective mode. This fragmentation may be compared to the Landau damping in a solid.

Before presenting specific applications, it is convenient to recall the main features of the electronic structure of clusters of the alkali metals. The population in the typical experiments in which the clusters form by aggregation of atoms in a supersaturated vapor shows magic sizes corresponding to a number of atoms $N = 2, 8, 18, 20, 34, 40 \ldots$ [48]. These magic numbers are explained by the formation of electronic shells in the common self-consistent potential of the cluster, which to a good approximation can be considered as a spherical droplet. The effective potential can be modeled by using the spherical jellium model. Then, as the cluster grows in size, the electronic shells become filled in the order $1s, 1p, 1d, 2s, 1f, 2p \ldots$. Clusters with filled electronic shells show an energy gap between the highest occupied and the lowest unoccupied molecular orbitals (HOMO–LUMO gap) and are evidently more stable than other clusters, giving rise to the observed magic numbers.

Most of the specific calculations to be reviewed next, and up to and including Section 16, use the Green's function method. However, when other techniques, like the matrix eigenvalue method, are used, this will be explicitly indicated. The results of a TDLDA calculation [26, 49] of the photoabsorption spectrum for Na_{20} within the spherical jellium model are shown in Fig. 3. The dotted curve is the spectrum obtained using the independent-particle susceptibility χ_0 in Eqs. (4) or (33), and the continuous line is the result for the interacting susceptibility χ. The ground-state electronic configuration of Na_{20} in the spherical jellium model is $1s^2 1p^6 1d^{10} 2s^2$, where the superscripts indicate the occupation of the electronic shells. Above these occupied subshells there are other unoccupied ones $1f, 2p, 1g, 2d, 3s \ldots$, and the peaks in the noninteracting spectrum represent allowed particle-hole excitations in which one electron is promoted from an occupied level to an unoccupied one. The transitions have been artificially broadened to simulate the effect of temperature (a full discussion of temperature effects is presented in Section 17). When electron–electron interactions are switched on (using the fully interacting χ), some particle-hole transitions are shifted in energy, just as Eq. (58) indicates, and other excitations lose their individual identities, merging into a collective resonance.

Compared to experiment, the LDA calculations for the spherical jellium model [6, 49] yield an insufficient red shift of the collective Mie resonance. For instance, Yannouleas et al. [50] predicted the plasmon position of Na_8 at 2.82 eV, to be compared with the peak experimentally observed at 2.53 eV [51]. This is also evident for Na_{20} in Fig. 4, where the calculated resonance lines [26] have been broadened to simulate finite temperature effects

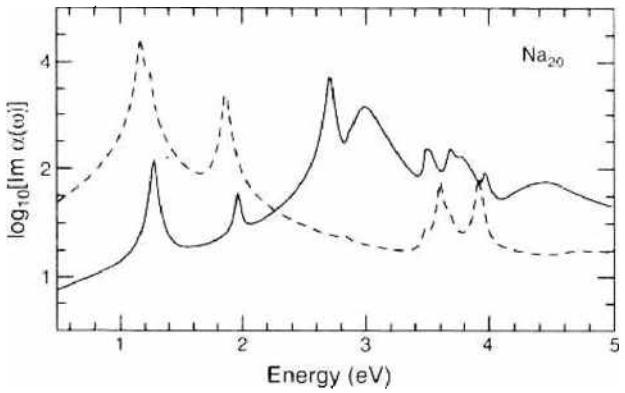


Figure 3. Imaginary part of the dynamical polarizability (per electron) of Na_{20}. The calculations employed the spherical jellium model and the time-dependent local density approximation. Dotted and continuous curves correspond to using the noninteracting (χ_0) and fully interacting (χ) dynamic susceptibilities, respectively. Transitions are broadened to simulate temperature effects. Reprinted with permission from [26], A. Rubio, J. A. Alonso, X. Blase, and S. G. Louie, *Int. J. Mod. Phys. B* 11, 2727 (1997). © 1997, World Scientific Publishing Company.

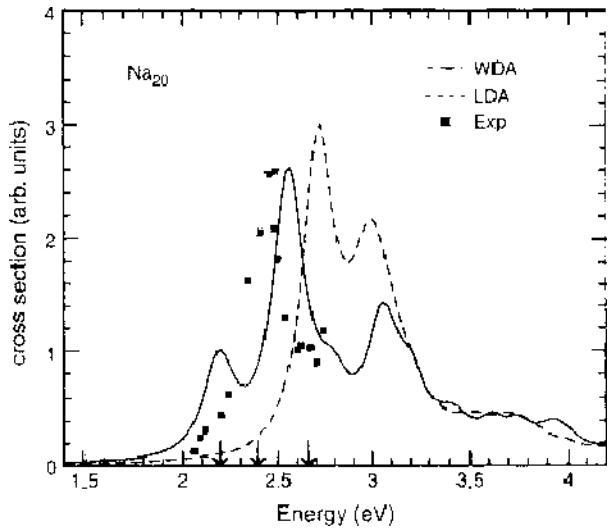


Figure 4. Calculated (local density approximation [LDA], weighted density approximation [WDA]) [26, 49] and experimental [53] photoabsorption cross sections (per electron) of Na_{20} in the spherical jellium model. Arrows mark the positions of observed peaks [56]. Reprinted with permission from [26], A. Rubio, J. A. Alonso, X. Blase, and S. G. Louie, *Int. J. Mod. Phys. B* 11, 2727 (1997), © 1997, World Scientific Publishing Company.

[6, 52]. The form of the LDA spectrum is similar to the experimental one [53], but the main LDA peak is displaced 0.3 eV to higher energy. The physical process underlying the fragmentation of the spectrum of Na_{20} is, however, not well described by the LDA. In the LDA calculation, fragmentation occurs because of the proximity to the plasmon line of a particle-hole excitation ($2s \rightarrow 3p$) with energy $\hbar\omega = 2.8$ eV. In addition, the $3p$ subshell is practically degenerate with the vacuum level, and the fragmented line is broadened by the proximity of transitions from the $2s$ states to scattering states in the energy continuum. In contrast, the experimental ionization threshold lies at an energy of 3.76 eV, more than 1 eV higher than the main plasmon peak. A calculation using a spherical jellium with a smoothed surface leads to an improvement of the position of the dipole resonance [54].

Replacing the LDA by a nonlocal description of exchange and correlation improves the results. Calculations [49] using the nonlocal WDA [55] (the continuous line in Fig. 4) shift the plasmon resonance to lower energies and place the main peak at 2.56 eV. This effect arises from a better description of the asymptotic (large r) behavior of the XC potential V_{xc} and from the improvement of the local field correction, or kernel K_{xc}, with respect to the LDA. In the WDA, the XC hole around an electron at position $\mathbf{r}_1$

$$n_{xc}(\mathbf{r}_1, \mathbf{r}_2) = n(\mathbf{r}_2)[g_{xc}(\mathbf{r}_1, \mathbf{r}_2) - 1] \tag{69}$$

presents two improvements with respect to the LDA. One is that the prefactor $n(\mathbf{r}_2)$ is not approximated by the local density $n(\mathbf{r}_1)$. In the LDA, the pair correlation function $g_{xc}(\mathbf{r}_1, \mathbf{r}_2)$ is replaced by $g_{xc}^h[|\mathbf{r}_1 - \mathbf{r}_2|; n(\mathbf{r}_1)]$; that is, the pair-correlation function for a homogeneous (h) electron gas with density n equal to the local density $n(\mathbf{r}_1)$. In the WDA, the functional form $g_{xc}^h(|\mathbf{r}_1 - \mathbf{r}_2|; n)$ for a homogeneous electron gas is preserved, but the effective density $n = \bar{n}(\mathbf{r}_1)$ is fixed at each point $\mathbf{r}_1$ by requiring the fulfillment of a sum rule stating that the integral of $n_{xc}(\mathbf{r}_1, \mathbf{r}_2)$ over the variable $\mathbf{r}_2$ equals the exact value of -1. This means that the effect of the XC hole is equivalent to removing a net charge of one electron from the neighborhood of the reference electron. The WDA prescription leads to an improved potential V_{xc}^{WDA} and kernel K_{xc}^{WDA}. Returning to Na_{20}, the more accurate V_{xc}^{WDA} also leads to an improved value of the ionization threshold, 3.27 eV, which does not interact with the plasmon because of the large separation between those two features. Instead, a shoulder develops at $\approx$2.7–2.8 eV and a pronounced secondary peak appears on the low-energy side of the resonance, at $\approx$2.2 eV, with both features being the result of the interaction of the plasmon with particle-hole excitations. Three resonances observed [56] at $\approx$2.19, 2.41, and 2.76 eV support the predicted WDA spectrum in the region below 2.9 eV. Calculations

including self-interaction corrections (SICS) [57, 58] also improve the results with respect to the LDA. SIC corrects for the spurious self-interaction of an electron with itself and improves the asymptotic behavior of V_{xc}. Saito et al. [57] introduced SIC in the kernel $K_{xc}[n](\mathbf{r},\mathbf{r}')$ by an Amaldi-type correction

$$K^{\mathrm{SIC}}(\mathbf{r}_1,\mathbf{r}_2)=\frac{N-1}{N}K^{\mathrm{LDA}}(\mathbf{r}_1,\mathbf{r}_2) \tag{70}$$

where N is the number of electrons. In contrast, Pacheco and Ekardt [58] followed the standard SIC scheme.

Interesting effects are expected when the plasmon excitation energy is close to the ionization threshold. This occurs for large-cluster anions [49, 59]. For small negatively charged clusters, the electron detachment threshold is low, and the energy of the collective resonance lies in the region of electronic excitations to the continuum of states, where Landau damping produces a broadening of the resonance. This is appreciated by comparing the calculated spectrum of Na_{10}^-, given in Fig. 5, with that for neutral Na_{20} in Fig. 4 (the two clusters are isoelectronic, with 20 electrons). The WDA was used in the calculations reported in Fig. 5. The electron detachment threshold of Na_{10}^- is indicated by the arrow in Fig. 5. As the cluster size increases, the plasmon approaches the region of discrete states. Then, when the detachment threshold of the negative cluster lies in the region of the plasmon resonance, electron emission becomes a decay mechanism that competes with the usual one of evaporation of atoms. The calculated photoabsorption cross sections of Na_{91} and Na_{197} are also given in Fig. 5. In those two clusters, the detachment threshold overlaps with the collective resonance. Reiners and Haberland [60] measured the photoabsorption cross section of Na_{91}^- and found a broad collective resonance centered at 2.65 eV (its width is 0.92 eV) whose decay can lead to two final channels: atom and electron emission. The calculated position of the collective resonance of Na_{91} in Fig. 5 is 2.69 eV, in good agreement with the experiment of Reiners and Haberland. Changing the net cluster charge, from anionic to neutral to cationic clusters, for a fixed total number of valence electrons has the effect of shifting slightly the plasmon to higher energies. For instance, the maximum of the resonance occurs at 2.65 eV

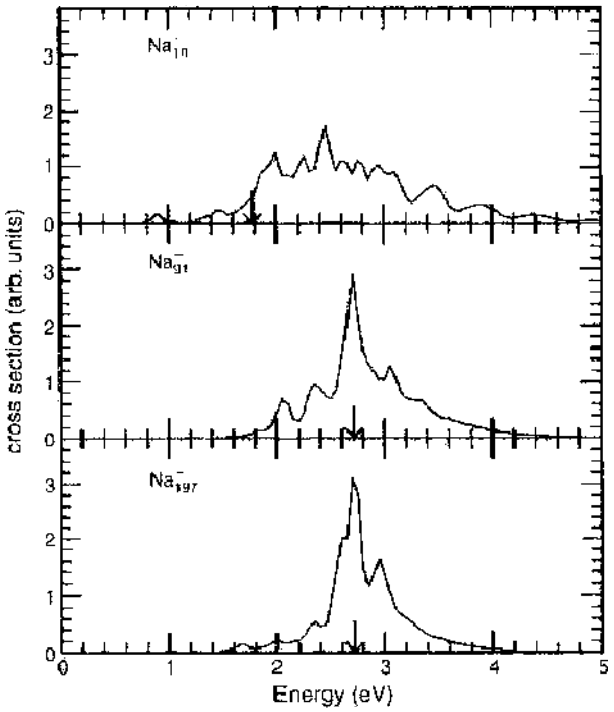


Figure 5. Calculated photoabsorption cross sections, per electron, of Na_N^- cluster anions. The clusters are described by the spherical jellium model. The nonlocal WDA approximation was used for exchange and correlation. The arrows mark the binding energy of the highest-occupied molecular orbital. Reprinted with permission from [26], A. Rubio, J. A. Alonso, X. Blase, and S. G. Louie, *Int. J. Mod. Phys. B* **11**, 2727 (1997) © 1997, World Scientific Publishing Company.

and 2.77 eV for Na_{91}^{+} and Na_{93}^{+}, respectively, and the corresponding width decreases from 0.92 eV to 0.51 eV.

11. SHAPE DEFORMATIONS: ROLE IN THE ABSORPTION SPECTRUM

A splitting of the collective resonance is observed in clusters with open electronic shells, which is a consequence of the static spheroidal deformations of the cluster shape. A double peak in the photoabsorption cross section has been observed for K and Na clusters [4, 61, 62] and for Ag clusters in the region $10 \leq N \leq 16$ [63]. The two modes correspond to excitations along the main axis of the spheroid and perpendicular to that axis, respectively. Figure 6 shows the experimental results of Borggreen et al. [62] for cationic Na clusters. These results reveal the systematics of cluster shapes after the spherical clusters with 8 and 20 electrons, respectively. The trend in the shapes goes spherical → prolate → oblate → spherical. This systematics is reproduced by total energy calculations within a spheroidal jellium model [64], and the splitting of the collective resonance is obtained by the TDLDA applied to the deformed clusters [65]. In larger clusters, it is difficult to disentangle the effects originating from static shape deformations from those resulting from the fragmentation mechanism discussed above.

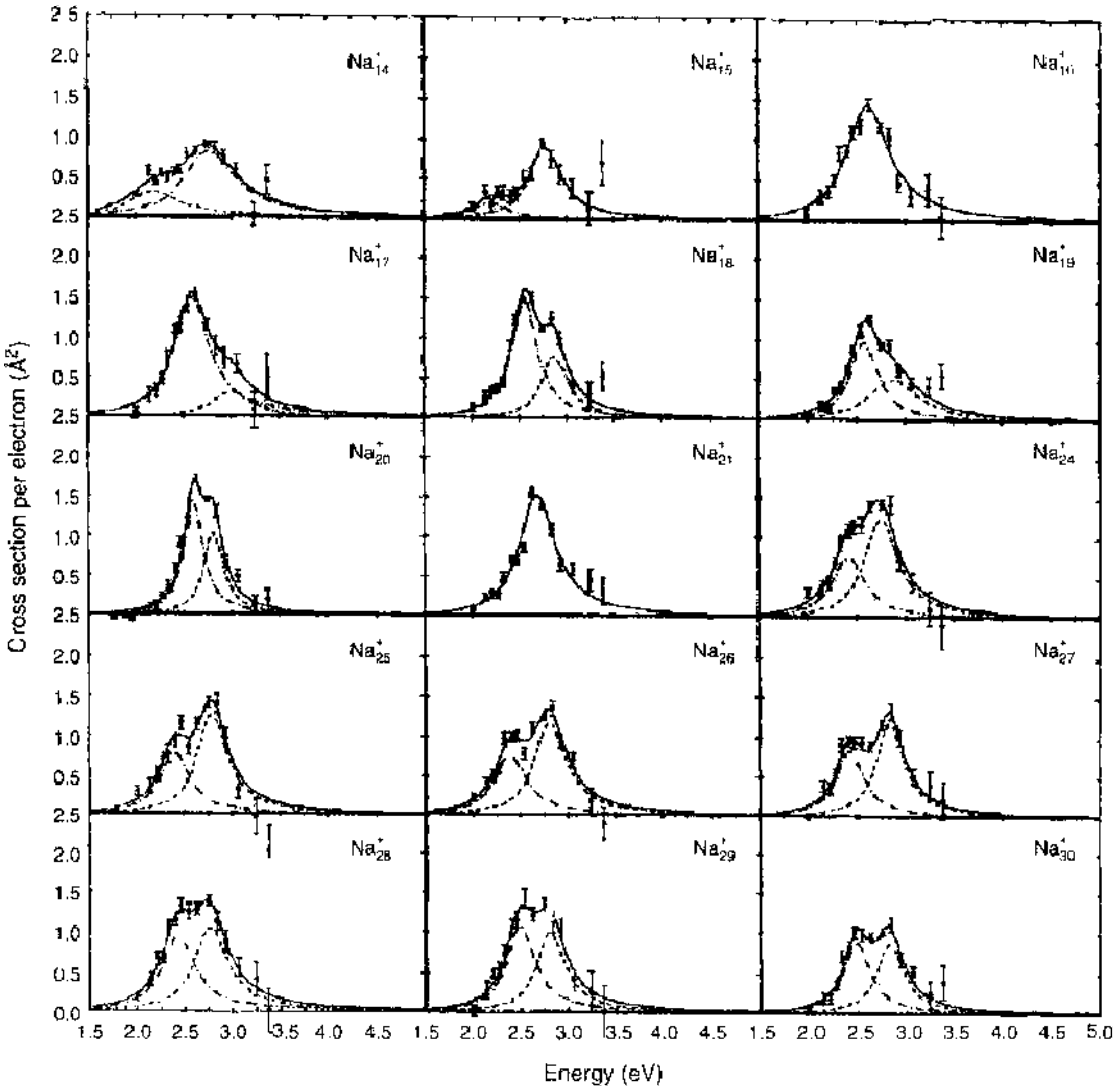


Figure 6. Experimental photoabsorption spectra of Na_N^+ cations. The curves are Lorentzian functions fitted to the data. Adapted with permission from [62], J. Borggreen, P. Choudhury, N. Kebaili, L. Lundsberg-Nielsen, K. Lützenkirchen, M. B. Nielsen, J. Pedersen, and H. D. Rasmussen, *Phys. Rev. B* 48, 17507 (1993). © 1993, American Physical Society.

A splitting of the dipole resonance into three peaks has been observed in some Na clusters [4, 62]. This is interpreted as corresponding to collective vibrations of the valence electrons in the directions of the principal axis of a triaxially deformed cluster and has motivated the extension of the jellium model to fully triaxial shapes [66–68]. The triaxial deformations of the uniform background can be classified in terms of the Hill–Wheeler coordinates β and γ [69]. β describes the overall quadrupole deformation; $\gamma = 0°$, 120°, and 240° describe prolate deformations; and $\gamma = 60°$, 180°, 300° oblate ones—all other values of γ give truly triaxial shapes. Let us consider Na_{12} and Na_{14}. The potential energy surfaces of those two clusters in the triaxially deformed jellium model have been calculated by Lauritsch et al. [66]. In addition to the shape deformation of the positive background, the jellium was also allowed to have a diffuse density profile at the surface [54]. The ground state of Na_{12} is triaxial, with deformation parameters $\beta = 0.54$, $\gamma = 15°$. This structure is energetically well separated from competing prolate and oblate configurations. Na_{14} is characterized by two axially symmetric isomers, prolate and oblate, respectively, which are almost degenerate in energy. The oblate minimum is rather soft in the γ-direction, whereas the prolate minimum predicts stiffer γ-vibrations. The pronounced shape isomerism found for both clusters bears some resemblance to that found by fully microscopic quantum chemical [70] and *ab initio* DFT calculations [71]. The resonance energies of the collective dipole excitations were obtained from the approximate expression

$$\omega_i^2 = \frac{1}{N m_e} \int n(\mathbf{r}) \frac{\partial^2}{\partial r_i^2} V_{\text{ext}}(\mathbf{r}) d^3 r \tag{71}$$

obtained from the RPA sum rules [6]. In this equation, V_{ext} is the electrostatic potential of the jellium background, m_e is the electronic mass, and i runs over the spatial directions; that is, $r_i = \{x, y, z\}$ for the triaxial clusters and $r_i = \{r, z\}$ for axial ones. The resonance energies, calculated for the ground state of Na_{12} and for the two degenerate minima of Na_{14} are given in Table 4. Three different energies are obtained for Na_{12}, reflecting its triaxial shape. The three energies are in qualitative agreement with the experimental peaks [4, 62], although the calculated energies are 10–15% too high because of the simple sum rule approximation used. Each of the two competing axial isomers of Na_{14} is characterized by a double-peak structure in which $\hbar\omega_1$ has double weight compared to $\hbar\omega_2$. The actual strength distribution will be an incoherent superposition of the two isomeric minima.

Kohl et al. [72] extended the calculations to a larger set of Na_N clusters ($N = 2$–20). They have confirmed the results of the spheroidal jellium model: prolate clusters after the magic numbers $N = 2$ and $N = 8$, and oblate ones before $N = 8$ and $N = 20$. However, a transition region formed by triaxial shapes was found separating the prolate and oblate regions. This region is very small between $N = 2$ and $N = 8$, containing only the cluster Na_5, and is larger in the region between $N = 8$ and $N = 20$. The triaxial minimum is well developed in Na_5, but the others are extremely soft, such that thermal fluctuations will easily wash out the triaxial signatures in the dipole resonance. For cationic Na_N^+ clusters, Kasperl et al. [67] have concluded that the signal of triaxiality on the resonance energies is faint for clusters larger than Na_9^+.

The mechanisms responsible for the width of the plasmon resonance have been investigated [73] for spherical clusters (Na_9^+, Na_{21}^+, Na_{41}^+) in the framework of the structurally averaged jellium model [74]. In this model, the effects of the ionic structure are added in an averaged manner: first as an additional potential on the electrons, second as an average Madelung energy in the volume, and third as an ionic surface energy. The two leading

Table 4. Calculated surface dipole plasmon energies (in electron volts) for the triaxial ground state of Na_{12} and for the oblate (OE) and prolate (PE) isomers of Na_{14}, obtained from the random phase approximation sum rule of Eq. (71) [66].

Na_{12} Triaxial			Na_{14} Oblate		Na_{14} Prolate	
$\hbar\omega_1$	$\hbar\omega_1$	$\hbar\omega_2$	$\hbar\omega_1$	$\hbar\omega_2$	$\hbar\omega_1$	$\hbar\omega_2$
2.857	3.238	3.313	2.531	3.401	3.102	2.313

mechanisms for the line broadening are the fragmentation of the resonance into nearby particle-hole transitions and splitting through thermal quadrupole fluctuations [52].

12. APPROXIMATE ACCOUNT OF THE IONIC STRUCTURE

The experimental measurements of the photoabsorption cross section of the closed shell cluster Na_{21}^+ show a plasmon resonance in the region 2.65–2.74 eV; more specifically, at 2.65 eV in Ref. [61], at 2.68 eV in Ref. [62], and at 2.74 eV in Ref. [75], with an averaged value of 2.69 eV. This cluster, having 20 electrons, is isoelectronic with Na_{20}, but the plasmon resonance occurs at a slightly higher energy in Na_{21}^+ because of the stronger confining potential. A TDLDA calculation for Na_{21}^+ within the jellium model, shown in the upper panel of Fig. 7 [76], gives the plasmon energy at 2.95 eV, overestimating its energy by 0.2–0.3 eV. Introducing in the calculation nonlocal corrections to exchange and correlation improves the position of the resonance (2.63 eV) but also leads to the appearance of a separated fragmentation peak. This fragmentation of the plasmon peak has not been observed in the experiments, although a shoulder was detected by Borgreen et al. [62] and by Reiners et al. [75] on the blue side of the peak.

The photoabsorption cross sections obtained with the spherically averaged pseudopotential model (SAPS) are plotted in the bottom panel of the same figure. The SAPS model [77, 78] provides an approximate description of the cluster structure, intermediate between the simple uniform background model and the full treatment of the geometry of the ionic skeleton. Consider a cluster with the ions at positions $\{\mathbf{R}_j\}$ $j=1,\ldots,N$. If each ion is replaced by a local pseudopotential, $v_{ps}(|\mathbf{r}-\mathbf{R}_j|)$, then the total external potential seen by the valence electron cloud is given by

$$V_{ps}(\mathbf{r})=\sum_{j=1}^{N} v_{ps}(|\mathbf{r}-\mathbf{R}_j|) \tag{72}$$

Experience with the spherical jellium model indicates that for clusters with a nearly spherical shape this external potential can be replaced by its spherical average about the cluster center

$$V_{ps}(\mathbf{r}) \rightarrow V^{SAPS}(r) \tag{73}$$

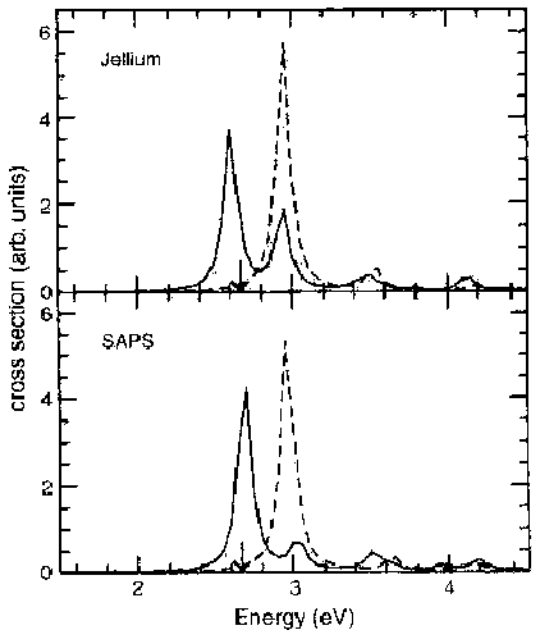


Figure 7. Calculated photoabsorption cross section of Na_{21}^+ in the jellium (upper panel) and SAPS (lower panel) models. Dashed and continuous curves correspond to local density approximation and nonlocal weighted density approximation treatments of exchange and correlation. The arrows indicate the experimental position of the plasmon resonance. Reprinted with permission from [26], A. Rubio, J. A. Alonso, X. Blase, and S. G. Louie, *Int. J. Mod. Phys. B* 11, 2727 (1997). © 1997, World Scientific Publishing Company.

This simplification greatly reduces the computational effort of calculating the electronic levels because the electrons now move in a spherically symmetric potential well. However the SAPS model goes beyond the Spherical Jellium model, as the radial structure of the cluster becomes reflected in the SAPS potential, and the ion–ion interaction is calculated for the true three-dimensional arrangement of the ions [78]. One can set limits of validity to the SAPS model. The cluster cannot be too small, because small clusters substantially deform away from the spherical shape. However, very large clusters have a tendency to form planar surface facets. The range of intermediate sizes is, then, the most appropriate one. Returning to Na_{41}^{+}, the structure of this cluster is rather spherical (O_h symmetry in a recent calculation [79]), so the use of the SAPS model is justified. The results are given in the lower panel of Fig. 7. The LDA again overestimates the experimental plasmon energy, but the nonlocal WDA calculation places the plasmon at 2.70 eV; this time the strong fragmentation has disappeared, and only small features remain at 3.0 eV and 3.5 eV, which correlate with the observed shoulders [62, 75]. The photoabsorption spectrum of Na_{41}^{+} has also been calculated [80], using the SAPS model and the WDA for exchange and correlation. The position and shape of the plasmon resonance is again in good agreement with experiment.

For Cs_8, a fragmentation peak has been observed near the surface plasmon resonance at 1.55 eV [81, 82]. The surface plasmon is obtained at 1.79 eV using the Spherical Jellium model. The SAPS model predicts a square antiprism as the ground-state geometry of Cs_8 [83]. By adjusting the cluster radius and the core radius of a pseudopotential developed by Manninen [84], the SAPS model leads to a good fit to the experimental plasmon peak and its fragmentation [81, 83]. A good fit was not obtained for the geometries tested other than the square antiprism.

In an extension of the SAPS model, Schöne et al. [85, 86] expand the total ionic pseudopotential of Eq. (72) in spherical harmonics about the center of the cluster

$$V_{ps}(\mathbf{r}) = V_0(r) + \sum_{l=1}^{\infty} \sum_{m=-l}^{l} V_{l,m}(r) Y_l^m(\theta, \varphi) \tag{74}$$

The first term V_0 is just the monopole, spherical part of the total ionic potential. This is the SAPS potential. The other part can be included perturbatively up to the second order on top of a SAPS calculation. Similar ideas, based on a perturbative introduction of geometrical effects beyond SAPS, were applied by Rubio et al. [87] to the fullerene molecule.

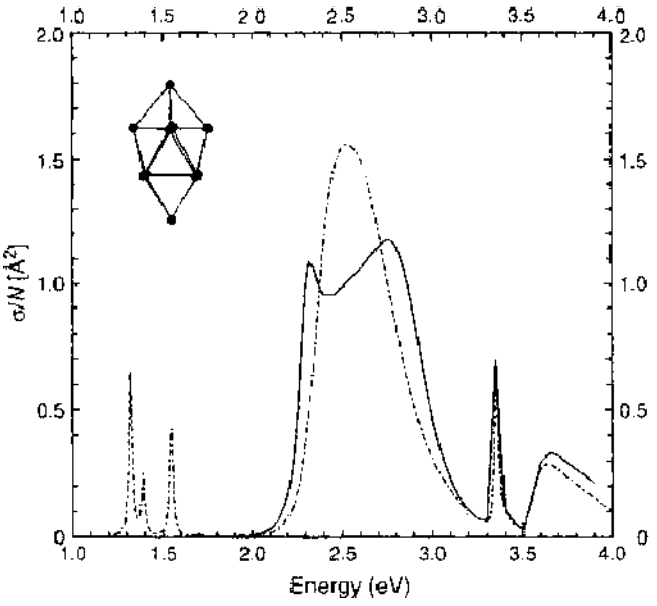


Figure 8. Photoabsorption cross section per electron for Na_{10} in the ground state configuration (the structure is given in the inset). The dashed line is from the SAPS model. The full symmetry is taken perturbatively in the calculation for the continuous line. Adapted with permission from [86]: W. D. Schöne, W. Ekardt, and J. M. Pacheco, *Phys. Rev. B* 50, 11079 (1994). © 1994, American Physical Society.

The perturbative theory can also be developed within the TDLDA formalism, and the method has been applied [85, 86] to calculate the optical spectrum of the closed-shell (Na_8, Na_8^+, Na_{18}) and open-shell clusters (Na_6, Na_{10}). For the closed-shell clusters, the perturbative calculations predict a photoabsorption spectrum in agreement with that of the SAPS model and justify the validity of this model for magic clusters. For small open-shell clusters there are differences, as expected. The geometry of Na_6 is a planar-like pentagonal pyramid (C_{5v} symmetry) [71]. The SAPS spectrum [85] gives a main absorption line at 2.3 eV. In contrast, the perturbative calculation for the C_{5v} geometry gives several peaks in the energy range 2.0–2.8 eV, in qualitative agreement with the experimental results of Wang et al. [88], and also with *ab initio* quantum chemical calculations [89]. Surprisingly, the spectrum of Na_{10}, given in Fig. 8, does not resemble that of the spheroidal jellium model [64, 65], and this appears to indicate a pronounced effect of the pseudopotentials beyond the jellium approximation. In summary, quantitative agreement with experiment was obtained by the perturbative method, but this method becomes difficult for clusters with low symmetry.

13. MIXED METAL CLUSTERS

Experiments with metal vapors [90, 91] have shown that some changes occur in the magic numbers of alkali metal clusters when these are doped with divalent impurities: Ba, Sr, Eu, Ca, Yb, Mg, and Zn dopants in Na clusters, and Mg, Hg, and Zn dopants in K clusters. These authors concentrated on the case of clusters containing a single dopant atom. For some impurities, a new magic cluster corresponding to 10 valence electrons is found; for instance, Na_8Mg, K_8Mg, K_8Hg, Na_8Zn, and K_8Zn, and at the same time the magic number corresponding to 18 valence electrons vanished. The single coordinate $\Delta n_+ = n_+^0(\text{impurity}) - n_+^0(\text{host})$; that is, the difference between the average conduction electron densities in the pure host and those in impurity metals, allows us to separate the doped clusters into two subsets [92]. Values of Δn_+ roughly higher than 0.008 e/a.u.^{3} induce changes in the magic numbers. In contrast, there are no changes for $\Delta n_+ < 0.008$ e/a.u.^{3}. The success of this coordinate indicates a simple extension of the Spherical Jellium model [92]. The impurity is embedded in a host cluster with a background density equal to $n_+^0(\text{host})$. The impurity, placed at the center of the host cluster, is then characterized by a positive background with density $n_+^0(\text{impurity})$ and radius R_{imp}. The radius of the doped cluster is determined by R_{imp} and the number of host atoms. Calculations for this jellium-on-jellium model show that the binding energy of the electrons with s-character ($l=0$) increases in comparison to the undoped cluster because the presence of the divalent impurity induces a more attractive effective potential in the central region of the cluster, whereas the binding energies of electrons with $l \geq 1$ change very little. The main effect on the electronic structure is that the magnitude of the gap between the $2s$ and $1d$ shells decreases with increasing Δn_+. When Δn_+ becomes roughly 0.008 e/a.u.^{3}, a reordering of these two subshells occurs and the filling order changes from the original $1s1p1d2s$ to the new one $1s1p2s1d$. Clusters with 10 valence electrons then have a closed-shell configuration $1s^21p^62s^2$, which accounts for the experimental observation of the new magic number. The magic number corresponding to 20 electrons, corresponding to the electronic configuration $1s^21p^62s^21d^{10}$, is still present, but evidently there is no shell-closing for 18 electrons.

The optical response of doped clusters has been studied by Yannouleas et al. using the jellium-on-jellium model [93]. The calculated 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, which carries 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 a smaller amount of strength than the lower-energy component at 2.63 eV. The fragmentation of the plasmon peak is caused by its degeneracy with the $2s \rightarrow 2p$ and $1p \rightarrow 3s$ particle-hole transitions. This result was obtained for an electron density parameter $r_s(\text{Zn}) = 1.15$ a.u. [see Eq. (20)], a value substantially smaller than the typical value of $r_s = 2.31$ a.u. for bulk Zn. The smaller value of $r_s(\text{Zn})$ produces a strong attractive potential at the impurity site, which leads to the downwards shift of those transitions required for the degeneracy with the plasmon to develop. The spectrum of pure Na_8 contains a single line at 2.53 eV, so the effect of the Zn impurity 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. Balbás et al. [94] have calculated the optical response of Na_8Zn using the SAPS model, with the Zn atom at the center of the cluster. Two peaks are obtained at 2.68 and 2.9 eV. The second peak was interpreted as a fragmentation of the surface plasmon induced by its proximity to the ionization threshold, an interpretation that differs from that proposed by Yannouleas et al. [93].

As another example of the influence of impurities on the optical reponse Yannouleas et al. considered the group formed by $K_{19}Rb$, K_{20}, and $K_{19}Na$, all of them having 20 valence electrons. The calculated optical spectrum of K_{20}, shown in Fig. 9, exhibits a split plasmon because of a degeneracy between the collective plasmon and a $2s \rightarrow 3p$ particle-hole transition. In the case of $K_{19}Na$, the Na impurity, with a smaller r_s than the host, shifts the $2s$ level downward, so the energy of the $2s \rightarrow 3p$ transition increases. At the same time, the energy of the plasmon remains unchanged, and the consequence is that 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.

The absorption spectrum of $Cs_{10}O$ shows two peaks at 1.39 and 1.54 eV [81, 82]. One can expect the $2p$ shell of the Oxygen atom to be filled by two electrons provided by the Cs atoms. In this way, Yannouleas and Broglia [95] have treated $Cs_{10}O$ as a system with eight active electrons. These authors employed a modified jellium potential in which the central part of the cluster was made less attractive to simulate the repulsive effect of the oxygen anion O^{2-}. The fragmentation of the plasmon was reproduced, although a difference of $\approx$0.2 eV remained between the positions of the theoretical and experimental peaks. Calculations using the SAPS model show that the absorption spectrum is sensitive to the radius of the cluster [83], and this can be adjusted to obtain a good description of the spectrum. Also, inclusion of the p-electrons of the oxygen atom is required, as particle-hole transitions from the p shell to the unoccupied $4s$ shell contribute to the fragmentation of the collective resonance. The cluster geometry used in the calculations was the one predicted by the SAPS model, with the oxygen atom at the center of a square pyramid and each of the five faces capped by one Cs atom.

A common phenomenon in binary metallic alloys is the preferential segregation of one of the components at the surface [96, 97]. Surface segregation affects also the collective electronic response of clusters. The ground-state structure of $Na_{31}K_{31}$ obtained with the SAPS model is composed of three layers surrounding a central Na atom. All the potassium atoms are on the surface, and the two inner layers have 11 and eight Na atoms, respectively. The calculated photoabsorption spectrum [98] shows a collective resonance peak at 2.1 eV. The tail of the resonance extends up to 3 eV and concentrates a sizable amount of oscillator

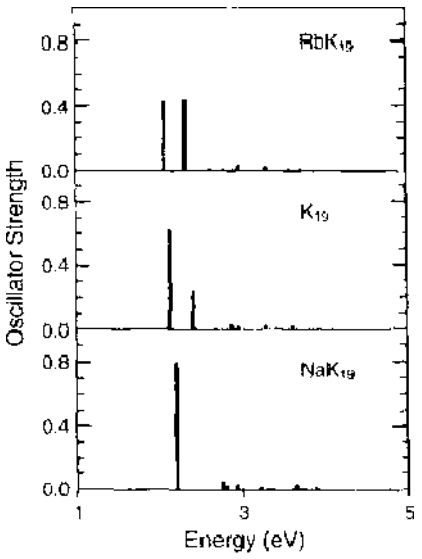


Figure 9. Calculated photoabsorption cross section of RbK_{19}, K_{20}, and NaK_{19}. Reprinted with permission from [93], C. Yannouleas, P. Jena, and S. N. Khanna, *Phys. Rev. B* 46, 9751 (1992). © 1992, American Physical Society.

strength. The position of the collective resonance is closer to the corresponding resonance of pure K clusters, calculated with the same method, compared to pure Na clusters, and this is a manifestation of the fact that the cluster surface, whose electron density contributes most to the collective excitation, is formed by K atoms. A calculation of the photoabsorption spectrum for other isomers allows us to analyze the sensitivity of the spectrum to the structural features. When the positions of some Na and K atoms are simply exchanged, preserving other features of the cluster geometry, a shift of the resonance peak to higher energies occurs as Na replaces K on the surface. More drastic variations of the structure produce pronounced changes in the spectrum, like the broadening and fragmentation of the plasmon peak. The optical response of mixed Li–Na clusters has also been studied [94]. In summary, the shape of the photoabsorption spectrum is sensitive to the cluster geometry and to the degree of segregation of one component to the surface, so, a comparison between measured and calculated spectra may be useful to elucidate segregation effects in mixed clusters.

14. SIMPLE MODEL FOR THE PHOTOABSORPTION SPECTRUM OF FULLERENES

Extensions of the jellium model have been applied to more complex systems, like carbon fullerenes [87]. Experimental studies [99–103] have shown the presence of two collective excitations in C_{60}, a broad one in the region 15–25 eV, associated with the σ electrons of the cage, and the other around 6 eV, which was interpreted as a collective oscillation of the π-electrons. The electrons in σ orbitals (three electrons per C atom) link neighboring atoms in the cage and provide most of the cohesion. In contrast, the π orbitals extend in and out of the cage and are more delocalized. For this reason, to describe the response of the π electrons, a model [87] has been used that simulates the C_{60} cage by a thin spherical shell of positive charge

$$n_+(r)=\begin{cases} n_1 & R-\Delta_1 \le r \le R+\Delta_1 \\ 0 & \text{otherwise} \end{cases} \tag{75}$$

with the observed radius $R=6.64$ a.u [104] and a width $2\Delta_1$ in the range 0.8–1.0 a.u., compensated by a distribution of π electrons (one π electron per C atom) self-consistently calculated. The constant n_1 is easily related to R, Δ_1 and the total number of electrons (60). The calculated plasmon resonances of C_{60} and C_{60}^+ are roughly in the right position [87, 105].

The model was extended to C_{60} coated by a layer of Na atoms [105]. In that case, the positive background was modeled as two concentric spherical thin layers, one on top of the other, with the inner one, of thickness $2\Delta_1$, representing the fullerene cage, and the outer one, of thickness Δ_2, representing the alkali coating; that is,

$$n_+(r)=\begin{cases} n_1 & R-\Delta_1 \le r \le R+\Delta_1 \\ n_2 & R+\Delta_1 \le r \le R+\Delta_1+\Delta_2 \\ 0 & \text{otherwise} \end{cases} \tag{76}$$

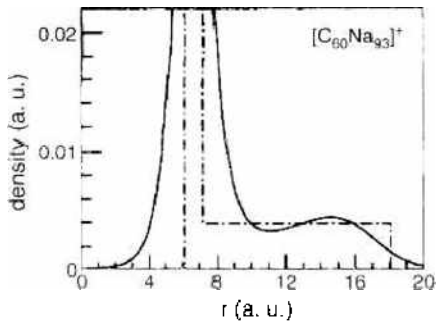


Figure 10. Electron density (continuous line) and positive background density (dashed-dotted line) as a function of distance to the cluster center in a simple continuous model for $[C_{60}Na_{93}]^+$. Adapted with permission from [105], A. Rubio, J. A. Alonso, J. M. López, and M. J. Stott, *Phys. Rev. B* 49, 17397 (1994). © 1994, American Physical Society.

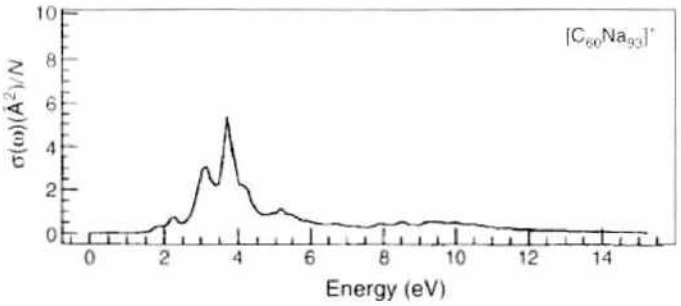


Figure 11. Calculated photoabsorption cross section, per electron, of $[C_{60}Na_{93}]^+$.

The background density n_2 of the external layer was fixed equal to its value for bulk Na metal, $n_2 = 0.004$ e/a.u.^{3}, and the thickness Δ_2 then depends on the number of coating Na atoms. The background density and the self-consistent electron density of $[C_{60}Na_{93}]^+$ are plotted in Fig. 10. The electron density shows two different decay lengths. This cluster has an outer surface that resembles a lot the surface of pure Na clusters. The development of the coating metallic layer becomes reflected in the calculated photoabsorption spectrum, shown in Fig. 11.

For small Na coverage, like in $[C_{60}Na_{13}]^+$, the interaction between the electrons originating from the Na atoms and the unoccupied levels of C_{60} is strong and the TDLDA calculation predicts a spreading of the π-plasmon of C_{60}, down to lower energies by more than 1 eV, and a reduction of its amplitude. This is mainly because of the fragmentation of the collective resonance caused by the interaction with particle-hole transitions. Then the characteristic features of the C_{60} plasmon progressively vanish as the number of coating Na atoms grows. At the same time, a new feature develops at lower energies that can be related to the surface plasmon of pure Na clusters. For $[C_{60}Na_{93}]^+$ the Na surface plasmon is well developed and the cluster responds much like a pure ionized Na cluster; the resonance has, nevertheless, a broad tail at high energies.

The fragmentation of the peak may have relevance for the interpretation of experiments by Martin and coworkers [106]. These authors have produced C_{60} clusters coated with Cs. For large coverage, say $[C_{60}Cs_N]^+$ with $N \approx 300$, the photoabsorption spectrum is very similar to that of large pure Cs_N clusters, except for having a more intense tail. As N decreases, a fragmentation of the plasmon occurs that resembles that which was obtained in the TDLDA calculations for $[C_{60}Na_N]^+$. Thus, the TDLDA based on the model of two spherical jellium slabs appears to afford a plausible explanation for the optical absorption spectrum of alkali-coated C_{60} clusters.

15. FULL ACCOUNT OF THE CLUSTER STRUCTURE

TDLDA calculations of the optical spectrum with a full account of the geometry of the cluster have also been performed. Apart from being more realistic compared to the other calculations discussed above, which retain only the overall shape characteristics, these new calculations can provide a useful method to determine the cluster geometries. Sometimes, the differences in binding energy between the ground state and some low-lying isomers obtained by *ab initio* calculations are so small as to cast doubts about the calculated lowest energy structure, and a comparison of the experimental optical spectrum and those calculated for several isomers with similar energies can help in the identification of the ground state.

A method based on a space and time representation of the response functions of extended systems has been developed that takes advantage of the rather sparse Hamiltonian matrix in a coordinate representation [107, 108]. The advantage is related to the localization range of the independent-particle susceptibility $\chi_s(\mathbf{r}, \mathbf{r}'; \omega)$ and other response functions, as localized objects are easily described in real space. In practice, the response functions of nonmetallic systems decay rapidly as $|\mathbf{r} - \mathbf{r}'| \to \infty$, so that for each $\mathbf{r}$, $\chi_s(\mathbf{r}, \mathbf{r}'; \omega)$ needs to be calculated only for $\mathbf{r}'$ inside a region of a certain radius R_{max} around $\mathbf{r}$. This is the origin of the success of recent N-linear methods (where N is the number of atoms in the unit cell) proposed to perform band-structure calculations for solids [109–113]. However, for metals and small-gap

semiconductors, the decay rate may be slow and R_{max} may span many unit cells. This problem can be solved by a mixed-space representation $\chi_{\mathbf{q}}(\mathbf{r},\mathbf{r}';\omega)$ of the response functions, where $\mathbf{r}$ and $\mathbf{r}'$ are restricted to a single cell and $\mathbf{q}$ spans the irreducible part of the Brillouin zone [108]. The method, described in Refs. [26] and [108], has been adapted to clusters by using a supercell formalism, in which the unit cell containing the cluster is periodically repeated in space, which allows us to expand the wave functions on a basis of plane waves. The volume of the cell containing the cluster has to be sufficiently large to avoid the interaction between clusters in neighbor cells; that is, a large part of the volume of the cell is empty.

The photoabsorption cross section of Na_8, calculated by this method [114], is in excellent agreement with experiment, with the calculated position of the plasmon peak at 2.55 eV and the experimental one at 2.53 eV [4, 51]. (The calculated value of the polarizability of Li_8 given in [114] is not correct. The correct value is 97 Å^3.) The TDLDA photoabsorption cross section of Li_8 is given by the continuous curve in Fig. 12. The calculated ground state structure, given in the inset, is a centered trigonal prism with an atom capping one of the lateral faces. The averaged value of the static dipole polarizability $(\alpha_{xx}+\alpha_{yy}+\alpha_{zz})/3$ is 97 Å^3, which is larger than the value of 63 Å^3 obtained in the spherical jellium model [114]. From classical arguments, a larger polarizability corresponds to a lower resonance frequency, so a redshift of the resonance with respect to the jellium value of 3.5 eV should be expected. Indeed, the effect of explicitly accounting for the cluster structure produces a redshift of 1 eV, which leads to very good agreement between the calculated resonance at 2.45 eV and the experimental value of 2.5 eV [115] (compare the position of the peak of the continuous curve with that of the arrow in Fig. 12). The nearly isotropic polarizability tensor explains the presence of a single resonance. The red shift can be tracked down to an increase of the electron effective mass, an effect of the Li ionic pseudopotential.

The computed structures of silicon clusters [114], given in the insets in Fig. 13, agree with other calculations [116, 117] and with Raman experiments [118]. A planar rhombus (D_{2h} symmetry) is obtained for Si_4, and two nearly isoenergetic structures for Si_6: a distorted octahedron (D_{4h}), and an edge-capped trigonal bypiramid (C_{2v}), with the octahedron being slightly more stable by ≈0.02 eV. Could structural isomers be distinguished from their absorption spectra? This question is analyzed in Fig. 13, where the calculated single-photon-averaged absorption profiles have been plotted for two different isomers of each of these two clusters, Si_4 and Si_6 (the second isomer of Si_4 is the tetrahedron). All the spectra have been broadened by 0.1 eV to simulate finite temperature effects [52]. The calculated spectra of different isomers are sufficiently different, and the conclusion is that the optical spectrum contains structural information, so the comparison between calculated

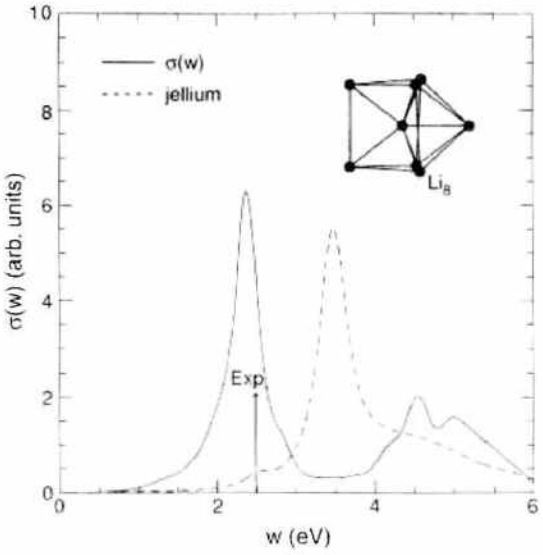


Figure 12. Photoabsorption cross section of Li_8. The dashed line is for the Jellium model and the continuous line is obtained with full account of the geometrical structure. The centroid of the experimental resonance is indicated by the arrow.

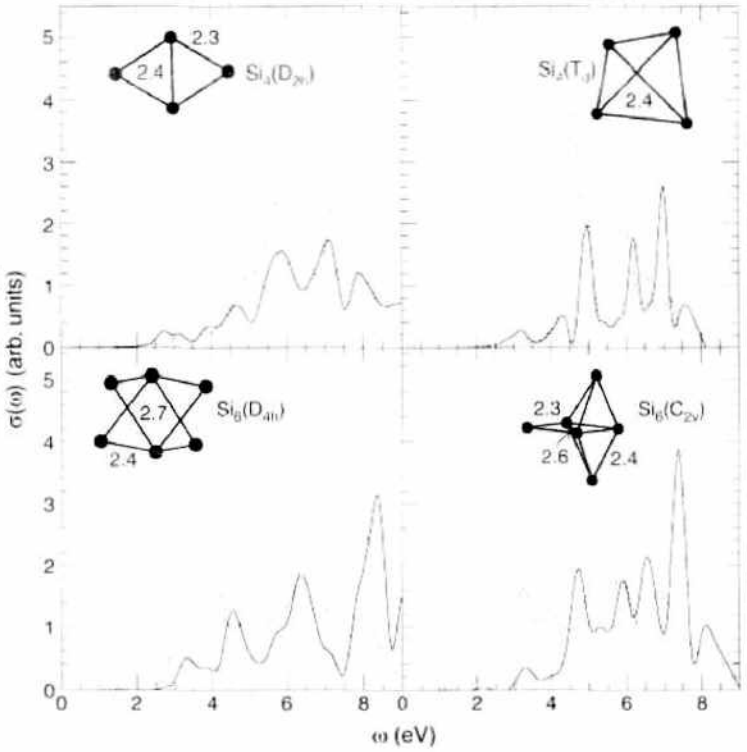


Figure 13. Averaged photoabsorption cross section of two different isomers of the Si_4 and Si_6 clusters, shown by the continuous line. The independent-particle spectra (dotted curves) have been included for comparison. All the spectra have been broadened by 0.1 eV. The cluster geometries are included in the insets, and some bondlengths are given in Angstroms. Reprinted with permission from [114]. A. Rubio, J. A. Alonso, X. Blase, L. C. Balbás, and S. G. Louie, *Phys. Rev. Lett.* 77, 247 (1996). © 1996, American Physical Society.

and measured optical spectra can provide a powerful tool to assign structures to the clusters. Even more striking differences in the spectra are obtained if each of the tensor components of the absorption cross section are considered separately instead of using the averaged component. XC effects play a minor role here as compared to the metallic clusters (the same observation is known to hold for the bulk). Surface and confinement effects are responsible for the appearance of absorption in the optical range seen in the TDLDA spectra for the two Silicon clusters. In fact, this effect could be related to the luminescence of porous silicon [119], showing that small clusters could present different and very interesting optical properties as compared to their bulk counterpart.

Other authors have been accounted for the cluster structure in the TDLDA, using different methods. Vasiliev et al. [39] have calculated the optical spectrum of small sodium clusters (Na_2, Na_4, and Na_8), using the coupling matrix method (see Section 6). The results obtained from the exact calculation [Eq. (59)] and the two approximate expressions [Eqs. (60) and (61)] are reproduced in Fig. 14. The approximate expressions do not give a good account of the experimental data. However, the exact TDLDA calculation remarkably reproduces the experimental spectral shape [51, 120–122], and the peak positions agree with experiment within 0.1–0.2 eV. The calculated spectrum of Na_2 exhibits three peaks in the 2–5-eV range. Na_4 has a rhombic shape, and its spectrum consists of three peaks in the 1.5–3.0-eV range and a broader feature at higher energy. The results are almost as accurate as the spectra calculated by the configuration interaction method [123]. For increasing cluster size, the spacing between the discrete lines decreases, evolving toward the collective plasmon. For Na_8, with the rather spherical structure of a bicapped octahedron, a single peak is already obtained. The importance of electronic screening is evident by comparing the different panels in the figure.

Pacheco and Martins [124] have studied the sensitivity of the calculated photoabsorption cross sections of the magic clusters Li_8, Li_{20}, Na_8, and Na_{20} to the use of different *ab initio* nonlocal pseudopotentials. The pseudopotential of Bachelet, Hamman, and Schlüter (BHS) [125] and the pseudopotential of Troulliers and Martins (TM) [126] were employed together

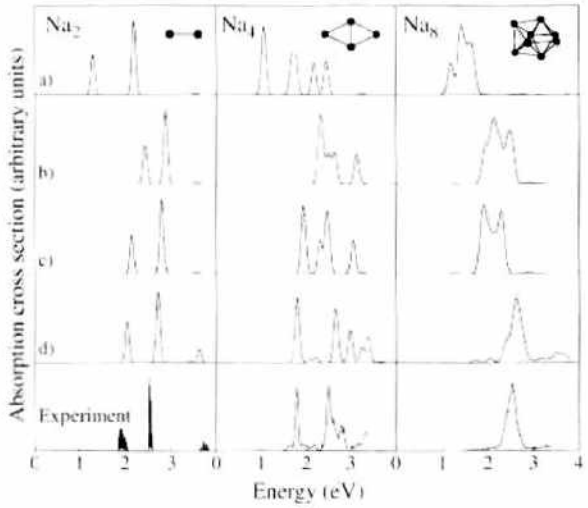


Figure 14. Calculated and experimental absorption spectra of sodium clusters. (a) Absorption spectra from the difference of Kohn–Sham eigenvalues. The other plots show spectra calculated from Eq. (61) (panel b), Eq. (60) (panel c), and Eq. (59) (panel d). All calculated spectra are broadened by 0.06 eV to simulate finite temperature. The experimental spectra are adapted from Refs. [51, 120–122]. Reprinted with permission from [39], I. Vasiliev, S. Ögut, and J. R. Chelikowsky, *Phys. Rev. Lett.* 82, 1919 (1999). © 1999, American Physical Society.

with the LDA for exchange and correlation. The different pseudopotentials lead to the same equilibrium geometries obtained previously for Li and Na clusters in the LDA: a bicapped octahedron for Na_8 [127] and a trigonal prism for Li_8 [114, 128]; the structures obtained for the 20-atom clusters are more difficult to classify. However, the average bond lengths depend on the input pseudopotential, and the bond lengths calculated with the BHS pseudopotential are smaller than those obtained with the TM pseudopotential. The difference is sizable for sodium clusters but minor in the lithium case. As a consequence of the different average bond lengths, differences are also found in the calculated static polarizabilities. The multipeaked line shapes of the photoabsorption cross sections are nearly identical, except for small overall energy shifts. Compared with the jellium LDA spectrum of Fig. 4, the spectrum of Na_{20} obtained by taking into account the full geometry of the ionic background is slightly shifted to lower energies; that is, in the correct direction.

Very small mixed clusters have been studied by quantum chemical *ab initio* methods. Motivated by the measurements of the optical absorption spectra of $LiNa_3$ and Li_2Na_2 [129], configuration interaction calculations have been performed for those two clusters [130], as well as Hartree–Fock calculations for the whole family Li_mNa_{4-m} [131]. The optical response was obtained in both cases from configuration interaction calculations for excited electronic states. The two theoretical studies predict planar rhombic forms as the most stable structures: slightly distorted for $LiNa_3$ and Li_3Na and undistorted for Li_4, Na_4 and Li_2Na_2. The photoabsorption spectrum is sensitive to m; that is, to the relative proportion of Li and Na. The spectrum of Na_4 resembles that obtained from the Mie–Drude theory for an ellipsoidal droplet with three different axes. However, as Li atoms replace Na atoms, the deviations from the Mie theory increase. The measured spectra of $LiNa_3$ and Li_2Na_2 are explained by the *ab initio* calculations. The calculations found low-lying isomers in each Li_mNa_{4-m} case corresponding to different ways of arranging the Li and Na atoms in the four vertices of the rhombus. The comparison between the experimental absorption spectrum of Li_2Na_2 and the spectra calculated for the three isomers detected confirmed that the best agreement is obtained for the lowest-energy isomer. However, the calculated spectra for the two singlet isomers of $LiNa_3$ are so similar that it was not possible to distinguish which isomer or whether a combination of both singlet isomers contributes to the measured spectrum.

TDLDA calculations for the whole $Na_{8-m}Li_m$ family [132] obtained single (plasmon) peaks at both ends of the series, Li_8 and Na_8, consistent with a spherically symmetric electron density for both clusters. The replacement of one or two atoms in the homogeneous clusters produces a spheroidal deformation of the density and introduces extra shoulders in the spectrum. In the middle of the series, for Li_5Na_3 and Li_4Na_4, the resonance is fragmented in

two close peaks. These two clusters have tetrahedral symmetry, and as a rule, the oscillator strength in the absorption spectrum of clusters with nonspherical shape is spread over a wide energy range, thus leading to a broader spectra.

16. OPTICAL RESPONSE OF FISSIONING CLUSTERS

Multiply charged metal clusters, like Na_N^{q+} are less stable than the corresponding neutrals Na_N because of the coulombic repulsion of the unbalanced positive charges, and they can easily experience a process of fission. The process of cluster fission has close analogies to the fission of nuclei [133]. In both cases, a charged droplet will become unstable toward the division into two or more fragments. In simple terms, the multiply charged cluster can be viewed as a droplet, which because of, the unbalanced excess positive charge, tends to deform through elongated shapes keeping the total volume constant. The shape deformation increases the surface area and then the surface energy of the cluster, however. This decreases the cohesion of the cluster and builds up an energy barrier that prevents the spontaneous fission for large cluster sizes, even when the sum of the energies of the fission products is lower than the energy of the multiply charged parent. For small clusters, in contrast, the coulombic repulsion may be so strong that there is no barrier, and fission occurs spontaneously. Optical spectroscopy has been proposed as a tool to follow the dynamics of fragmentation along the fission path [134]. As an example, let us consider the symmetric fission of Na_{18}^{2+}; that is,

$$Na_{18}^{2+} \rightarrow Na_9^+ + Na_9^+ \tag{77}$$

The results of TDLDA calculations of the optical spectrum [134] along the fission path are given in Fig. 15. The cylindrically average pseudopotential (CAPS) model was employed to calculate the ionic and electronic distributions. The CAPS model can be viewed as an extension of the SAPS model considered in Section 12. The CAPS model is based on the observation that most clusters of simple metals have axial symmetry. Taking up this idea, Montag and Reinhard [135] have proposed replacing the total external ionic pseudopotential by its cylindrical average. Using cylindrical coordinates, the CAPS potential then becomes

$$V^{\text{CAPS}}(z,\rho) = \sum_{j=1}^{N} \bar{v}_{\text{ps}}(z,\rho;z_j,\rho_j) \tag{78}$$

$$\bar{v}_{\text{ps}}(z,\rho;z_j,\rho_j) = \frac{1}{2\pi}\int_0^{2\pi} v_{\text{ps}}(|\mathbf{r}-\mathbf{R}_j|)d\varphi \tag{79}$$

A proper choice of the z axis is critical for the success of the method. Montag and Reinhard considered the inertia tensor $\hat{\mathbf{I}}$ of the ionic distribution and identified the z-axis with the principal axis of $\hat{\mathbf{I}}$ whose momentum I_i deviates most from the average momentum $\bar{I} = (I_1+I_2+I_3)/3$. Because the electrons see an axially symmetric potential, their wave functions separate accordingly as

$$\phi_{n\mu}(\rho,\varphi,z) = R_{n\mu}(\rho,z)e^{-i\mu\varphi} \tag{80}$$

The calculation of the ground-state geometry and the electronic structure then proceeds by an interlaced iteration of the Kohn–Sham equations and the ionic stationary conditions.

The energy of the ground state of Na_{18}^{2+} corresponds to the point labeled A in the energy curve of Fig. 15. Its ionic configuration (not shown in the figure) hints at preformed subunits Na_{14}^+ and Na_3^+, which reflects the high stability of the magic two-electron cluster Na_3^+ [136]. Nevertheless, the barriers for the asymmetric and symmetric fission channels are practically equal [136, 137]. Several cluster configurations along the symmetric fission channel are shown on the upper right side of Fig. 15, together with the corresponding electronic distributions. To calculate those configurations, the ions of the cluster are divided in two groups, each one with nine ions (to simulate the symmetric fragmentation), and a constraint is set up on the distance d between the centers of mass of these two fragments. The electronic distribution

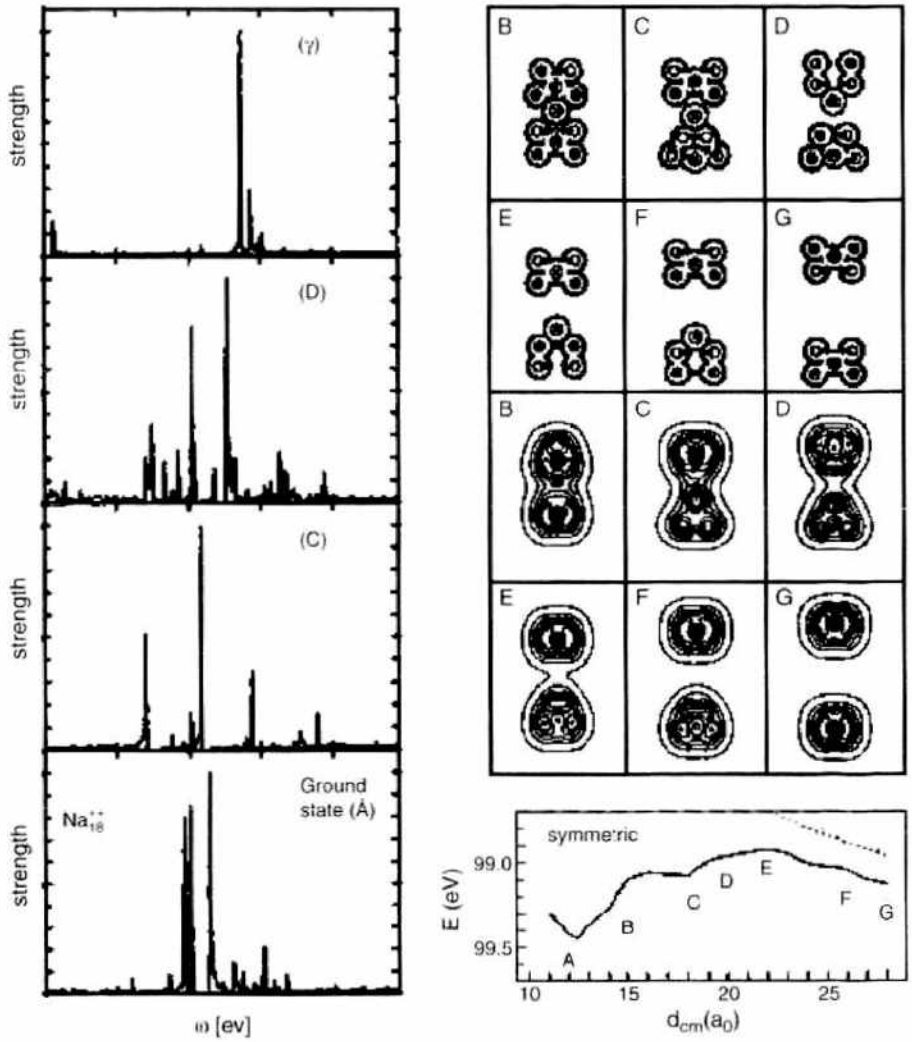


Figure 15. The symmetric fission of Na_{18}^{++} illustrated by plots of the ionic distribution (upper right) and equidensity plots of the electron density (middle right) at various stages indicated over the fission barrier (lower right). The left side shows the calculated optical response at some of those stages. Reprinted with permission from [134], J. A. Alonso, M. Barranco, F. Garcias, P. G. Reinhard, and E. Suraud, in "Fission Dynamics of Atomic Clusters and Nuclei" (J. da Providencia, D. M. Brink, F. Karpechine, and F. B. Malik, Eds.), p. 163. World Scientific, Singapore 2001. © 2001, World Scientific.

and the ionic configuration are then optimized for each value of d in an increasing series of fixed distances d. A marked structural change occurs between configurations A and B (where symmetry is established). This rearrangement corresponds to a large difference in energy between the two configurations. Ionic scission takes place at stage D only shortly before the saddle configuration E, whereas the electron cloud breaks up rather late after the saddle point, around configuration F.

The linear optical response, shown on the left side of Fig. 15, gives an enlightening picture of the various stages along the fission path. The ground-state configuration A exhibits a pronounced resonance between 2 and 2.2 eV. With increasing cluster deformation along the fission path we first observe (basically up to stage D) a gradual increase in the fragmentation of the spectrum, but still centered around 2 eV. The connectivity of the electron cloud still existing in stages C and D has the effect of centering the dipole spectrum around 2 eV, whereas the ions are already fully separated at stage D. A marked change occurs when the electron densities separate at stage F and beyond. The fragmentation of the spectrum disappears and is replaced by the clean plasmon resonance of free Na_9^+ around 2.7–2.8 eV. The optical response thus provides a tool of analysis of the various configurations along the fission path. It might be experimentally feasible to follow such a fission path by femtosecond time-resolved recording of the electronic response following short laser pulses, as done

nowadays for mapping the isomerization paths of biological photoreceptors and chemical reaction path of molecules [138, 139]. One might even hope to have access to fission time scales and to thus be able to estimate viscosity effects, in a way somewhat similar to the nuclear case.

17. THERMAL LINE BROADENING

To compare with experiment, most TDDFT calculations perform an ad hoc broadening of the photoabsorption spectrum: The spectral lines, calculated for a static geometrical configuration of the cluster, are broadened through convolution with Gaussian or Lorentzian functions [50, 65, 114], but a first-principles description of absolute magnitudes, peak positions, and line broadening can be achieved by calculating the photoabsorption cross sections along finite temperature molecular dynamics simulation trajectories. Shape fluctuations were introduced to account for the line broadening of simple metal clusters [52], with good results for Na clusters. A step forward was taken by Pacheco and Schöne [140], by treating in perturbation theory the deviation of the ionic potential from being spherical. The Monte Carlo simulations performed in this work mimic the effect of temperature on the line width and on the position of the resonance peaks.

However, true first-principles calculations have only been done recently by Moseler et al. [141]. They have calculated the TDLDA optical spectra of Na_3^+, Na_5^+, and Na_9^+ at finite temperatures. The excitation energies were calculated by solving the eigenvalue problem [38] of Eq. (59) and averaging the calculated cross sections for a time propagation of 10 ps. The results are shown in Fig. 16. The structure of Na_3^+ is an equilateral triangle. Its calculated spectrum of at $T = 100$ K, given by the histograms in panel (a) of the figure, shows two peaks. The low-energy peak originates from two transitions ($\hbar\omega_1 = \hbar\omega_2 = 2.65$ eV) from the occupied s-like state to two empty p-like states with orbitals in the plane of the cluster. The three relevant orbitals are shown in the inset on the left of panel (a), where they are labeled as 1, 2, and 3, respectively. The other peak, centered at an energy of 3.41 eV, is the result of the excitation to the other p-like orbital, perpendicular to the cluster plane (see the inset on the right side of the same panel). The positions of the two calculated peaks agree well with the experimental spectrum, given by the continuous line [142]. The intensity and width of the low-energy peak are correctly predicted, but the measured high-energy peak is less intense than the calculated one. The reason is that the experimental cross section $\sigma(\omega)$ is determined by measuring the photodepletion of the Na_3^+ intensity because of dissociation following the absorption of one photon. The two p-like states in the cluster plane are antibonding, so excitation into these states promotes dissociation, but excitation into the p-like state perpendicular to the cluster plane does not have a direct destabilizing effect. Thermal motion distorts the symmetry of Na_3^+ and lifts the degeneracy of the ω_1 and ω_2 transitions. In addition, the mean bond distance in the cluster is anticorrelated with the average of ω_1 and ω_2, and also with ω_3. As a consequence, the line width can be explained by the combined effect of the line splitting caused by symmetry breaking (called degeneracy lifting by Moseler et al. [141]) and breathing vibrations (or spectral sweeping mechanism).

At a temperature of 100 K, the dynamics of Na_5^+ distort the ground-state D_{2d} symmetry (compare structures α and β in the inset in panel b, where the left triangle of structure β is more elongated compared to structure α) and fragments the absorption line at $\hbar\omega = 2.8$ eV into two spectral lines (compare the oscillator strengths given in the upper-right inset in panel b). The increase of T to 300 K results in bent geometries (structure γ in panel c). In this case, the low-energy line at ≈ 2 eV is fragmented. All the lines are further broadened by the effect of the breathing modes. For larger clusters, thermal isomerization leads to another line-broadening mechanism, which adds to the other two mechanisms discussed above. The ground state of Na_9^+ is oblate (the contours of constant electron density have the shape of an ideal oblate spheroid). As a consequence, the absorption lines are distributed in a bimodal manner, with further broadening arising from the line fragmentation and breathing mechanisms. By increasing the temperature to 450 K, the spectrum transforms to one with a single broad maximum. The main reason for the change in shape is the transformation between the low-temperature ground-state structure α (a tricapped trigonal prism) and the

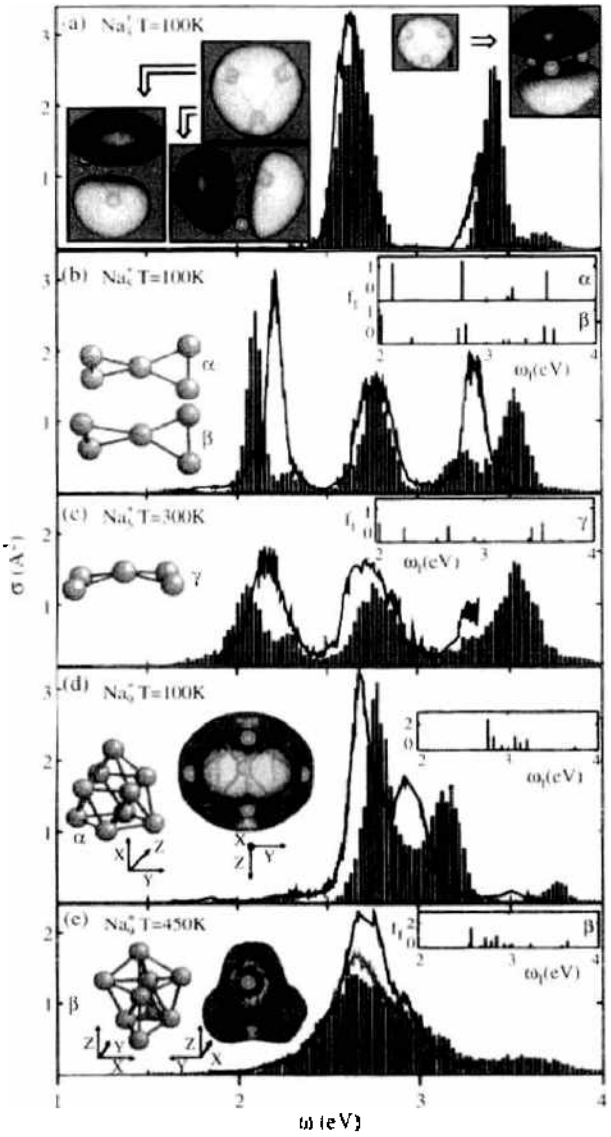


Figure 16. Comparison of theoretical (histograms) and experimental (solid curves) [142] photoabsorption cross sections of Na_5^+ at 100 K (panel a), Na_5^+ at 100 K (panel b), Na_5^+ at 300 K (panel c), Na_5^+ at 100 K (panel d), and Na_5^+ at 450 K (panel e). Insets in panel (a) indicate constant density contour plots corresponding to electronic orbitals relevant for the discussion of the spectrum (see text). The contour plots in panels (d) and (e) correspond to the total electron density and reflect the shape of the cluster. Insets on the upper right sides of panels (b), (c), (d), and (e) give the oscillator strength for the cluster structures shown on the left side. Reprinted with permission from [141], M. Moseler, H. Häkkinen, and U. Landman, *Phys. Rev. Lett.* 87, 053401 (2001). © 2001, American Physical Society.

isomer labeled β. The static spectrum of this isomer is shifted to lower energies compared to that of the ground state, and the broad spectrum at 450 K is the result of contributions from both isomers.

18. FULL SOLUTION OF THE TIME-DEPENDENT KOHN–SHAM EQUATIONS

18.1. Theory

Another efficient method exists for calculating the optical spectrum in finite systems, based on directly solving the TDKS equations of DFT in response to the external laser field. This method, originally used for the study of nuclear reactions [143], has been applied to clusters [144–147]. The set of equations to be solved for the combined electron and ion dynamics is formed by the TDKS Eqs. (30) and (31), together with the Newton equations for the motion

of the ions

$$m_\alpha \frac{d^2\mathbf{R}_\alpha}{dt^2} = \mathbf{F}_\alpha(\mathbf{R}_\alpha, t) \tag{81}$$

The applied perturbing field $V_{\text{ext}}(\mathbf{r}, t)$ appearing in Eq. (30) is, in this case, the laser potential $V_{\text{laser}}(\mathbf{r}, t)$ describing the classical time-dependent external electromagnetic field acting on the system. In the Newton's equations for the ions, $\mathbf{R}_\alpha$ stands for the position of the ion labeled α, m_α for its mass, and $\mathbf{F}_\alpha$ is the instantaneous force on that ion. This force is calculated through the Ehrenfest theorem

$$\mathbf{F}_\alpha(\mathbf{R}_\alpha, t) = -\left\langle \Psi(t) \middle| \frac{\partial}{\partial \mathbf{R}_\alpha} \hat{H} \middle| \Psi(t) \right\rangle \tag{82}$$

This is just the extension of the Hellmann–Feynman theorem to the time-dependent domain. There are no Pulay corrections to this expression when the Kohn–Sham wavefunctions are expanded in a regular grid space, but this would not be the case if one used localized basis sets [146, 148, 149].

The ground state of the electronic system in the nuclear equilibrium configuration, that is, the solution of the ground-state Kohn–Sham equations

$$\left[-\frac{1}{2}\nabla^2 + V_H(\mathbf{r}) + V_{\text{xc}}(\mathbf{r}) + V_N(\mathbf{r})\right]\psi_i(\mathbf{r}) = \varepsilon_i \psi_i(\mathbf{r}) \tag{83}$$

$$n(\mathbf{r}) = \sum_{i=1}^{N} |\psi_i(\mathbf{r})|^2 \tag{84}$$

is the starting point for the time-dependent simulations. The Kohn–Sham orbitals are then propagated in time as

$$\psi_i(\mathbf{r}, t + \Delta t) = e^{-\int_t^{t+\Delta t} \hat{H}_{\text{KS}}(t')dt'} \psi_i(\mathbf{r}, t) \tag{85}$$

In this method, only occupied states need to be propagated, and there is no need of computing empty states.

To obtain the linear optical absorption spectrum, one can follow an scheme proposed by Yabana and Bertsch [144, 147] and excite all frequencies of the system by giving a small momentum $\mathbf{k}$ to the electrons. This is achieved by initially transforming the wave functions according to

$$\psi_i(\mathbf{r}, \delta t) = e^{ikz} \psi_i(\mathbf{r}, 0) \tag{86}$$

and then propagating these wavefunctions for some (finite) time. The spectrum can then be obtained from the expression of the dipole strength function $S(\omega)$

$$S(\omega) = \frac{2\omega}{\pi} \operatorname{Im} \alpha(\omega) \tag{87}$$

where the dynamical polarizability $\alpha(\omega)$ is essentially the Fourier transform of the dipole moment of the system

$$\alpha(\omega) = \frac{1}{k} \int e^{i\omega t} [D(t) - D(0)] dt \tag{88}$$

With this definition, the Thomas–Reiche–Kuhn f-sum rule for the number of electrons, N, is given by the integral

$$N = \int S(\omega) d\omega \tag{89}$$

This sum rule can be used to check the quality of the calculations. Another check is energy conservation, which the TDDFT respects when there is no external field applied.

For the purposes of obtaining nonlinear optical properties, the evolution of the system under the influence of a laser field, treated in the dipole approximation (although this constraint can be removed), is followed. The emitted harmonic spectrum can then be calculated from the acceleration of the dipole moment [150, 151]

$$H(\omega) \propto \left| \int dt e^{i\omega t} \frac{d^2}{dt^2} D(t) \right|^2 \tag{90}$$

In the implementation of Rubio and coworkers, who have developed a code called OCTOPUS [146], charge is absorbed at the boundaries of the simulation region during the propagation. This can be simulated by using an imaginary absorbing potential [152], which is added to the Kohn–Sham potential; that is,

$$V_{\text{eff}}(\mathbf{r}, t) = V_{\text{KS}}(\mathbf{r}, t) - iV_{\text{abs}}(\mathbf{r}) \tag{91}$$

where $V_{\text{abs}}(\mathbf{r})$ is zero in the inner region of the simulation box and rises smoothly until reaching the edges. By adjusting both the height and the shape of the potential, one can select which momenta are absorbed and prevent the unwanted reflections at the boundary. Another option for simulating the absorption is to use a mask. In this case, the wavefunction is multiplied at each time step by a function that is 1 in the inner region of the simulation box and gradually goes to 0 at the borders. The absorbed charge can be interpreted as the ionization probability and can be used to estimate the photo-electron spectra. The box size has to be big enough that the physical system is not perturbed by the absorbing boundaries.

18.2. Applications to Carbon Clusters

Medium-size carbon clusters are predicted to have a wide variety of isomers with the form of cages, bowls, planar graphitic structures, rings, and linear chains. The theoretical and experimental study of the different isomers may help to understand the way fullerenes form [153]. The smallest possible fullerene, consisting of 12 pentagons with no graphitic hexagons intercalated, is an isomer of C_{20}. Other low-energy isomers of C_{20} include a bowl (which can be considered as a fragment of C_{60}), several rings, and other closed three-dimensional arrangements. Prinzbach et al. [154] have reported the production of the cage and bowl members of the family. The smallest fullerene cannot be expected to form spontaneously, but has been produced from the precursor $C_{20}H_{20}$. The bowl was produced in the same way, and photoelectron spectroscopy has been used to distinguish between the different species.

It is very difficult to make reliable theoretical predictions of the most stable structure of C_{20}. In fact, different levels of theory favor different isomers. At the Hartree–Fock level, the ring is the ground state, followed by the bowl and the cage [155]. DFT in the LDA approximation reverses the order, predicting the cage as the lowest-energy structure [156, 157]. The use of better functionals based on the generalized gradient approximation (GGA) does not clarify matters: the ordering of the isomers depends on the correction used [155–157]. Quantum Monte Carlo (QMC) and coupled cluster (CC) methods have also been applied in an attempt to resolve the issue, yielding bowl–ring–cage ordering using the former method [155] and cage–bowl–ring ordering using the latter [158]. Furthermore, it seems that the results are sensitive to the pseudopotential [158]. Changing slightly the pseudopotential cutoff radius may actually reverse the ordering of the isomers. Another complication is that entropy effects affect the relative stability, and the calculated free energies as a function of the temperature [159] have been used to assign the dominant species generated in experiments by vaporizing graphite or prepared from precursors. Thus, it is important to find experimental methods to determine the structure that are sensitive enough to be usable with the available cluster beam intensities, and optical spectroscopy fulfills the requirements. The geometrical structures of six members of the C_{20} family are given in Fig. 17: the smallest fullerene (cage), which is a Jahn–Teller distorted dodecahedron; the ring; the bowl; and three cagelike structures, labeled as (d), (e), and (f). Structures (d) and (f), related by a Stone–Wales transformation [160], are composed of four hexagons, four pentagons, and

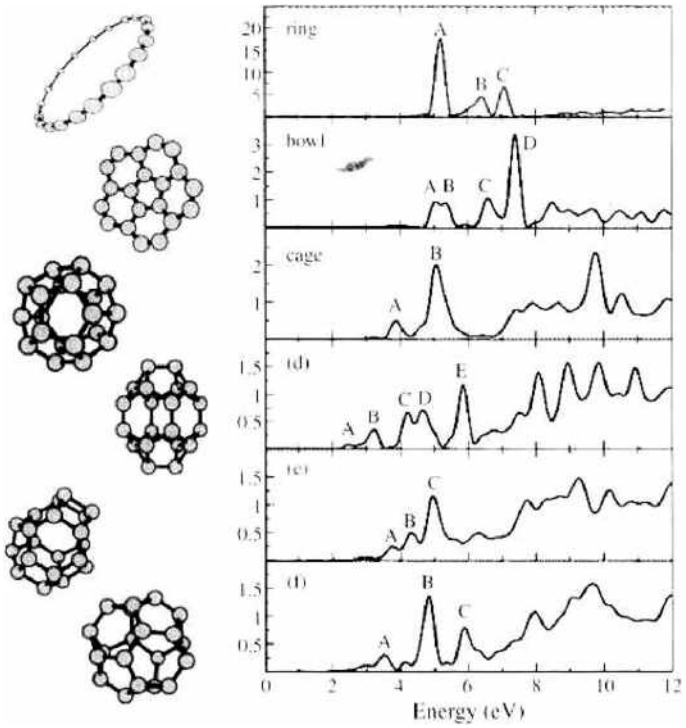


Figure 17. Dipole strength function, in eV^{-1}, for several isomers of C_{20}. Adapted with permission from [162], A. Castro, M. A. L. Marques, J. A. Alonso, G. F. Bertsch, K. Yabana, and A. Rubio, *J. Chem. Phys.* 116, 1930 (2002), © 2002, American Institute of Physics.

four four-membered rings. These structures are the six isomers with lower energy as calculated by Jones within the LDA approximation [161]. Other structures, such as bicyclic rings, chains, and tadpoles may be favored by entropy at high temperature and have been observed experimentally. However, neither of them seem to be a possible low-temperature ground state. The results of TDDFT calculations of the optical absorption [162] are also shown in Fig. 17. Table 5 gives the energies and strengths of the lowest transitions with appreciable strength. These results were obtained by solving the TDKS equations, with the LDA for exchange and correlation, and representing the wavefunctions on a uniform spatial grid. The norm-conserving soft-core pseudopotentials of Troullier and Martins [126] were used to avoid explicit consideration of the 1*s* core electrons.

The dipole strength functions shown have been averaged over all orientations of the system. In the case of the ring, the response in the direction perpendicular to the ring plane is almost negligible below 8 eV compared with the response within the plane. Also, for the quasi-planar bowl isomer, the perpendicular response is extremely weak in that energy range. Although present molecular beam experiments are not able to discriminate between the different spatial directions, the averaged spectra are still sufficiently different to discriminate between the different structures without ambiguity. Two regions can be distinguished in all the graphs: the peaks that can be seen in the near ultraviolet, and a broad absorption that starts at around 7.5 eV. Focusing attention on the lower-energy peaks, the ring exhibits the largest optical gap in the spectrum and also the strongest collective transition. The bowl also has a high optical threshold, larger than 5 eV, but the intensity of the first significant

Table 5. Energies, in electron volts, and strengths (between parentheses) of selected peaks in the optical response of the C_{20} isomers of Fig. 15 [162].

	Ring	Bowl	Cage	d	e	f
A	5.20 (5.4)	5.05 (0.7)	3.88 (0.2)	2.47 (0.03)	3.77 (0.1)	3.53 (0.1)
B	6.42 (1.4)	5.35 (0.7)	5.07 (1.3)	3.23 (0.1)	4.33 (0.2)	4.84 (0.7)
C	7.09 (2.0)	6.60 (0.7)		4.21 (0.3)	4.96 (0.5)	5.89 (0.3)
D		7.41 (2.3)		4.67 (0.4)		
E				5.86 (0.4)		

transition is an order of magnitude weaker than in the ring. The relative intensities of the peaks, the fact that the first excitation is divided into two for the bowl, and the relative strength of the excitations in the 6–7 eV region can all be used to distinguish the bowl isomer from the ring.

The spectra of the four three-dimensional isomers start at much lower energy and are more similar to each other, which is expected from their similar geometries. The fullerene isomer exhibits two peaks at 3.9 and 5.1 eV, with the second peak much stronger than the first one. Most of the strength concentrates above the ionization threshold (7.5 eV) and has a broad plateau starting at around 7 eV. This is different from planar-like isomers, in which an important fraction of the strength appears below 7 eV. Isomer (d) can be distinguished by the presence of a transition at quite low energy, 2.5 eV, as well as by the fragmentation into many states going up to 6 eV. Isomer (e) differs from the fullerene cage by the presence of a transition (labeled "B" in the figure) between the transitions that would be seen in the cage. The spectrum of isomer (f) is similar to that of the fullerene cage up to the second peak, but it is shifted down by about 0.3 eV. This is close to the border of where the TDDFT energies are reliable. However, isomer (f) also has a third peak near 6 eV in a region in which there is a gap in the fullerene cage spectrum, and that difference would be definitive.

The benzene molecule, C_6H_6, was studied using the OCTOPUS code [146] and its predecesor code [163]. The results from the OCTOPUS code are shown in Fig. 18. The main features are the narrow peak at about 7 eV, which corresponds to the transition between the π and π^* orbitals (also shown in the figure) and that is a characteristic of carbon conjugated compounds, and the broad feature above 9 eV, which corresponds to the $\sigma \rightarrow \sigma^*$ transition. The method works well for all organic molecules studied (see the OCTOPUS Web page [146] for a database).

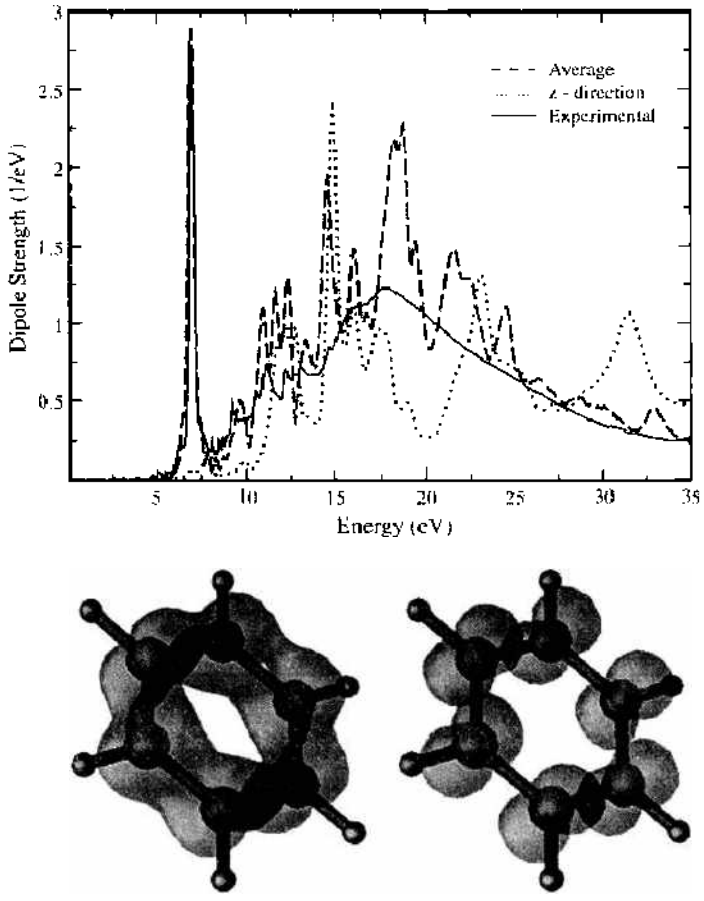


Figure 18. Optical absorption of the benzene molecule calculated with the OCTOPUS code. Experimental result are from Koch [164]. The bottom panel shows π and π^* Kohn–Sham orbitals. Transitions between these two states are responsible for the large absorption peak at 7 eV. Reprinted with permission from [146], M. A. L. Marques, A. Castro, G. F. Bertsch, and A. Rubio, *Computer Phys. Commun.* 151, 60 (2003). © 2003, Elsevier.

18.3. Applications to Biomolecules

In addition to the spectacular advances over the last years in the characterization of structural and dynamical properties of biomolecules by a combination of quantum mechanical and classical molecular mechanics methods, the theoretical understanding of the interaction of those molecules with external time-dependent fields is in its infancy in spite of the large amount of experimental work on photoactive molecules. In particular, processes related to vision and photosynthesis rely on a subtle interplay between optical absorption in the photoactive center and its decaying mechanism through the coupling to the internal vibrational modes of the molecule, including isomerization processes as well as coupling to the environment (supporting protein and solvent).

In this context the green flourescent protein (GFP) has been studied experimentally in various environments (in solution as well as in vacuo) and has been found to exhibit a rich and complex behavior that is the subject of much current debate. The measured optical absorption spectrum of the wild-type (wt) GFP shows two main resonances at 2.63 and 3.05 eV [165, 166] (see Fig. 19), that are attributed to two different thermodynamically stable protonation states of the chromophore (neutral and negative configurations of the chromophore, respectively). So far, *ab initio* quantum chemistry has not been able to provide satisfactory agreement with the spectroscopic data, and thus it has not contributed too much to confirm or rule out various possible scenarios of photodynamics in GFP.

A good description of the optical properties of the GFP photoreceptor has been achieved in recent work [167] using an approach combining a quantum-mechanical molecular-mechanics (QM-MM) method to obtain the structure with TDDFT to treat the electronic excitations. The structures were optimized using a hybrid QM-MM method [168, 169] with a semiempirical Hamiltonian [170] to describe the quantum subsystem. The QM region was formed by three amino acid sequences, Ser65, Tyr66, and Gly67. The frontier between the QM and MM regions was treated within the H-link approximation. In this approach, a hydrogen atom is included whenever the frontier between the QM and MM regions passes through a chemical bond. This H-link atom is forced during the minimization to be on the line along the frontier bond and does not interact with the MM atoms. The optimized structure of the chromophore with the most important neighbor residues is shown in Fig. 20.

In contrast, the anionic form of the chromophore was obtained by deprotonation of the Tyr66 and protonation of Glu222. The computed photoabsorption spectra of the GFP neutral and anionic chromophores, shown in Fig. 19, are in excellent agreement with experiment, assuming the presence of the two forms, protonated and deprotonated, of the photoreceptor in an approximate 4:1 ratio. Furthermore, it can be seen in the inset of Fig. 19 that light polarized along the x-direction is responsible for the lowest optical transition. The molecule is nearly transparent to visible light polarized along the other two orthogonal directions. GFP turns out to be a rather anisotropic molecule in the visible range, a property that could be used to enhance the photodynamic processes in well-oriented GFP samples for

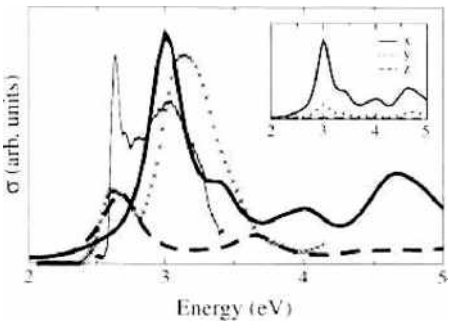


Figure 19. Computed photoabsorption cross section of the neutral (thick solid line) and anionic (thick dashed line) chromophores. For comparative purposes the anionic results have been divided by four. Experimental results at 1.6 K (thin solid lines) [165] and room temperature (crosses) [166] are also given. The inset shows a decomposition of the calculated spectrum of the neutral chromophore in the three directions, showing the inherent anisotropy of the green fluorescent protein molecule. Reprinted with permission from [167], M. A. L. Marques, X. López, D. Varsano, A. Castro, and A. Rubio, *Phys. Rev. Lett.* 90, 258101 (2003). © 2003, American Physical Society.

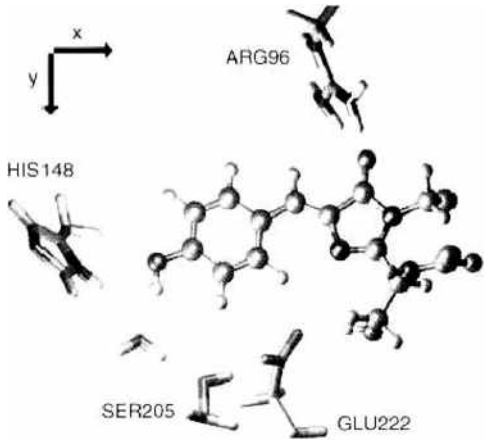


Figure 20. Optimized structure of the neutral chromophore and its closest charged residues inside the green fluorescent protein: His148, Arg96 (positive), and Glu222 (negative). Reprinted with permission from [167], M. A. L. Marques, X. López, D. Varsano, A. Castro, and A. Rubio, *Phys. Rev. Lett.* **90**, 258101 (2003). © 2003, American Physical Society.

opto-electronic devices. This new approach holds great promise for future applications in biochemistry and biophysics, as it is able to handle not only the optical response but also ultrashort femtosecond electron–ion dynamics. Preliminary calculations [171] for the optical spectra of the DNA basis are also in excellent agreement with available experimental data.

18.4. Local Orbitals

A method has been derived [172] on the basis of the description of the electronic wavefunctions using a linear combination of atomic orbitals (LCAO). Because the size of the LCAO basis is small, the scheme leads to matrices with a size substantially smaller than other methods. In that method, use is made of the SIESTA code [173, 174] to compute the initial wavefunctions and the Hamiltonian matrix for each time step. In the SIESTA code, the core electrons are replaced by norm-conserving pseudopotentials, and the basis set is a linear combination of numerical atomic orbitals (NAOs), constructed from the eigenstates of the atomic pseudopotentials.

To calculate the linear optical response, the explicit time evolution of the wavefunctions is followed. A bounded system in a weak electric field is first considered and the ground state of the system is calculated using standard time-independent DFT. This is achieved in practice by including a perturbation $\Delta H = -\mathbf{E}\cdot\mathbf{x}$ and setting a small value for the field. Then, the field is switched off at time $t = 0$ and the occupied Kohn–Sham eigenstates are propagated in time by solving the TDKS equation with the ALDA for XC [172].

The calculation of the optical response of Na_8 is a difficult test for codes that use localized orbitals because of the delocalized electron density of this cluster. The basis set used by Tsolakidis [172] includes 13 NAOs per atom: two radial shapes to represent the $3s$ states plus a polarization p shell with confinement radii $r_s = r_p^{\text{pol}}$ equal to 12.2 a.u., and two additional $3p$ and $3d$ shells with confinement radii $r_p = r_d$ of 10.0 a.u. The shape of the calculated imaginary part of the linear dynamical polarizability is in good agreement with other calculations [39] and with experiment [51]. However, the maximum of the plasmon is obtained at 2.8 eV; that is, the peak is shifted by 0.27 eV to higher energies with respect to experiment. This shift seems to be related to the extension of the LCAO basis: The more confined the orbitals, the larger the shift.

C_{60} is characterized by strong electron–electron interactions as a result of the confinement. Thirteen NAOs were used for this cluster: two different radial shapes for the description of the $2s$ states, followed by another two for the $2p$, plus an additional shell of d orbitals. The calculated spectra show a small dependence of the confinement radii. The dipole strength function obtained from the time evolution of the dipole moment is shown in Fig. 21, for energies up to 60 eV. The main features of the spectrum are the low-energy transitions arising from the π electrons and the σ and π transitions in the region of 14–27 eV. In the

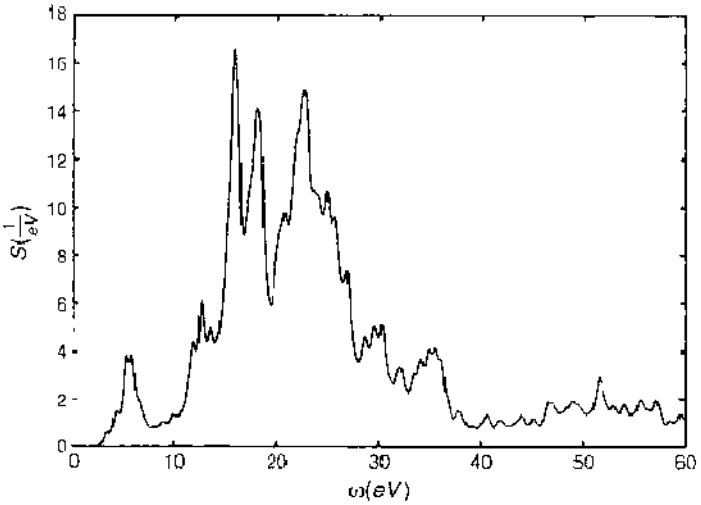


Figure 21. Calculated dipole strength function of C_{60} versus energy. Reprinted with permission from [172], A. Tsolakidis, D. Sánchez-Portal, and R. M. Martin, *Phys. Rev. B* 66, 235416 (2002). © 2002, American Physical Society.

low-energy region, peaks appear at 3.5, 4.4, 5.4, and 5.8 eV, which agrees well with those observed in the experiments [175]. The integration of the dipole strength over energy gives a sum rule strength of 223.8, whereas an exact fulfillment of the sum rule should give a strength of 240. The sum rule is then satisfied up to 93.2%, which reflects the incompleteness of the basis set, which fails to reproduce some of the high-energy excitations.

The nonperturbative nature of the method allows, for large values of the applied field, to obtain nonlinear polarizabilities. Tsolakidis et al. [172] also developed a method for the calculation of the imaginary part of the integrated frequency dependent second-order nonlinear polarizability. The advantage of the explicit time method is that it uses the same operations as in the linear case. The disadvantage is that, unlike the linear case, in which each Fourier component is independent, the nonlinear response depends on the detailed spectrum of the applied field. The method developed is valid for a field that is a step function, and it was applied to C_{60}.

19. ASSESSMENT OF EXCHANGE–CORRELATION FUNCTIONALS FOR THE CALCULATION OF OPTICAL PROPERTIES

When the excitation energies are obtained from the time-dependent density, that is, by solving the TDKS equations [see Eq. (30)], an approximated time-dependent XC potential $V_{xc}(\mathbf{r},t)$ has to be used. This potential is expected to be much simpler to model than the XC kernel $K_{xc}[n](\mathbf{r},t,\mathbf{r}',t')$, which is required in the other formulation; that is, when the excitations are obtained from the position of the poles of the KS linear response function.

The simplest and most commonly used approximation is again the ALDA in which the static LDA XC potential is used in the TDKS equations but is evaluated at the time-dependent density; that is,

$$V_{xc}^{\mathrm{ALDA}}(\mathbf{r},t)=V_{xc}^{\mathrm{LDA}}[n(\mathbf{r},t)] \tag{92}$$

One can also start with the static GGA [176, 177]

$$E_{xc}^{\mathrm{GGA}}=\int d^3r\, n(\mathbf{r})\varepsilon_{xc}^{\mathrm{GGA}}[n(\mathbf{r}),\nabla n(\mathbf{r})] \tag{93}$$

where $\nabla n(\mathbf{r})$ is the gradient of the density at the point $\mathbf{r}$ and $\varepsilon_{xc}^{\mathrm{GGA}}$ is an analytic function of $n(\mathbf{r})$ and $\nabla n(\mathbf{r})$ with some free parameters that are either fitted to experiment or determined by fulfilling some exact sum rules. Following, then, the same reasoning, it is straightforward to construct adiabatic GGA potentials. Unfortunately, the onset of absorption calculated either with the adiabatic LDA or GGA functionals are typically below the observed ones

(by several electron volts in the case of atoms). This problem is once more related to the wrong asymptotic behavior of the effective Kohn–Sham potential, which goes exponentially to zero instead of having the correct behavior $-1/r$ for neutral systems. This is a result of the insufficient correction of the self-interaction part of the Hartree potential. As a simple way to correct the asymptotic part of the adiabatic potential, one can use the adiabatic approximation of van Leeuven and Baerends (known as LB94) [178]. These authors applied the Becke GGA construction [176] not to the derivation of the exchange energy functional (as in Becke's original work), but to the modeling of the XC potential directly. By imposing the correct asymptotic behavior to the XC potential, much better ionization potentials, and better energy eigenvalues in general, have been obtained. However, the high-lying excitation energies calculated with the LB94 potential are usually overestimated for small molecules, and the performance for low-lying states is less accurate than for ALDA [179].

To overcome some of the difficulties encountered when trying to write V_{xc} as an explicit functional of the density, orbital-dependent xc-functionals have been introduced. Those functionals are written explicitly in terms of the Kohn–Sham orbitals, albeit remaining implicit density functionals because of the Runge–Gross theorem (see Section 4 and Ref. [27]). The EXX functional, which is a typical member of this family, is obtained by expanding the action functional A_{xc} in powers of e^2 (e is the electronic charge), and retaining the lowest-order term only; that is, the exchange term. This is given by the Fock integral

$$A_{xc}^{EXX} = -\frac{1}{2}\sum_{j,k}^{occ}\int_{t_0}^{t_1} dt \int d^3r \int d^3r' \frac{\psi_j^*(\mathbf{r}',t)\psi_k(\mathbf{r}',t)\psi_j(\mathbf{r},t)\psi_k^*(\mathbf{r},t)}{|\mathbf{r}-\mathbf{r}'|} \tag{94}$$

From this action functional, one determines the local Kohn–Sham potential by a series of chain rules for functional derivatives. The procedure is called the OEP, and the derivation of the time-dependent version of the OEP equations [42] is very similar. The final equation that determines the EXX potential has the form of an integral equation, and its solution poses a hard numerical problem. By performing an approximation first proposed by Krieger, Li, and Iafrate [41], it is possible to obtain a semianalytical solution of the integral equation. The kernel corresponding to the OEP was presented in Eq. (64). The SIC-LDA [180] is another example of an orbital-dependent functional

$$E_{xc}^{SIC} = E_{xc}^{LDA}[n_\uparrow(\mathbf{r}), n_\downarrow(\mathbf{r})] - \sum_{\sigma}\sum_{i=1}^{N} E_{xc}^{LDA}[|\psi_{i\sigma}(\mathbf{r})|^2, 0] - \frac{1}{2}\sum_{\sigma}\sum_{i=1}^{N}\int d^3r d^3r' \frac{|\psi_{i\sigma}(\mathbf{r})|^2|\psi_{i\sigma}(\mathbf{r}')|^2}{|\mathbf{r}-\mathbf{r}'|} \tag{95}$$

In this functional, σ is the spin index ($\uparrow, \downarrow$). The first term in the functional is the LDA approximation; in its spin-dependent version, the second term subtracts the self-interaction part of the XC functional, and the last term exactly cancels the self-interaction part of the Hartree term (that is the spurious interaction of an electron with itself that appears in the classical Coulomb interaction expressed in terms of the total electron density). The time-dependent generalization of $V_{xc}^{SIC-LDA}$ is immediate.

19.1. Metal Clusters

The performance of the adiabatic $V_{xc}(\mathbf{r},t)$ functionals discussed above has been compared for small sodium and silane molecules [145]. All the calculated optical spectra of Na_2 (LDA, GGA, EXX, SIC-LDA, and LB94) are quite similar, regardless of the XC potential used. Those spectra show three peaks in the 2–5-eV range, and they compare quite well with experiment, although the DFT peaks are all shifted toward higher energies by amounts ranging from 0.2 to 0.4 eV. Because the electronic transition energies can be expressed [see Eq. (57)]

$$\Omega_{cv} = \varepsilon_c - \varepsilon_v + \left\langle \psi_v\psi_c \left| \frac{1}{|\mathbf{r}_1-\mathbf{r}_2|} \right| \psi_v\psi_c \right\rangle + \Delta_{xc} \tag{96}$$

the shift can be understood as resulting from the competition between the Coulomb repulsion contribution to the response and the electron-hole attraction in the XC part. For a given $v \to c$ transition, the ALDA approximation introduces only an effective static attractive

electron-hole interaction (the expressions are more complicated for other kernels)

$$\Delta_{ss'} = \int d^3r \psi_s^*(\mathbf{r})\psi_{s'}(\mathbf{r}) \frac{\partial V_{xc}(\mathbf{r})}{\partial n} \psi_{s'}^*(\mathbf{r})\psi_s(\mathbf{r}) \tag{97}$$

but it is clear that the effective attractive interaction is not complete. Dynamic effects may be needed in the kernel to recover this minor effect. Also, temperature effects may introduce a broadening of the spectrum as well as a shift of the peaks to lower frequencies [140, 141]. The functional giving the best results, although by a small margin, is EXX, whereas the strongest departure from the experiment is found for LB94.

For Na_4, shown in Fig. 22, all the calculations yield similar spectra. The spectrum consists of three peaks in the 1.5–3.5-eV range and a broader feature around 4.5 eV. The comparison with the experimental peak positions is quite good, although the calculated peaks appear, again, shifted to higher energies by ≈0.2 eV. The deviation of the LB94 is a little bit larger. In general, the errors in the calculations with all these functionals are larger for the high-energy peaks, which involve transitions to states near the ionization threshold. The figure also contains the results of a GW quasiparticle many-body calculation including excitonic effects [181] (see Section 21).

19.2. Silanes

The two simplest hydrogen-terminated silicon clusters, silane (SiH_4) and disilane (Si_2H_6), pose a much harder challenge that the alkali clusters because of the presence of p electrons and also because of the hydrogen atoms. As expected, the different XC functionals lead to dissimilar results [145]. The HOMO-LUMO gaps obtained for SiH_4 with the different functionals are 8.10 eV (LDA), 8.12 eV (GGA), 8.40 eV (LB94), 7.70 eV (SIC-LDA), and 8.77 eV (EXX). The differences between them are not large, although it is woth noticing the smallest value is obtained for SIC and the largest ones for LB94 and EXX. All those gaps are, however, smaller than the value of 13.0 eV obtained in a GW calculation [182]. For Si_2H_6 the HOMO-LUMO gaps are 6.76 eV (LDA), 6.80 eV (GGA), 6.58 eV (LB94), 5.98 eV (SIC), and 7.17 eV (EXX), and again the variations are not large. However, the main difference between those methods is a nearly rigid shift of the spectrum of the LDA and GGA energy eigenvalues with respect to LB94 and EXX. The upward shift leads to lower electronic binding energies. The SIC spectrum is also shifted, but much less so.

The experimental absorption spectrum of silane [183], given in Fig. 23, has three peaks between 8 and 12 eV, followed by a much broader feature at higher energies. The peaks derive from a Jahn–Teller splitting of the triply degenerate $2t_2 \rightarrow 4s$ transition. The spectra obtained with the LDA and GGA functionals (upper panel in the figure) are quite similar to each other, and the onset of absorption is underestimated by around 1 eV with respect to

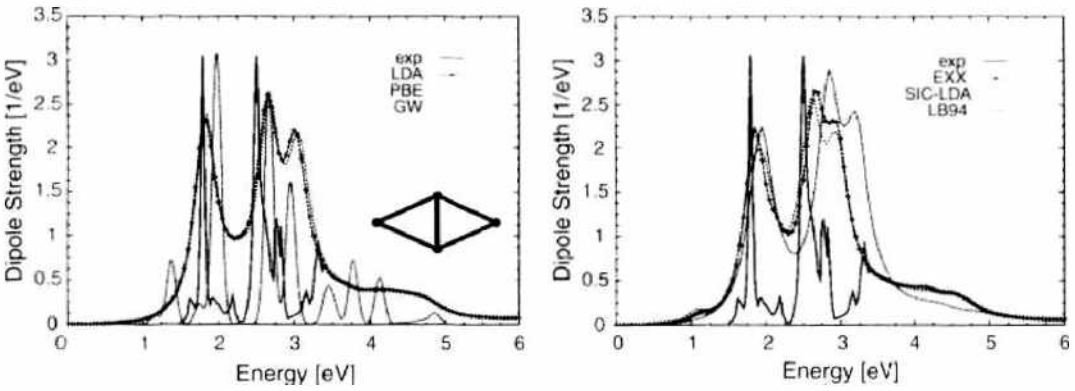


Figure 22. Averaged dipole strength of Na_4. The curve labeled "exp" shows the experimental photodepletion spectrum of Wang et al. [51]. GW shows the results of a many-body calculation including self-energy and excitonic effects [181], and PBE indicates the results of GGA calculations with the functional of Perdew, Burke, and Ernzerhof [177]. The other curves (LDA, EXX, SIC-LDA, and LB94) correspond to calculations with functionals explained in the text. The geometrical structure of Na_4 is shown in the inset. Reprinted with permission from [145], M. A. L. Marques, A. Castro, and A. Rubio, *J. Chem. Phys.* 115, 3006 (2001). © 2001, American Institute of Physics.

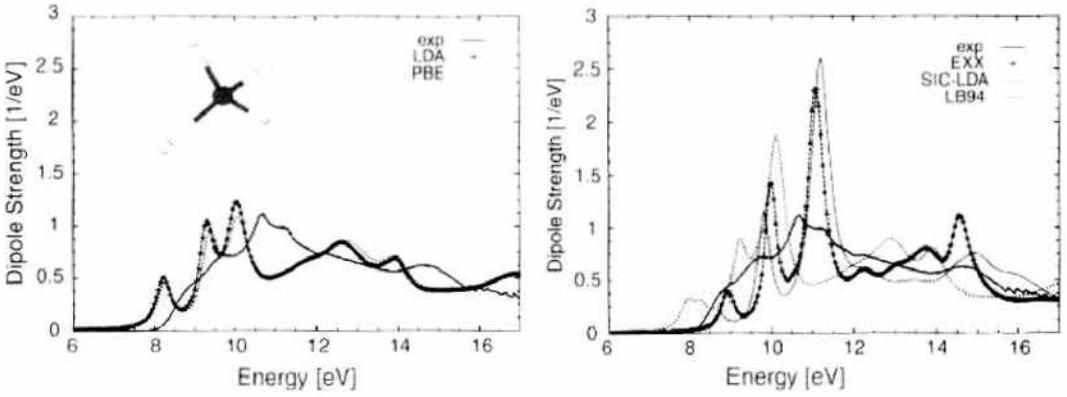


Figure 23. Averaged dipole strength of SiH_4. The experimental curve (exp) is from Itoh [183]. PBE indicates the results of generalized gradient approximation calculations with the functional of Perdew, Burke, and Ernzerhof [177]. The other curves (LDA, EXX, SIC-LDA, and LB94) correspond to calculations with functionals explained in the text. The geometrical structure of SiH_4 is shown in the inset. Reprinted with permission from [145], M. A. L. Marques, A. Castro, and A. Rubio, *J. Chem. Phys.* 115, 3006 (2001). © 2001, American Institute of Physics.

experiment. However, the SIC spectrum (given in the lower panel) is unphysically shifted to lower energies, and its first peak is split. The shifts of the SIC, LDA, and GGA spectra could be anticipated from the low HOMO-LUMO gaps. The LB94 and EXX functionals behave quite well: The onset of absorption is now correct and the error in the position of the first three peaks is reduced by a factor of two from the LDA or GGA results. The LDA, GGA, LB94, and EXX spectra of disilane (Si_2H_6) are all very similar and consist of five peaks in the interval 7–12 eV, followed by a broader feature at higher energies. The overall comparison with experiment is slightly better than for silane. The SIC-LDA functional yielded again a quite unreasonable spectrum.

20. NONLINEAR PROCESSES

20.1. Clusters in Strong Laser Fields

Progress in laser technology has opened new lines of research in the domain of nonlinear cluster dynamics. Lasers offer an ideal tool for spanning various dynamical regimes, ranging from the linear regime with plasmon-dominated dynamics (already discussed in previous sections) to the semilinear regime of multiphoton absorption processes [184, 185] and to the strongly nonlinear regime of Coulomb explosion [186, 187]. From the theoretical side, only theories based on DFT have been able to deal with such different situations and dynamical regimes for clusters. Those calculations [188] have exploited the numerical experience acquired in nuclear physics.

Let us sketch the various steps appearing in the response of highly excited metal clusters. In addition to the irradiation by intense femtosecond laser pulses [189], one can consider another class of rapid, intense, excitations: collisions with energetic, highly charged ions [190]. In both cases, the excitation takes place in times between tens of femtoseconds down to below 1 fs. This time is directly comparable to characteristic timescales of the valence electron cloud, and consequently the cluster response is primarily of electronic nature. The first phase of the reaction is a direct emission of electrons and an oscillation of the collective plasmon. This first phase is characterized by timescales on the order 1–10 fs. In a second stage, still of purely electronic nature, damping of the collective electronic motion takes place, both by means of Landau-like damping (i.e., excitation of resonances in the continuum) and by electron–electron collisions. The timescales associated to these effects are variable depending on the cluster size (Landau-like damping) and the deposited excitation energy (electron–electron collisions). Landau-like damping takes 10–20 fs and collisional effects around 10–100 fs. After that, the electronic degrees of freedom slowly couple to the ionic motion and may lead to the explosion of the charged cluster on long times of several hundred femtoseconds. Two mechanisms operate here: the first one is the coulombic repulsion resulting from the net charge of the cluster following ionization, and the second

one corresponds to energy exchanges between the hot electron cloud and the still-cold ions. The two effects interfere constructively to activate ionic motion and to lead to evaporation, fission, or fragmentation. Thermal evaporation of electrons proceeds on a very long timescale, usually slower than ionic processes like monomer evaporation and fragmentation. It can become competitive in the 100 fs range only for very hot clusters.

Two examples are now considered, to illustrate the various stages of the excitation and response of metal clusters in the nonlinear regime. The first example places the focus on the electronic response. Figure 24 represents the first stage in the response of Na_{93}^{+} irradiated by an intense laser pulse [134]; that is, the electronic response. The cluster structure is treated in the jellium approximation and the laser pulse is modeled by a ramp pulse (trapezium shape) with a total duration of 100 fs. The intensity is $I = 1010$ W/cm^2, and the photon frequency, $\hbar\omega = 3.1$ eV, is slightly above the Mie resonance for this cluster. The response depends crucially on the actual laser frequency [188, 191]. For laser frequencies sufficiently far away from the plasmon resonance, the dipole response follows closely the pulse profile and disappears when the laser pulse profile vanishes. On the contrary, for laser frequencies close to the Mie resonance, the laser may attach the resonance. This results in a sizable electron emission, and the dipole response survives the laser pulse, as it generates a true eigenfrequency of the system. The example considered in Fig. 24 corresponds to a situation in which the plasmon actually comes into play during the process. For the first 50 fs, the laser pulse remains above resonance and the electronic dipole moment $D(t)$ follows the profile of the pulse (see the upper panel in the figure). The laser intensity is still enough to induce the ionization of the cluster. The charge deficit makes the electronic cloud of the cluster more compressed (the ionic background does not change in the model), and this shifts the plasmon resonance toward higher frequencies, and thus closer to the laser frequency. From about 50 fs on, the Mie plasmon couples resonantly with the laser, and this leads to a substantial increase in ionization. The process reaches a peak until the violent electron emission produces a

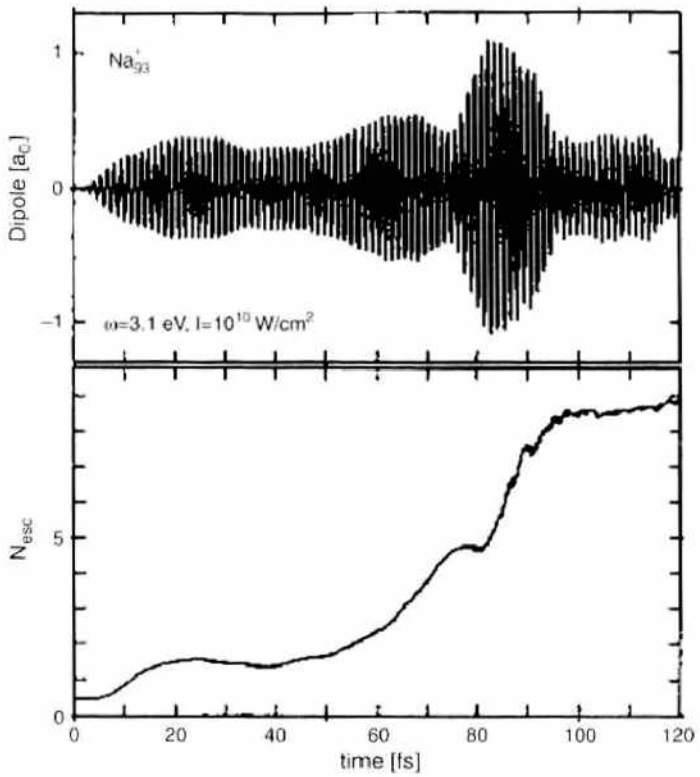


Figure 24. Electronic response of Na_{93}^{+} as a function of time (femtoseconds) to a 100 fs laser pulse of peak intensity 10^{10} W/cm^2. The upper panel gives the dipole moment (in atomic units) along the axis of laser polarization, and the lower panel gives the number of emitted electrons. Reprinted with permission from [134], J. A. Alonso, M. Barranco, F. Garcías, P. G. Reinhard, and E. Suraud, in "Fission Dynamics of Atomic Clusters and Nuclei" (J. da Providencia, D. M. Brink, F. Karpechine, and F. B. Malik, Eds.), p. 163, World Scientific, Singapore 2001. © 2001, World Scientific.

further blueshift of the plasmon frequency that detunes the plasmon from the laser. Still, even after the pulse has been switched off and electron emission has leveled off, the electron cloud continues performing collective oscillations at the actual plasmon frequency of the system (i.e., the frequency is consistent with the net charge of the cluster).

This example illustrates the role of the plasmon resonance in triggering ionization. For the moderately short pulses considered, the ion positions remain practically frozen and do not interfere with the ionization process. However, this is not the case when longer pulses are considered. Indeed, experiments [187] in platinum clusters indicate that the highly charged cluster rapidly undergoes a Coulomb expansion with a timescale of 100–500 fs. An interference can thus occur between the laser pulse and the ionic motion, which may enhance the ionization.

A second example, the excitation of Na_{41}^{+} subjected to a long laser pulse (240 fs) of frequency $\hbar\omega_{laser} = 2.86$ eV, shows how the ionic motion can interfere with the excitation process. The excitation of this cluster has been simulated using the TDLDA [192]. The results are presented in Fig. 25. An explicit account of the ions is needed, and the ion–electron

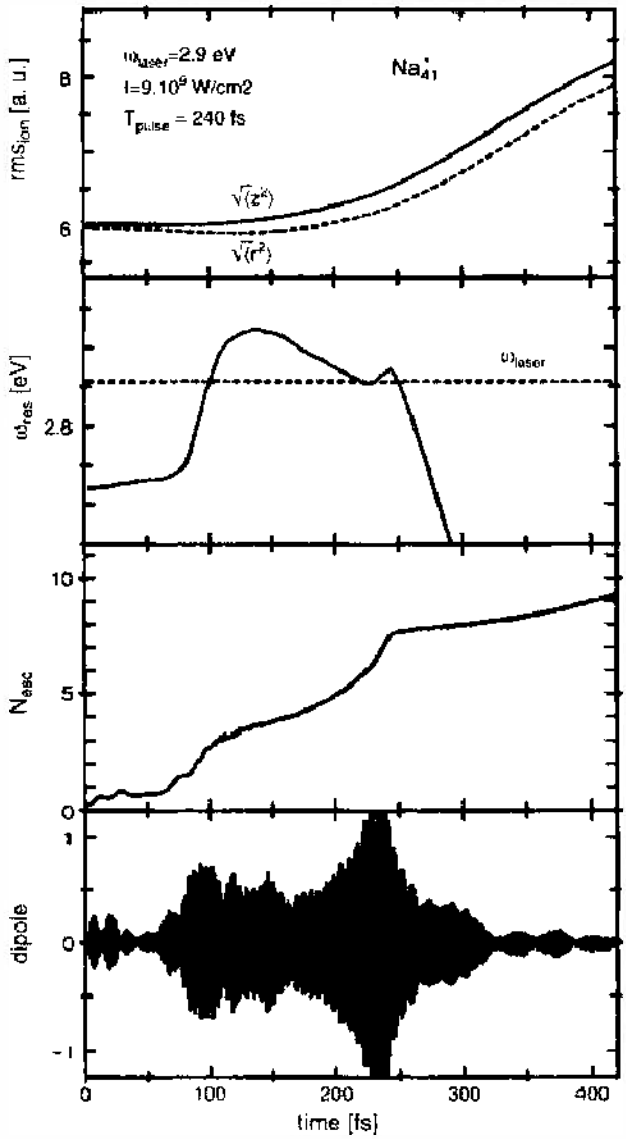


Figure 25. Time-dependent local density approximation simulation of the excitation of Na_{41}^{+} with a laser of frequency $\hbar\omega_{laser} = 2.86$ eV, intensity $I = 9 \times 10^{9}$ W/cm^2, and pulse length 240 fs. From top to bottom: global extension of the ionic distribution in z-(along laser polarization) and axial r-direction (transverse to laser polarization), average resonance frequency $\omega_{res}(t)$ for the actual structure and charge state, number of emitted electrons N_{esc} and dipole signal. Reprinted with permission from [192], E. Suraud and P. G. Reinhard, *Phys. Rev. Lett.* 85, 2296 (2000). © 2000, American Physical Society.

interaction was modeled with pseudopotentials. The third panel, giving the number of electrons emitted, N_{esc}, shows that ionization takes place in several steps. Again, in a first phase, lasting for about 80 fs, the response is fully electronic and is characterized by a low ionization, but the net charge of the cluster shifts the plasmon resonance upward until it comes into resonance with the laser. This results in a sudden increase in ionization around 100 fs, leaving the cluster in a state with a net charge 5+. Up to that stage, the situation is similar to the previous case of Na_{93}^{+}. From then on, ionization proceeds at a slower pace until another burst of electrons shows up around 250 fs, again stripping about five electrons. The lowest panel of the figure gives the electric dipole signal $D(t)$. It is clear that large slopes in ionization (N_{esc}) are correlated with large dipole amplitudes, which again reflects resonant conditions.

A relation between the two observables is observed by plotting in the second panel the instantaneous plasmon frequency $\omega_{res}(t)$, calculated at each time t for the instantaneous structure and charge of the cluster. The laser frequency is the dashed horizontal line in the same panel. The correlation between large slopes in N_{esc} and resonant conditions is noticeable. The first coincidence at time 100 fs reflects the blueshift of the plasmon resulting from the first stage of ionization and corresponds to an electronic effect. It also triggers the time at which a sizeable Coulomb expansion of the ionic distribution starts (see also uppermost panel), and it is noteworthy that this occurs rather soon (about 50 fs) after the violent initial charging. The Coulomb expansion in turn leads to a red shift of the resonance [see Eq. (9)], which is responsible for the second coincidence at 230 fs. The system thus acquires a much higher charge state and ends up in a violent Coulomb explosion.

Other nonlinear phenomena related to the coupled electron–ion dynamics correspond to the selective dissociation of molecules, isomerization, and the possibility of having an inverse Landau damping process; that is, an electron-hole excitation decaying by emitting a plasmon [193].

20.2. High Harmonic Generation

By irradiating an atom, a molecule, or a surface with a high-intensity laser, an electron may absorb several photons and then return to its original state by emitting a single photon. The emitted photon will have a frequency that is a multiple number of the laser frequency. This process is known as high harmonic generation. Because the emitted high energy photons maintain a high coherence, they can be used as a source for X-ray lasers. Figure 26 shows the harmonic spectrum of the helium atom. The experimental data, from Miyazaki and Sakai [194], is represented by the squares. The solid line gives the theoretical results [195] obtained from the TDDFT using the EXX functional. The spectrum shows a series of peaks that

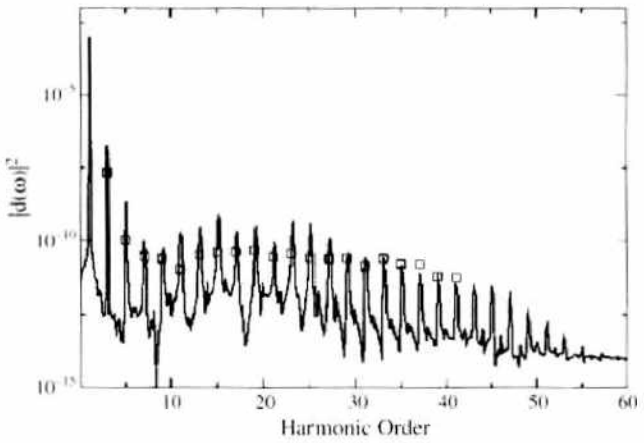


Figure 26. Calculated harmonic spectrum for He at $\lambda = 616$ nm and intensity equal to 3.5×10^{14} W/cm^2. The squares represent the experimental data [194], normalized to the value of the thirty-third harmonic of the calculated spectrum. Reprinted with permission from [195], C. A. Ullrich, S. Erhard, and E. K. U. Gross, in "Super Intense Laser Atom Physics (SLAP IV)" (H. G. Muller and M. A. Fedorov, Eds.). Kluwer, Amsterdam, 1996. © 1996, Kluwer.

first decrease in amplitude until a plateau is reached that extends to very high frequency. The frequencies of the peaks are odd multiples of the laser frequency. The even multiples are forbidden by symmetry. All theoretical approaches based on perturbation theory would produce a harmonic spectrum that decays exponentially. TDDFT, however, reproduces well the measured intensities.

High harmonic generation can, as indicated above, be used as a source of soft X-ray lasers. To this end, one should optimize the laser parameters, frequence, intensity, and so forth to increase the intensity of the emitted harmonics and to extend the plateau as far as possible. TDDFT "computer experiments" can be helpful for this task. Erhard and Gross [196] have modeled the irradiation of H atoms with lasers of a common frequency and different intensities. They found that the amplitude of the harmonics increases as the intensity of the laser increases, until a maximum amplitude is reached for a given intensity. Beyond that laser intensity, the aplitude of the harmonics decreases because the probability for ionization becomes substantial.

21. MANY-BODY TECHNIQUES: GW QUASIPARTICLES

We have seen above (Section 5) that the eigenvalues of the Kohn–Sham equations of DFT cannot be interpreted as electron removal energies. A successful development to compute excitations in electronic systems has been a first-principles self-energy approach in which the quasiparticle (qp) energies are determined directly by calculating the contribution of the dynamical polarization of the surrounding electrons. There are no parameters in this theory, and the qp energies and wavefunctions are determined by solving a Schrödinger-like equation

$$(T+V_{\text{ext}}+V_H)\psi_i^{\text{qp}}(\mathbf{r})+\int d\mathbf{r}'\Sigma(\mathbf{r},\mathbf{r}';E_i)\psi_i^{\text{qp}}(\mathbf{r}')=E\psi_i^{\text{qp}}(\mathbf{r}) \tag{98}$$

where T is the kinetic energy operator, V_{ext} is the external potential resulting from the ions, V_H is the Hartree potential of the electrons, and Σ is the self-energy operator where all the many-body exchange and correlation effects are included. The self-energy operator describes an effective potential on the qp resulting from the response of the other electrons in the system. In general, Σ is nonlocal, energy-dependent, and non-Hermitian, with the imaginary part giving the lifetime of the quasiparticles. In practical implementations (the so-called GW approximation [197–199]), the self-energy operator is taken to be the first order term in a series expansion in terms of the screened Coulomb interaction W and the dressed Green's function G of the electron as

$$\Sigma(\mathbf{r},\mathbf{r}';E)=\frac{i}{2\pi}\int d\omega e^{-i\delta\omega}G(\mathbf{r},\mathbf{r}';E-\omega)W(\mathbf{r},\mathbf{r}';\omega) \tag{99}$$

where δ is a positive infinitesimal. Vertex corrections are not included in this approximation [197–199]. The quasiparticle energy is usually computed within the framework of DFT as

$$E_i=\varepsilon_i^{\text{DFT}}+\langle\psi_i^{\text{DFT}}|\Sigma(E_i)-V_{\text{xc}}^{\text{DFT}}|\psi_i^{\text{DFT}}\rangle \tag{100}$$

where $V_{\text{xc}}^{\text{DFT}}$ is the XC potential within DFT and $\varepsilon_i^{\text{DFT}}$ and ψ_i^{DFT} are the corresponding orbital energies and wavefunctions. Two major components of the theory are the time-ordered one-electron Green's function

$$G(\mathbf{r},\mathbf{r}';E)=\sum_i\frac{\psi_i(\mathbf{r})\psi_i^*(\mathbf{r}')}{E-E_i-i\delta} \tag{101}$$

($\delta=0^+$ for $E<\mu$, and $\delta=0^-$ for $E>\mu$, where μ is the Fermi energy) and the dynamically screened Coulomb interaction

$$W(\mathbf{r},\mathbf{r}';E)=\frac{1}{\Omega}\int d\mathbf{r}''\varepsilon(\mathbf{r},\mathbf{r}'';E)\frac{1}{|\mathbf{r}''-\mathbf{r}'|} \tag{102}$$

where Ω is the volume and $\varepsilon(\mathbf{r},\mathbf{r}'';E)$ is the time-ordered dielectric response function of the system. In general, the Hedin's procedure is a self-consistent scheme to first order in W. The general equations are (the numbers stand for spatial and temporal coordinates and v for the

bare coulomb interaction)

$$\Sigma(12)=i\int G(13)\Gamma(324)W(41)d(34) \tag{103}$$

$$W(12)=v(12)+\int v(13)P(34)W(42)d(34) \tag{104}$$

$$P(12)=-i\int G(13)G(41)\Gamma(324)d(34) \tag{105}$$

$$\Gamma(123)=\delta(12)\delta(13)+\int\frac{\delta\Sigma(12)}{\delta G(45)}G(46)G(75)\Gamma(673)d(4567) \tag{106}$$

The polarizator $P(12)$ is the time-ordered counterpart of the retarded response functions. The main idea of the GW approach is to start the iteration of these equations neglecting vertex corrections; that is, $\Gamma(123)$ is assumed to be a diagonal operator. In this way, the polarizator and self-energy operators read as $P=iGG$ and $\Sigma=iGW$, explaining why the approximation is called GW. Adding on top of these equations the Dyson equation for the Green function

$$G=G^{\mathrm{DFT}}+G^{\mathrm{DFT}}\Sigma G \tag{107}$$

where G^{DFT} is the DFT Green's function obtained from the Kohn–Sham orbital energies and wavefunctions, we can see that, even at the GW level, we have a many-body self-consistent problem. Most GW applications do this self-consistent loop by varying the energy of the quasiparticle but keeping the wavefunction fixed equal to the DFT wavefunction.

Using standard many-body techniques, one can write the XC energy and get from it the XC potential. In this way, we can extract an exact equation for the XC potential in terms of the XC part of the self-energy operator [200]

$$\int d\mathbf{r}_1 V_{\mathrm{xc}}(\mathbf{r}_1)\int dE G_0(\mathbf{r},\mathbf{r}_1,E)G(\mathbf{r}_1,\mathbf{r},E)=\iint d\mathbf{r}_1 d\mathbf{r}_2\int dE G_0(\mathbf{r},\mathbf{r}_1,E)\Sigma_{\mathrm{xc}}(\mathbf{r}_1,\mathbf{r}_2,E)G(\mathbf{r}_2,\mathbf{r},E)$$

If only the exchange term is kept in Σ and G is replaced by G_0 we obtain the local potential in the Hartree–Fock approximation. When applying this equation to bulk Si, it is found that the overlap between the quasiparticle wavefunctions and the LDA-wavefunction is higher than 99%, which justifies choosing the DFT Green's functions in the GW procedure. However, this agreement is less satisfactory for II–VI or III–V semiconductors with semicore levels close to the valence complex and for transition metal systems. In those cases, it is necessary to iterate the GW equations to get good quasiparticle energies and wavefunctions, and from these the corresponding XC potential. The formalism can be extended to the time-dependent domain to derive correlation contributions to the time-dependent XC potential [201]. The inclusion of vertex corrections, that is, the inclusion of Γ for P, can be achieved through a second iteration of Hedin's equations. For a review comparing many-body approaches and TDDFT schemes, see Ref. [31].

The quasiparticle energies of small closed-shell sodium and potassium clusters have been calculated [33] using a spherical jellium model description of the cluster structure. The results for Na_{20} are compared in Table 6 with the DFT energy eigenvalues. The quasiparticle

Table 6. Calculated quasiparticle energies of Na_{20} in the spherical jellium model [33] compared with Kohn–Sham local density approximation (LDA) [33] and weighted density approximation (WDA) [32] energy eigenvalues.

	GW	DFT-LDA	DFT-WDA
E_{1s}	-5.8 ± 0.4	-5.1	-5.8
E_{1p}	-5.2 ± 0.4	-4.4	-5.1
E_{1d}	-4.4 ± 0.1	-3.5	-4.2
E_{2s}	-3.8 ± 0.05	-2.9	-3.7
E_{1f}	-1.7 ± 0.1	-2.4	

Note: DFT, density-functional theory.

energies of the occupied states (shells $1s, 1p, 1d$ and $2s$) are more negative than the LDA eigenvalues. The opposite occurs for the quasiparticle energy E_{1f} of the lowest unoccupied shell, and consequently the energy gap is larger in the GW calculation. The result found for bulk semiconductors and insulators is similar [202]. As expected, the absolute value of the highest occupied quasiparticle energies of the Na and K clusters are closer to the experimental ionization potentials than the LDA Kohn–Sham eigenvalues [33]. It is useful to notice that a DFT calculation using the nonlocal WDA for XC effects has been able to give Kohn–Sham energy eigenvalues of occupied states in very good agreement with the quasiparticle energies [32]. The WDA results are given in the last column of Table 6. The description of the ionization potentials of the Na and K clusters, as approximated by the absolute value of the HOMO eigenvalue, becomes also substantially improved with respect to the LDA. The improvement is the result of the more accurate account of the asymptotic part of the XC potential.

The results of a GW calculation of the optical spectrum of Na_4 by Onida et al. [181] including the rhombic structure of the cluster are given in Fig. 22. The LDA energy eigenvalues and orbitals form the input for the evaluation of the RPA-screened Coulomb interaction W and the self-energy Σ. The quasiparticle energies are then evaluated in first-order perturbation theory in $\Sigma - V_{xc}$. The GW corrections to the energies of the empty states are between 0.75 and 0.9 eV. The HOMO and the LUMO are lowered by 1.55 and 1.4 eV, respectively. This leads to a HOMO-LUMO gap of 3.0 eV, to be compared to the small LDA gap of 0.55 eV. These large corrections are expected for a system in which screening is weak. The calculation of Onida et al. also includes excitonic effects, which is the binding interaction between the electron and hole quasiparticles that occurs when electron-hole pairs are created, as in optical excitations. Starting with the quasiparticle energies, the excitonic effects are computed from an effective two-particle interaction, which includes the screened electron-hole interaction and also an unscreened exchange term [203]. These excitonic effects are expected to be large when screening is weak, and in fact, absorption occurs in Na_4 at energies substantially smaller than the quasiparticle gap. In summary, the figure reveals good agreement with experiment for both the positions (within 0.2 eV) and the relative oscillator strengths of the main peaks of the photoabsorption spectrum. The final results are in similar agreement with experiment as the calculations based on TDLDA discussed in Section 19.1, and Fig. 22 puts in evidence this comparison.

ACKNOWLEDGMENTS

Work supported by McyT of Spain (Grants MAT2001-0946 and MAT2002-04499-C02-01), Junta de Castilla y León (Grant CO01/102), and EC through the Research Training Networks NANOPHASE (HPRN-CT-2000-00167) and COMELCAN (HPRN-CT-2000-00128). AR also thanks the support of UPV (Grant UPV 00206.215-1369/2001), and JAA acknowledges the hospitality and support of Donostia International Physics Center. The authors have benefited from fruitful discussions and collaborations with A. Castro, M. A. L. Marques, G. F. Bertsch, K. Yabana, S. G. Louie, E. K. U. Gross, G. Onida, L. Reinig, P. M. Echenique, A. Marini, X. López, D. Varsano, M. J. López, M. J. Stott, M. Barranco, and F. Garcias.

REFERENCES

1. R. M. Dreizler and E. K. U. Gross, "Density Functional Theory. An Approach to the Quantum Many Body Problem." Springer, Berlin, 1990; R. G. Parr and W. Yang, "Density Functional Theory of Atoms and Molecules." Oxford University Press, Oxford, 1989.
2. M. A. L. Marques and E. K. U. Gross, in "A Primer in Density Functional Theory" (C. Fiolhais, F. Nogueira, and M. Marques, Eds.), Lecture Notes in Physics 620, p. 144, Springer, Berlin, 2003.
3. G. Onida, L. Reinig, and A. Rubio, *Rev. Mod. Phys.* 74, 601 (2002).
4. K. Selby, M. Vollmer, J. Masui, V. Kresin, W. A. de Heer, and W. D. Knight, *Phys. Rev. B* 40, 5417 (1989).
5. W. A. de Heer, K. Selby, V. Kresin, J. Masui, M. Vollmer, A. Chatelain, and W. D. Knight, *Phys. Rev. Lett.* 59, 1805 (1987).
6. M. Brack, *Rev. Mod. Phys.* 65, 677 (1993).

7. G. F. Bertsch and R. A. Broglia, "Oscillations in Finite Quantum Systems," Cambridge University Press, Cambridge, 1994.
8. M. Goldhaber and E. Teller, *Phys. Rev.* 74, 1046 (1948).
9. G. Mie, *Ann. Phys. (Leipzig)* 25, 377 (1908).
10. O. Bohigas, A. M. Lane, and J. Martorell, *Phys. Reports* 51, 267 (1979).
11. D. Bohm and D. Pines, *Phys. Rev.* 92, 609 (1953).
12. D. J. Thouless, *Nucl. Phys.* 22, 78 (1961).
13. W. A. de Heer, *Rev. Mod. Phys.* 65, 611 (1993).
14. M. Brack, *Phys. Rev. B* 39, 3533 (1989).
15. A. Rubio, L. C. Balbás, L. Serra, and M. Barranco, *Phys. Rev. B* 42, 10950 (1990).
16. S. Lundqvist and N. H. March, Eds., "Theory of the Inhomogeneous Electron Gas," Plenum Press, New York, 1983.
17. J. A. Alonso and L. A. Girifalco, *Phys. Rev. B* 17, 3735 (1978).
18. J. H. Parks and S. A. McDonald, *Phys. Rev. Lett.* 62, 2301 (1989).
19. L. Serra, F. Garcías, M. Barranco, J. Navarro, L. C. Balbás, A. Rubio, and A. Mananes, *J. Phys. Cond. Matter* 1, 10391 (1989).
20. L. Serra, F. Garcías, N. Barberán, M. Barranco, J. Navarro, and A. Rubio, *Z. Phys. D* 19, 89 (1991).
21. P. Hohenberg and W. Kohn, *Phys. Rev.* 136, B864 (1964).
22. W. Kohn and L. J. Sham, *Phys. Rev.* 140, A1133 (1965).
23. E. K. U. Gross, J. Dobson, and M. Petersilka, in "Density Functional Theory II" (R. F. Nalewajski, Ed.), Topics in Current Chemistry, Vol. 181, p. 81, Springer, Berlin, 1986.
24. M. J. Stott and E. Zaremba, *Phys. Rev. A* 21, 12 (1980).
25. A. Zangwill and P. Soven, *Phys. Rev. A* 21, 1561 (1980).
26. A. Rubio, J. A. Alonso, X. Blase, and S. G. Louie, *Int. J. Mod. Phys. B* 11, 2727 (1997).
27. E. Runge and E. K. U. Gross, *Phys. Rev. Lett.* 52, 997 (1984).
28. C. O. Almbladh and U. von Barth, *Phys. Rev. B* 31, 3231 (1985).
29. J. P. Perdew and M. Levy, *Phys. Rev. B* 56, 16021 (1997).
30. D. P. Chong, O. V. Gritsenko, and E. J. Baerends, *J. Chem. Phys.* 116, 1760 (2002).
31. L. C. Balbás, J. A. Alonso, and G. Borstel, *Z. Phys. D* 6, 219 (1987).
32. L. C. Balbás, J. A. Alonso, and A. Rubio, *Europhys. Lett.* 14, 323 (1991).
33. S. Saito, S. B. Zhang, S. G. Louie, and M. L. Cohen, *Phys. Rev. B* 40, 3643 (1989).
34. R. O. Jones and O. Gunnarsson, *Rev. Mod. Phys.* 61, 689 (1989).
35. M. Petersilka, U. J. Gossmann, and E. K. U. Gross, *Phys. Rev. Lett.* 76, 1212 (1996).
36. A. L. Fetter and J. D. Walecka, "Quantum Theory of Many-Body Systems," Mc Graw Hill, New York, 1971.
37. T. Grabo, M. Petersilka, and E. K. U. Gross, *J. Molec. Struct. (Theochem)* 501, 353 (2000).
38. M. Casida, in "Recent Developments and Applications in Density Functional Theory" (J. Seminario, Ed.), p. 391, Elsevier, Amsterdam, 1996.
39. I. Vasiliev, S. Ögut, and J. R. Chelikowsky, *Phys. Rev. Lett.* 82, 1919 (1999).
40. E. U. K. Gross and W. Kohn, in "Advances in Quantum Chemistry" (S. B. Trickey, Ed.), Academic Press, San Diego, 1990.
41. J. B. Krieger, Y. Li, and G. J. Iafrate, *Phys. Rev. A* 45, 101 (1992).
42. C. A. Ullrich, U. Gossmann, and E. K. U. Gross, *Phys. Rev. Lett.* 74, 872 (1995).
43. Y. H. Kim and A. Görling, *Phys. Rev. Lett.* 89, 096402 (2002); *Phys. Rev. B* 66, 035114 (2002).
44. S. Kurth and U. Von Barth (unpublished).
45. C. E. Moore, *Nat. Stand. Ref. Data Ser. (U.S. Nat. Bur. Stand.)* 35 (1971).
46. E. K. U. Gross, T. Kreibig, M. Lein, and M. Petersilka, in "Electron Correlations and Material Properties" (A. Gonis, N. Kioussis, and M. Ciftan, Eds.), Plenum Press, New York, 1999.
47. S. J. A. Van Gisbergen, F. Kootstra, P. R. T. Schipper, O. V. Gritsenko, J. G. Snijders, and E. J. Baerends, *Phys. Rev. A* 57, 2556 (1998).
48. W. A. De Heer, W. D. Knight, M. Y. Chou, and M. L. Cohen, *Solid State Phys.* 40, 93 (1987).
49. A. Rubio, L. C. Balbás, and J. A. Alonso, *Phys. Rev. B* 46, 4891 (1992).
50. C. Yannouleas, R. A. Broglia, M. Brack, and P. F. Bortignon, *Phys. Rev. Lett.* 63, 255 (1989).
51. C. R. C. Wang, S. Pollack, D. Cameron, and M. M. Kappes, *J. Chem. Phys.* 93, 3787 (1990).
52. Z. Penzar, W. Ekardt, and A. Rubio, *Phys. Rev B* 42, 5040 (1990).
53. K. Selby, V. Kresin, J. Masui, M. Vollmer, W. A. de Heer, A. Scheideman, and W. Knight, *Phys. Rev. B* 43, 4565 (1991).
54. A. Rubio, L. C. Balbás, and J. A. Alonso, *Z. Phys. D* 19, 93 (1991).
55. J. A. Alonso and N. A. Cordero, in "Recent Developments and Applications of Modern Density Functional Theory" (J. M. Seminario, Ed.), Theoretical and Computational Chemistry, Vol. 4, p. 239, Elsevier, Amsterdam, 1996.
56. V. Bonacic-Koutecky, P. Fantucci, C. Fuchs, C. Gatti, J. Pittner, and S. Polezzo, *Chem. Phys. Lett.* 213, 522 (1993).
57. S. Saito, G. F. Bertsch, and D. Tomanek, *Phys. Rev. B* 43, 6804 (1991).
58. J. M. Pacheco and W. Ekardt, *Ann. Physik.* 1, 255 (1992); *Phys. Rev. B* 47, 6667 (1993).
59. J. A. Alonso, A. Rubio, and L. C. Balbás, *Philos. Mag. B* 69, 1037 (1994).
60. T. Reiners and H. Haberland, *Phys. Rev. Lett.* 77, 2440 (1996).
61. C. Brechignac, P. Cahuzac, F. Charlier, M. de Frutos, and J. Leignier, *Chem. Phys. Lett.* 189, 28 (1992).

62. J. Borggreen, P. Choudhury, N. Kebaili, L. Lundsberg-Nielsen, K. Lützenkirchen, M. B. Nielsen, J. Pedersen, and H. D. Rasmusen, *Phys. Rev. B* 48, 17507 (1993).
63. J. Tiggesbäumker, L. Köller, H. O. Lutz, K. H. and Meiwes-Broer, *Chem. Phys. Lett.* 190, 42 (1992).
64. Z. Penzar and W. Ekardt, *Z. Phys. D.* 17, 69 (1990).
65. W. Ekardt and Z. Penzar, *Phys. Rev. B* 43, 1322 (1991).
66. G. Lauritsch, P. G. Reinhard, J. Meyer, and M. Brack, *Phys. Lett. A* 160, 179 (1991).
67. S. Kasperl, C. Kohl, and P. G. Reinhard, *Phys. Lett. A* 206, 81 (1995).
68. M. Koskinen, O. P. Lipas, and M. Manninen, *Z. Phys. D* 35, 285 (1995).
69. D. L. Hill and J. A. Wheeler, *Phys. Rev.* 89, 1102 (1953).
70. V. Bonacic-Koutecky, P. Fantucci, and J. Koutecky, *Chem. Rev.* 91, 1035 (1991).
71. U. Röthlisberger and W. Andreoni, *J. Chem. Phys.* 94, 8129 (1991).
72. C. Kohl, B. Montag, and P. G. Reihnard, *Z. Phys. D* 35, 57 (1995).
73. B. Montag and P. G. Reihnard, *Phys. Rev. B* 51, 14686 (1995).
74. B. Montag, P. G. Reihnard, and J. Meyer, *Z. Phys. D* 32, 125 (1994).
75. T. Reiners, W. Orlik, Ch. Ellert, M. Schmidt, and H. Haberland, *Chem. Phys. Lett.* 215, 357 (1993).
76. A. Rubio, L. C. Balbás, and J. A. Alonso, *Z. Phys. D* 26, 284 (1993).
77. M. P. Iñiguez, J. A. Alonso, and L. C. Balbás, *Solid State Commun.* 57, 85 (1986).
78. M. P. Iñiguez, M. J. López, J. A. Alonso, and J. M. Soler, *Z. Phys. D* 11, 163 (1989).
79. I. A. Solovyov, A. V. Solovyov, and W. Greiner, *Phys. Rev. A* 65, 053203 (2002).
80. J. A. Alonso, L. C. Balbás, and A. Rubio, *Int. J. Quantum Chem.* 56, 499 (1995).
81. H. Fallgreen and T. P. Martin, *Chem. Phys. Lett.* 168, 233 (1990).
82. H. Fallgreen, K. M. Brown, and T. P. Martin, *Z. Phys. D* 36, 65 (1996).
83. A. Rubio, L. C. Balbás, and J. A. Alonso, *Phys. Rev. B* 45, 13657 (1992).
84. M. Manninen, *Phys. Rev. B* 34, 6886 (1986).
85. W. D. Schöne, W. Ekardt, and J. M. Pacheco, *Phys. Rev. B* 50, 11079 (1994).
86. W. D. Schöne, W. Ekardt, and J. M. Pacheco, *Z. Phys. D* 36, 65 (1996).
87. A. Rubio, J. A. Alonso, J. M. López, and M. J. Stott, *Physica B* 183, 247 (1993).
88. C. R. Wang, S. Pollack, T. A. Dahlseid, and G. M. Korestky, *J. Chem. Phys.* 96, 7931 (1992).
89. V. Bonacic Koutecky, J. Pittner, C. Scheuch, M. F. Guest, and J. Koutecky, *J. Chem. Phys.* 96, 7938 (1992).
90. M. Kappes, M. Schär, C. Yeretzian, U. Heiz, A. Vayloyan, and E. Schumacher, in "Physics and Chemistry of Small Clusters" (P. Jena, B. K. Rao, and S. N. Khanna, Eds.), NATO ASI Series B, Vol. 158, p. 263. Plenum Press, New York, 1987.
91. C. Yeretzian, *J. Phys. Chem.* 99, 123 (1995).
92. C. Baladrón and J. A. Alonso, *Physica B* 154, 73 (1988).
93. C. Yannouleas, P. Jena, and S. N. Khanna, *Phys. Rev. B* 46, 9751 (1992).
94. L. C. Balbás, A. Rubio, and M. B. Torres, *Z. Phys. D* 31, 269 (1994).
95. C. Yannouleas and R. Broglia, *Europhys. Lett.* 15, 843 (1991).
96. J. A. Alonso and N. H. March, "Electrons in Metals and Alloys." Academic Press, London, 1989.
97. N. Q. Lam, H. Hoff, H. Wiedersich, and L. E. Rehn, *Surf. Sci.* 149, 517 (1985).
98. A. Bol, G. Martín, J. M. López, and J. A. Alonso, *Z. Phys. D* 28, 311 (1993).
99. H. Ajie, M. M. Alvarez, S. J. Anz, R. D. Beck, F. Diederich, K. Fostiropoulos, D. R. Huffman, W. Kratschmer, Y. Rubin, K. E. Schriver, D. Sensharma, and R. L. Whetten, *J. Phys. Chem.* 94, 8630 (1990).
100. J. W. Keller and M. A. Coplan, *Chem. Phys. Lett.* 193, 89 (1992).
101. J. R. Heat, R. F. Curl, and R. E. Smalley, *J. Chem. Phys.* 87, 4236 (1987).
102. I. V. Hertel, H. Steger, J. De Vries, B. Weiser, C. Menzel, B. Kamke, and W. Kamke, *Phys. Rev. Lett.* 68, 784 (1992).
103. S. Leach, M. Vervloet, A. Despres, E. Bréheret, J. Hare, T. J. Dennis, H. W. Kroto, R. Taylor, and D. R. M. Walton, *Chem. Phys. Lett.* 160, 451 (1992).
104. J. M. Hawkins, A. Meyer, T. A. Lewis, S. Loren, and F. J. Hollander, *Science* 252, 312 (1991).
105. A. Rubio, J. A. Alonso, J. M. López, and M. J. Stott, *Phys. Rev. B* 49, 17397 (1994).
106. S. Frank, N. Malinowski, F. Tast, M. Heinebrodt, I. M. Billas and T. P. Martin, *Z. Phys. D* 40, 250 (1997).
107. H. N. Rojas, R. W. Godby, and R. J. Needs, *Phys. Rev. Lett.* 74, 1827 (1995).
108. X. Blase, A. Rubio, M. L. Cohen, and S. G. Louie, *Phys. Rev. B* 52, 2225 (1995).
109. W. Yang, *Phys. Rev. Let.* 66, 1438 (1991).
110. G. Galli and M. Parrinello, *Phys. Rev. Lett.* 69, 3547 (1992).
111. P. Ordejón, D. A. Drabold, M. P. Grumbach, and R. M. Martin, *Phys. Rev. B* 48, 16464 (1993).
112. W. Kohn, *Phys. Rev. Lett.* 76, 3168 (1996).
113. D. Sánchez-Portal, P. Ordejón, E. Artacho, and J. M. Soler, *Int. J. Quantum Chem.* 65, 453 (1997).
114. A. Rubio, J. A. Alonso, X. Blase, L. C. Balbás, and S. G. Louie, *Phys. Rev. Lett.* 77, 247 (1996); Erratum 77, 5442 (1996).
115. J. Blanc, V. Bonacic-Koutecky, M. Broyer, J. Chevaleyre, Ph. Dugourd, J. Koutecky, C. Scheuch, J. P. Wolf, and L. J. Wöste, *Chem. Phys.* 96, 1793 (1992).
116. N. Binggeli and J. R. Chelikowsky, *Phys. Rev. Lett.* 75, 493 (1995).
117. J. R. Chelikowsky, N. Troullier, and Y. Saad, *Phys. Rev. Lett.* 72, 1240 (1994).
118. E. C. Honea, A. Ogura, C. A. Murray, K. Raghavachari, W. O. Sprenger, M. F. Jarrold, and W. L. Brown, *Nature* 366, 42 (1993).
119. A. G. Cullis and L. T. Canham, *Nature* 353, 335 (1991).

120. ... R. C. Wang, S. Pollack, D. Cameron, and M. M. Kappes, *Chem. Phys. Lett.* 165, 26 (1990).
121. ... P. Sinha, *Proc. Phys. Soc. London* 62, 124 (1949).
122. A. R. Fredrickson and W. W. Watson, *Phys. Rev.* 30, 429 (1927).
123. ... Bonacic-Koutecky, J. Pittner, C. Fuchs, P. Fantucci, and J. Koutecky, *J. Chem. Phys.* 104, 1427 (1996).
124. ... M. Pacheco and J. L. Martins, *J. Chem. Phys.* 106, 6039 (1997).
125. G. B. Bachelet, D. R. Hamman, and M. Schüller, *Phys. Rev. B* 26, 4199 (1982).
126. N. Troullier and J. L. Martins, *Phys. Rev. B* 43, 1993 (1991).
127. ... L. Martins, J. Buttet, and R. Car, *Phys. Rev.* 31, 1804 (1985).
128. M. W. Sung, R. Kawai, and J. H. Weare, *Phys. Rev. Lett.* 73, 3552 (1994).
129. S. Pollack, C. R. C. Wang, T. A. Dahlseid, and M. M. Kappes, *J. Chem. Phys.* 96, 4918 (1992).
130. V. Bonacic-Koutecky, J. Gaus, M. Guest, and J. Koutecky, *J. Chem. Phys.* 96, 4934 (1992).
131. T. A. Dahlseid, M. M. Kappes, J. A. Pople, and J. Ratner, *J. Chem. Phys.* 96, 4924 (1992).
132. M. D. Deshpande, D. G. Kanhere, P. V. Panat, I. Vasiliev, and R. M. Martin, *Phys. Rev. A* 65, 053204 (2002).
133. T. Näher, S. Bjornholm, S. Frauendorf, F. Garcias, and C. Guet, *Phys. Rep.* 285, 245 (1997).
134. ... A. Alonso, M. Barranco, F. Garcias, P. G. Reinhard, and E. Suraud, in "Fission Dynamics of Atomic Clusters and Nuclei" (J. da Providencia, D. M. Brink, F. Karpechine, and F. B. Malik, Eds.), p. 163. World Scientific, Singapore, 2001.
135. B. Montag and P. G. Reinhard, *Phys. Lett. A*, 193, 380 (1994).
136. B. Montag and P. G. Reinhard, *Phys. Rev. B* 52, 16365 (1995).
137. A. Lyalin, A. Solovyov, and W. Greiner, *Phys. Rev. A* 65, 043202 (2002).
138. H. Tee, V. A. Lobastov, U. M. Gomez, B. M. Godson, R. Srinivasan, C. Y. Ruan, and A. H. Zewail, *Science* 291, 458 (2001).
139. V. S. Sundström Ed., "Femtosecond Chemistry and Femtobiology: Ultrafast Reaction Dynamics of Atomic Scale Resolution," Nobel Symposium, Vol. 101. Imperial College Press, London, 1996.
140. J. M. Pacheco and W. D. Schöne, *Phys. Rev. Lett.* 79, 4986 (1997).
141. M. Moseler, H. Häkkinen, and U. Landman, *Phys. Rev. Lett.* 87, 053401 (2001).
142. M. Schmidt, C. Ellert, W. Kronmüller, and H. Haberland, *Phys. Rev. B* 59, 10970 (1999).
143. H. Flocard, S. E. Koonin, and M. S. Weiss, *Phys. Rev. C* 17, 1682 (1978).
144. K. Yabana and G. F. Bertsch, *Phys. Rev. B* 54, 4484 (1996).
145. M. A. L. Marques, A. Castro, and A. Rubio, *J. Chem. Phys.* 115, 3006 (2001).
146. M. A. L. Marques, A. Castro, G. F. Bertsch, and A. Rubio, *Computer Phys. Commun.* 151, 60 (2003). OCTOPUS, http://www.tddft.org/programs/octopus.
147. K. Yabana and G. F. Bertsch, *Z. Phys. D* 42, 219 (1997); *Phys. Rev. A* 58, 2604 (1999); *Phys. Rev. A* 60, 1271 (1999); *Phys. Rev. A* 60, 3809 (1999).
148. M. di Ventra and S. Pantelides, *Phys. Rev. B* 61, 16207 (2000).
149. M. di Ventra, S. Pantelides, and N. Lang, *Phys. Rev. Lett.* 88, 046801 (2002).
150. E. Brabec and F. Krausz, *Rev. Mod. Phys.* 72, 545 (2000).
151. M. Protopapas, C. H. Keitel, and P. L. Knight, *Rep. Prog. Phys.* 60, 389 (1997).
152. T. Nakatsukatsa and K. Yabana, *J. Chem. Phys.* 114, 2550 (2000).
153. N. S. Goroff, *Acc. Chem. Res.* 29, 77 (1996).
154. H. Prinzbach, A. Weiler, P. Landenberger, F. Wahl, J. Wörth, L. T. Scott, M. Gelmont, D. Olevano, and B. Von Issendorff, *Nature* 407, 60 (2000).
155. J. C. Grossmann, L. Mitas, and K. Raghavachari, *Phys. Rev. Lett.* 75, 3870 (1995).
156. R. O. Jones and G. Seifert, *Phys. Rev. Lett.* 79, 443 (1997).
157. K. Raghavachari, D. L. Strout, G. K. Odom, G. E. Scuseria, J. A. Pople, B. G. Johnson, and P. M. W. Gill, *Chem. Phys. Lett.* 214, 357 (1992).
158. E. J. Bylaska, P. R. Taylor, R. Kawai, and J. H. Weare, *J. Phys. Chem.* 100, 6966 (1996).
159. J. Lu, S. Re, Y. Choe, S. Nagase, Y. Zhou, R. Han, L. Peng, X. Zhang, and X. Zhao, *Phys. Rev. B* 67, 125415 (2003).
160. A. J. Stone and D. J. Wales, *Chem. Phys. Lett.* 28, 501 (1986).
161. R. O. Jones, *J. Chem. Phys.* 110, 5189 (1999).
162. A. Castro, M. A. L. Marques, J. A. Alonso, G. F. Bertsch, K. Yabana, and A. Rubio, *J. Chem. Phys.* 116, 1930 (2002).
163. K. Yabana and G. F. Bertsch, *Int. J. Quantum Chem.* 75, 55 (1999).
164. E. E. Koch and A. Otto, *Chem. Phys. Lett.* 12, 476 (1972).
165. T. M. H. Creemers, A. J. Lock, V. Subramaniam, T. M. Jovin, and S. Völker, *Proc. Natl. Acad. Sci. U.S.A.* 97, 2974 (2000); *Nat. Struct. Biol.* 6, 557 (1999).
166. S. B. Nielsen, A. Lapierre, J. U. Andersen, U. V. Pedersen, S. Tomita, and L. H. Andersen, *Phys. Rev. Lett.* 87, 228102 (2001).
167. M. A. L. Marques, X. López, D. Varsano, A. Castro, and A. Rubio, *Phys. Rev. Lett.* 90, 258101 (2003).
168. B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, *J. Comput. Chem.* 2, 187 (1983).
169. M. J. Field, P. A. Bash, and M. Karplus, *J. Comp. Chem.* 11, 700 (1990).
170. M. J. S. Dewar, E. Zoebisch, E. F. Healy, and J. J. P. Stewart, *J. Am. Chem. Soc.* 107, 3902 (1985).
171. D. Varsano, A. Castro, M. A. L. Marques, and A. Rubio (unpublished).
172. A. Tsolakidis, D. Sánchez-Portal, and R. M. Martin, *Phys. Rev. B* 66, 235416 (2002).
173. E. Artacho, D. Sánchez-Portal, P. Ordejón, A. Garcia, and J. M. Soler, *Phys. Status Solidi B* 215, 809 (1999).

174. P. Ordejón, *Phys. Status Solidi B* 217, 335 (2000).
175. R. Bauernschmitt, R. Ahlrichs, F. H. Hennrich, and M. M. Kappes, *J. Am. Chem. Soc.* 120, 5052 (1998).
176. A. D. Becke, *Phys. Rev. A* 38, 3098 (1988).
177. J. P. Perdew, K. Burke, and M. Ernzerhof, *Phys. Rev. Lett.* 77, 3865 (1996).
178. R. Van Leeuwen and E. J. Baerends, *Phys. Rev. A* 49, 2421 (1994).
179. M. E. Casida and D. R. Salahub, *J. Chem. Phys.* 113, 8918 (2000).
180. J. P. Perdew and A. Zunger, *Phys. Rev. B* 23, 5408 (1981).
181. G. Onida, L. Reinig, R. W. Godby, R. Del Sole, and W. Andreoni, *Phys. Rev. Lett.* 75, 818 (1995).
182. J. G. Grossman, M. Rohlfing, L. Mitas, S. G. Louie, and M. L. Cohen, *Phys. Rev. Lett.* 86, 472 (2001).
183. U. Itoh, Y. Toyoshima, H. Onuki, N. Washida, and T. Ibuki, *J. Chem. Phys.* 85, 4867 (1986).
184. H. Hohmann, C. Callegari, S. Furrer, D. Grosenick, E. E. B. Campbell, and I. V. Hertel, *Phys. Rev. Lett.* 73, 1919 (1994).
185. R. Schlipper, R. Kusche, B. von Issendorff, and H. Haberland, *Phys. Rev. Lett.* 80, 1194 (1998).
186. M. Lezius, S. Dobosz, D. Normand, and M. Schmidt, *Phys. Rev. Lett.* 80, 261 (1998).
187. L. Köller, M. Schumacher, J. Köhn, S. Teuber, J. Tiggesbäumker, and K. H. Meiwes-Broer, *Phys. Rev. Lett.* 82, 3783 (1999).
188. F. Calvayrac, P. G. Reinhard, E. Suraud, and C. Ullrich, *Phys. Reports* 337, 493 (2000).
189. T. Baumert and G. Gerber, *Adv. At. Mol. Opt. Phys.* 35, 163 (1995).
190. F. Chandezon, C. Guet, B. A. Huber, D. Jalabert, M. Maurel, E. Monnand, C. Ristori, and J. C. Rocco, *Phys. Rev. Lett.* 74, 3784 (1995).
191. C. A. Ullrich, P. G. Reinhard, and E. Suraud, *J. Phys. B.* 30, 5043 (1997).
192. E. Suraud and P. G. Reinhard, *Phys. Rev. Lett.* 85, 2296 (2000).
193. M. García, A. Castro, J. A. Alonso, and A. Rubio (unpublished).
194. K. Miyazaki and H. Sakai, *J. Phys. B* 25, L83 (1992).
195. C. A. Ullrich, S. Erhard, and E. K. U. Gross, in "Super Intense Laser Atom Physics (SLAP IV)" (H. G. Muller and M. A. Fedorov, Eds.). Kluwer, Amsterdam, 1996.
196. S. Erhard and E. K. U. Gross, in "Multiphoton Processes" (L. Lambropoulos and H. Walther, Eds.), IOP, Bristol, 1996.
197. L. Hedin, *Phys. Rev.* 139, A796 (1965).
198. L. Hedin and S. Lundqvist, in "Solid State Physics" (F. Seitz, D. Turnbull, and H. Ehrenreich, Eds.), Vol. 23, p. 1. Academic Press, New York, 1969.
199. R. W. Godby, M. Schlüter, and L. J. Sham, *Phys. Rev. Lett.* 56, 2415 (1986).
200. L. J. Sham and M. Schlüter, *Phys. Rev. Lett.* 51, 1888 (1983).
201. R. Van Leeuwen, *Phys. Rev. Lett.* 76, 3610 (1996).
202. M. S. Hybertsen and S. G. Louie, *Phys. Rev. B* 38, 5390 (1986).
203. L. J. Sham and T. M. Rice, *Phys. Rev.* 144, 1519 (1982).

CHAPTER 5

Modeling of Photonic Crystals

Wounjhang Park

Department of Electrical and Computer Engineering, University of Colorado, Colorado, USA

CONTENTS

1. INTRODUCTION

Since the first proposals by Yablonovitch [1] and John [2], photonic crystal (PC) or photonic band gap (PBG) material has quickly emerged as one of the most important classes of materials for photonic applications. PC is an artificially made material in which the dielectric constant, or the index of refraction, is periodically modulated. As a result, the electromagnetic wave undergoes Bragg scattering by the dielectric structure in much the same way as the electrons in a periodic potential are produced by a crystal. As a consequence, the dispersion relation is modified to form a photonic band structure that bears strong resemblance to the electronic energy band structure. Some structures even exhibit PBGs in which no photon modes exist, just as some crystals exhibit electronic energy band gaps in which no electronic states are present. The complete PBG, in which light propagation is prohibited in all directions, is particularly important, with extremely high technological potential, and

ISBN: 1-58883-049-7

Handbook of Theoretical and Computational Nanotechnology
Edited by Michael Rieth and Wolfram Schommers
Volume 7: Pages (263–327)

thus the efforts to identify the structures exhibiting complete PBG have been a major driving force in PC research. The technological importance of complete PBG lies primarily in the possibility of deliberately introducing defects that subsequently create localized modes within the PBG. Such defect modes would then act as waveguides or cavities, depending on their exact geometry. Because the features in PC structures have dimensions on the order of the optical wavelength, these photonic defect waveguides and cavities can be made with nanoscale geometries. Furthermore, in these structures the light is strongly confined by the photonic band structure, sharply contrasting the weak confinement with the index difference in conventional waveguides. Therefore, PCs present exciting new possibilities of realizing large-scale integrated optical devices, analogous to the large-scale integrated circuits that revolutionized the electronics and computing industry several decades ago. Since the first demonstration of near-perfect waveguiding around a sharp 90° corner in a PC waveguide [3], the research on the integrated optical devices based on PC structures has been extensive, ranging from analysis of mode structures in PC waveguides [4, 5] to design of new devices such as channel add/drop filters [6] to coupling with optical fibers [7].

In addition to controlling the propagation of light, the photonic band structure also affects the generation of light through the strongly modified photon density of states. Spontaneous emission of an atom, as the name suggests, has long been believed to be an intrinsic, unalterable property. However, spontaneous emission is, in fact, a result of interaction between atom and vacuum, and it can be greatly enhanced or suppressed if the vacuum states of photons are modified. The research on the atom–vacuum system, often referred to as the cavity quantum electrodynamics (QED), has led to the prediction and demonstration of many novel optical phenomena [8]. A hallmark application of cavity QED effects is the thresholdless laser, in which all spontaneously emitted photons are coupled to the lasing mode, and thus the light output increases linearly with the pump power. The key to achieving the thresholdless laser is to increase the spontaneous emission factor β, which is given by

$$\beta = \frac{\lambda^4}{4\pi^2 n^3 V \Delta\lambda} \tag{1}$$

where n is the refractive index, V is the mode volume, λ is the emission wavelength, and $\Delta\lambda$ is the linewidth [9]. Obviously, one can increase β by decreasing the mode volume and simultaneously increasing the cavity Q factor, which is inversely proportional to $\Delta\lambda$. Although a variety of structures such as micropillars and microdisks have been developed, the PC provides the most efficient platform for creating very small cavities with high Q factors. A recent report on a two-dimensional PC cavity showed a Q factor as high as 2800 and a mode volume as small as $0.43(\lambda/n)^3$ [10]. In addition to the strong light confinement by PBG, PCs also offer high design flexibility that allows tailoring of mode structure to yield the strongest coupling between light emitter and cavity modes [11]. It is thus believed that the PC is one of the best material platforms for high-efficiency light sources.

Another exciting phenomenon observed in PC structures is the extraordinary refraction and dispersion characteristics, which may become orders of magnitude greater than in conventional optical materials. This novel phenomenon was first reported by Lin et al. in the millimeter wave region [12]. Kosaka et al. then constructed a three-dimensional Si/SiO_2 PC structure using a self-replicating sputtering technique and observed a similar super prism effect at optical wavelengths [13]. Both of these structures possess sixfold rotational symmetry, which results in a highly anisotropic dispersion curve with corresponding sixfold rotational symmetry. This anisotropy is responsible for the giant refraction characteristics. In Ref. [13], it was reported that a ray entering the crystal with a small incident angle (7°) experiences a strong refraction with an extraordinarily large refraction angle (as high as 70°). Furthermore, the dispersion relation can also become highly nonlinear, resulting in a superprism effect in which closely spaced wavelength components are widely dispersed. It was recently reported that two wavelength components differing only by 1% could be separated by as much as 50°, an angular dispersion 500 times greater than conventional prism [14]. In addition, a PC may also exhibit a negative refractive index, causing the incident ray refracted back to the side on which it was incident. This unconventional phenomenon is a

direct consequence of the periodic index modulation in PCs, which results in strong mixing of diffracted waves with various reciprocal lattice vectors [15]. It should be noted that these unconventional refraction and dispersion properties do not require the existence of PBG, which significantly relieves material and fabrication requirements. These novel properties present an exciting possibility of achieving nanoscale optical devices that can collect, focus, disperse, switch, and steer light in a highly integrated geometry.

In all of the recent developments in PC-related research, theoretical modeling has been the most critical component. It has led the early developments, providing the basic information about the photonic band structures and the resultant optical properties. In recent years, the experimental research for realizing many novel PC structures has significantly intensified. However, the theoretical modeling study still plays a crucial role as a design tool while it continues to explore the new frontiers of the exciting new phenomena PCs might exhibit. In this review, we give an overview of the theoretical framework, numerical modeling techniques, and recent research findings of the theoretical modeling studies.

2. THEORETICAL BACKGROUND

As in any problem involving electromagnetic waves, the theoretical treatment of PC begins with the four Maxwell's equations

$$\begin{aligned} \nabla \cdot \mathbf{D} &= 4\pi\rho \\ \nabla \cdot \mathbf{B} &= 0 \\ \nabla \times \mathbf{E} + \frac{1}{c}\frac{\partial \mathbf{B}}{\partial t} &= 0 \\ \nabla \times \mathbf{H} - \frac{1}{c}\frac{\partial \mathbf{D}}{\partial t} &= \frac{4\pi}{c}\mathbf{J} \end{aligned} \tag{2}$$

The information of the medium in which the fields are present is included by the constitutive relations that relate **D** to **E** and **B** to **H**

$$\begin{aligned} \mathbf{D} &= \varepsilon\mathbf{E} \\ \mathbf{H} &= \mu\mathbf{B} \end{aligned} \tag{3}$$

Here, ε and μ may be functions of spatial coordinates and also of frequency. When dealing with nonmagnetic, lossless, and sourceless medium, which is often the case for PCs, we can simply put $\mu = 1$, $\rho = 0$, and $\mathbf{J} = 0$. This condition also gives that ε is purely real. We note that we make these assumptions for illustrative purposes. As we shall see later, the formalism can be fairly easily extended to the more general cases such as metal-dielectric PCs, where some of the above assumptions are not valid. We now employ the familiar technique of assuming harmonic time dependence for **E** and **H** fields. The two divergence equations in Eq. (2) are then simply reduced to the transversality condition, whereas the two curl equations in Eq. (2) may be transformed into a pair of wave equations

$$\begin{aligned} \nabla \times \nabla \times \mathbf{E}(\mathbf{r}) &= \varepsilon(\mathbf{r})\frac{\omega^2}{c^2}\mathbf{E}(\mathbf{r}) \\ \nabla \times \left\{\frac{1}{\varepsilon(\mathbf{r})}\nabla \times \mathbf{H}(\mathbf{r})\right\} &= \frac{\omega^2}{c^2}\mathbf{H}(\mathbf{r}) \end{aligned} \tag{4}$$

The resemblance of Eq. (4) to the Schrödinger equation is quite remarkable. In the equation for E field, for example, we may regard the second-order differential operator in the left-hand side as the kinetic energy term, while in the right-hand side $\varepsilon(\mathbf{r})$ acts as the potential energy and ω as the energy eigenvalue. This recognition is particularly important for modeling of PCs, as one may then explore the possibility of using techniques similar to those developed and widely used in solid-state physics. In fact, many modeling techniques currently used for PCs were originally developed for electronic systems and later adapted for PCs. However, there are two crucial differences that must be taken into account. First, the E and

H fields are vectors with two degrees of freedom for polarization, whereas the electronic wavefunction in Schrödinger equation is a scalar quantity. The full incorporation of the vector nature of the E and H fields is critical for accurate prediction of photonic band structure. As a matter of fact, during the early theoretical development there was some confusion about the existence of complete PBG in some structures as a result of incorrect prediction by scalar wave approximation. In addition, the dielectric constant ε appears in a product with the frequency, and thus the periodic potential and energy eigenvalue are intrinsically tangled together, contrasting the Schrödinger equation, where the two are independent.

The periodicity of PC is contained in the functional form of $\varepsilon(\mathbf{r})$, which satisfies

$$\varepsilon(\mathbf{r}) = \varepsilon(\mathbf{r} + \mathbf{R}) \tag{5}$$

where $\mathbf{R}$ is a lattice translation vector generated by the primitive lattice vectors $\mathbf{a}_i$, such that $\mathbf{R} = n_1\mathbf{a}_1 + n_2\mathbf{a}_2 + n_3\mathbf{a}_3$, where n_i's are integers. Mathematically, the periodicity may be expressed by translational symmetry; that is, the system is invariant under certain translations defined by the primitive lattice vectors. This translational symmetry then allows us to invoke Bloch's theorem and write the fields as

$$\mathbf{E}_\mathbf{k}(\mathbf{r}) = \mathbf{u}_\mathbf{k}(\mathbf{r})e^{i\mathbf{k}\cdot\mathbf{r}} \tag{6}$$

where $\mathbf{u}$ is a function with the same periodicity as the lattice, $\mathbf{u}(\mathbf{r}) = \mathbf{u}(\mathbf{r} + \mathbf{R})$. A key characteristic of the Bloch states is that the states with certain wave vectors are identical. To be precise

$$\mathbf{E}_{\mathbf{k}+\mathbf{G}}(\mathbf{r}) = \mathbf{E}_\mathbf{k}(\mathbf{r}) \tag{7}$$

if $\mathbf{G} = m_1\mathbf{b}_1 + m_2\mathbf{b}_2 + m_3\mathbf{b}_3$ where $\mathbf{b}_j$'s are the reciprocal lattice vectors defined by $\mathbf{a}_i \cdot \mathbf{b}_j = 2\pi\delta_{ij}$. This is a direct consequence of the real space periodicity, which causes the wave vector space, or equivalently, its Fourier-transformed space or the reciprocal space, also periodic. As a result, we only need to consider a unit cell in the reciprocal space, which is conventionally taken to be a Wigner–Seitz unit cell, often called the first Brillouin zone. Finding all allowed frequencies for the $\mathbf{k}$ vectors within the first Brillouin zone yields the dispersion relation $\omega(\mathbf{k})$, or the photonic band structure, from which most optical properties may be predicted. The photonic band structure also gives rise to the photonic density of states (DOS), which could be strongly modified from the case of homogeneous medium. The DOS is, by definition, the number of allowed states at a given frequency ω, and thus may be found by counting all allowed states.

$$\rho(\omega) = \int_{BZ} d^3k\delta[\omega - \omega(\mathbf{k})] \tag{8}$$

In the free space, the dispersion relation is simply given by $\omega = c\mathbf{k}$, from which one finds a DOS quadratically increasing with angular frequency [16]

$$\rho(\omega) = \frac{\omega^2}{2\pi^2 c^3} \tag{9}$$

In a PC structure, however, the DOS is strongly modified in such a way that the DOS becomes zero inside the PBG, whereas it tends to be strongly enhanced near the band edges. The modified DOS subsequently affects the light–matter interaction and plays an important role in controlling spontaneous emission rate, as discussed later.

Before we start the discussion of various numerical techniques used for computing the photonic band structure, it is profitable to investigate the symmetry properties of photonic bands, which is the most effectively done by using group theory. A PC structure is often composed of a highly symmetric arrangement of dielectric materials. For example, one of the most widely studied two-dimensional PC structures is a triangular array of air holes in a background dielectric medium. From the geometry, we find that there exist six rotations and six reflections that leave the structure invariant. They are the six rotations by angles that are multiples of 60° (E, C_2, $2C_3$, and $2C_6$) and three reflections in the planes along the

nearest-neighbor directions ($3\sigma_v$) and three more reflections in the planes along the next nearest-neighbor directions ($3\sigma_d$). These symmetry operations form the point group C_{6v}.

Following the group theoretical approach originally developed in solid-state physics [17], we can then proceed to identify the subgroup leaving the given wave vector **k** invariant, which is called the group of **k** vectors. Then, the eigenmodes having the given **k** vector are specified by the irreducible representations of this group of **k** vectors. In general, the quantum states are specified by the irreducible representations of the symmetry group containing all symmetry operations that leave the Schrödinger equation invariant, but when dealing with the Bloch states as we are now, substitution of the Bloch functions such as those given in Eq. (6) gives rise to an additional term involving the wave vector **k**. Therefore, the resulting wave equation is invariant under the operations, leaving both the **k** vector and the potential, or the dielectric function $\varepsilon(\mathbf{r})$, in our case, invariant. This process is analogous to the well-known hydrogen atom problem, in which one can first invoke the full spherical symmetry to determine the angular part of the wavefunction in terms of the spherical harmonics and then deal with the resulting radial part of the Schrödinger equation to obtain the radial part of the wavefunction. In this case, we first use Bloch's theorem to account for the periodicity from unit cell to unit cell and then deal with the resulting wave equation to determine the wavefunction within one unit cell.

To proceed further, we now need to specify the group of **k** vectors and this requires us to consider particular **k** points in the first Brillouin zone. For the triangular crystal currently being considered, the reciprocal lattice is also a triangular lattice, and therefore the first Brillouin zone is defined by a regular hexagon. The center of the Brillouin zone is called the Γ point and corresponds to $\mathbf{k} = 0$. Other high-symmetry points are K and M points, which correspond to the corners of the hexagon and the midpoints of the sides, respectively. It is immediately clear that there are six equivalent K and M points. Examining the symmetry for the Γ point shows that it remains invariant under all symmetry operations in the full point group C_{6v}, which is thus identified as the group of **k** vectors for the Γ point. This means that any eigenmodes having $\mathbf{k} = 0$ can be specified by the irreducible representations of C_{6v}, or equivalently the symmetry of these eigenmodes may be classified by the irreducible representations of C_{6v}. The point group C_{6v} has six irreducible representations; four of them are one-dimensional, and two are two-dimensional. Using the Mulliken notation, they are A_1, A_2, B_1, B_2, E_1, and E_2, where the A's and B's are one-dimensional representations and the E's are two-dimensional representations.

The transformation properties of these representations under the symmetry operations in C_{6v} are found in the character table shown in Table 1. The lowest band at the Γ point corresponds to $\omega = 0$ mode, which is invariant under all symmetry operations in C_{6v}. This mode obviously belongs to the identity representation A_1. For the second band with the E-field vector polarized within the two-dimensional plane (transverse electric [TE] mode), we show in Fig. 1a the real part of H_z field distribution calculated by the plane wave (PW) method that will be described in the following section (these figures were generated using a freely available software program [18]). The field profile clearly retains the sixfold rotational symmetry of the PC lattice. A closer examination reveals that the field changes its sign under C_6 and C_2 rotations and also under σ_d reflections, although it remains invariant under C_3 and σ_v. Inspection of the character table shows that this is exactly the transformation property of B_2 to which this mode belongs.

This process may be continued for the classification of higher modes, and the character table indicates that all of the eigenmodes at the Γ point are either nondegenerate (belonging to the A's and B's) or double degenerate (belonging to the E's). As we move away from the Γ point, the symmetry of the **k** point is lowered. For example, we find the M point remains invariant under operations E, C_2, σ_v (reflection in the vertical plane along one of the second nearest-neighbor directions; i.e., the Γ–M direction), and σ_v' (reflection in the vertical plane perpendicular to σ_v), which form the point group C_{2v}. For the intermediate **k** points lying along the Γ–M direction, the symmetry and is even lower and represented by the point group $C_{1h} = \{E, \sigma_v\}$. Similarly, the symmetry of the K point is C_{3v}, and for the **k** points lying along Γ–K directions, the symmetry is C_{1h}. The eigenmode with a particular

Table 1. Character table for the point group (a) C_{6v}, (b) C_{2v}, and (c) C_{1h}.

(a)

	E	C_2	$2C_3$	$2C_6$	$3\sigma_d$	$3\sigma_v$
A_1	1	1	1	1	1	1
A_2	1	1	1	1	−1	−1
B_1	1	−1	1	−1	−1	1
B_2	1	−1	1	−1	1	−1
E_1	2	−2	−1	1	0	0
E_2	2	2	−1	−1	0	0

(b)

	E	C_2	σ_v	σ_v
A_1	1	1	1	1
A_2	1	1	−1	−1
B_1	1	−1	1	−1
B_2	1	−1	−1	1

(c)

	E	σ_h
A	1	1
B	1	−1

k vector is then represented by one of the irreducible representations of the point group describing the symmetry of the **k** point.

One can identify the symmetry of the eigenmode by inspecting the field profiles, as we did above, but one may also predict the symmetry from the symmetry reduction scheme. Group theory provides a powerful means of predicting the symmetry reduction and consequent energy level splitting by using the character tables [17]. For example, when we move away from the Γ point toward the M point, the symmetry of the **k** vector is lowered from C_{6v} to C_{1h} and the group theoretical reduction formula predicts that the mode described by the irreducible representation B_2 of C_{6v} should be reduced to the one represented by irreducible representation A of C_{1h}. Equivalently, it is said that the symmetry B_2 of point group C_{6v} for eigenmodes at the Γ point is compatible with the symmetry A of the point group C_{1h} for eigenmodes with **k** values along the Γ–M direction. Thus, the second band for the **k** points lying along the Γ–M direction has the symmetry of irreducible representation A of C_{1h}. Collecting all possible reduction schemes of all irreducible representations of C_{6v} constitutes

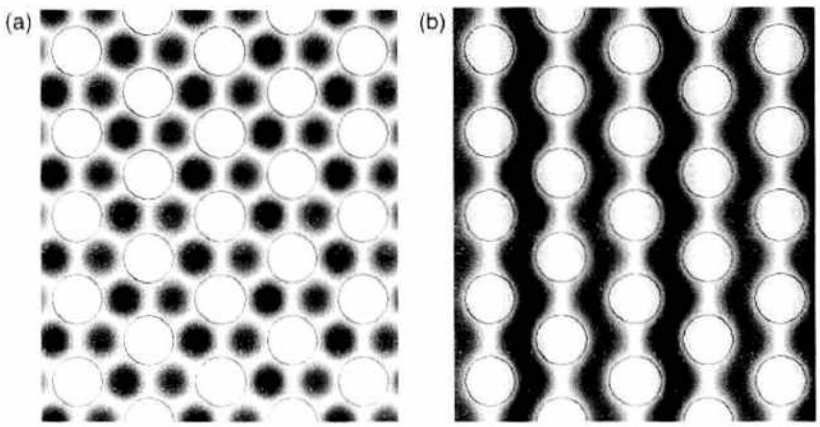


Figure 1. Color plots of (a) the real part of H_z field for the second lowest band at the Γ point and (b) the real part of H_z field for the second band at the M point. Blue, white, and red indicate negative, zero, and positive values, respectively.

the compatibility table for the symmetries of wavefunctions at the Γ point and at $\mathbf{k}$ points along Γ–M.

Similarly, one can obtain the compatibility table for the modes at the M point and for those with $\mathbf{k}$ values lying on Γ–M. As shown in Table 1, the eigenmodes at the M point may be classified by the irreducible representations A_1, A_2, B_1, and B_2 of the point group C_{2v}. Straightforward application of reduction formula using the characters shown in Table 1 reveals that among the four one-dimensional irreducible representations A_1, A_2, B_1, and B_2 of C_{2v}, A_1 and B_1 are reduced to A of C_{1h} and A_2 and B_2 are reduced to B. Thus, we expect that the second band at the M point should possess the symmetry corresponding to either A_1 or B_1. In Fig. 1b, we show the H_z field profile for the second band at the M point, which is invariant under the reflection σ_v in the plane along the Γ–M direction (the horizontal direction in Fig. 1b) and is anti-symmetric under the 180° rotation, C_2, and the reflection σ_v' in the plane perpendicular to that of σ_v (the vertical direction in Fig. 1b). In comparison with the character table of C_{2v} given in Table 1, this mode belongs to the irreducible representation B_1, which is consistent with the compatibility relationship.

As demonstrated by the example we just discussed, the eigenmodes of a PC structure possess transformation properties that may be derived from the symmetry of the PC lattice and that could be effectively described by using the group representation theory. This technique is also applicable to the localized modes associated with defects, for which the defect mode symmetries are determined by the symmetry of the defect structure. An example that dramatically shows the effectiveness of the symmetry argument is the existence of the uncoupled mode [19]. Consider a band along the Γ–M direction that has symmetry corresponding to irreducible representation B of C_{1h}. Inspection of the character table of C_{1h}, shown in Table 1, reveals that irreducible representation B is antisymmetric under the reflection σ_v, which is a plane containing the Γ–M segment and z-axis. Consider now that a plane wave traveling along the Γ–M direction is incident on this PC lattice. The incident wave is clearly symmetric under the reflection σ_v, and is thus incapable of exciting the B band because of the symmetry mismatch. If there is a range of frequency in which only the uncoupled modes are present, then the transmission coefficient will be zero even though the density of states is nonzero. This is a completely different mechanism of suppressing transmission without the presence of a PBG.

3. MODELING TECHNIQUES

3.1. Plane Wave Method

The PW method is the most widely used technique in PC modeling where Eq. (4) is solved by expanding the dielectric constant and the periodic part of the Bloch function into Fourier series over reciprocal lattice vectors.

$$\varepsilon(\mathbf{r}) = \sum_{\mathbf{G}} \varepsilon_{\mathbf{G}} e^{i\mathbf{G}\cdot\mathbf{r}} \tag{10}$$

$$\mathbf{E}(\mathbf{r}) = \sum_{\mathbf{G}} E_{\mathbf{G}} e^{i(\mathbf{k}+\mathbf{G})\cdot\mathbf{r}} \quad \text{and} \quad \mathbf{H}(\mathbf{r}) = \sum_{\mathbf{G}} \mathbf{H}_{\mathbf{G}} e^{i(\mathbf{k}+\mathbf{G})\cdot\mathbf{r}} \tag{11}$$

Substitution of Eq. (10) and (11) into Eq. (4) converts the original differential equations into matrix eigenvalue equations. We realize that only one of the two eigenvalue equations in Eq. (4) needs be solved because once the $\mathbf{E}$ or $\mathbf{H}$ field is known the other follows directly from the original Maxwell's equations. Let us first consider the wave equation for the $\mathbf{E}$ field for which the matrix eigenvalue equation is

$$(\mathbf{k}+\mathbf{G}) \times (\mathbf{k}+\mathbf{G}) \times \mathbf{E}_{\mathbf{G}} + \frac{\omega^2}{c^2} \sum_{G'} \varepsilon_{\mathbf{G}\mathbf{G}'} \mathbf{E}_{\mathbf{G}'} = 0 \tag{12}$$

where $\varepsilon_{\mathbf{G}\mathbf{G}'} \equiv \varepsilon_{\mathbf{G}-\mathbf{G}'}$. Equation (12) is an infinite dimensional generalized eigenvalue equation of the form $\mathbf{A}\mathbf{x} = \lambda\mathbf{B}\mathbf{x}$, where both $\mathbf{A}$ and $\mathbf{B}$ are Hermitian and $\mathbf{B}$ is positive definite. When the system possesses inversion symmetry so that $\varepsilon(\mathbf{r}) = \varepsilon(-\mathbf{r})$, the Fourier coefficients $\varepsilon_{\mathbf{G}}$

become real and also the matrix **B** becomes real symmetric, which then allows the use of standard numerical techniques to solve the problem. One may, of course, attempt to recast Eq. (12) into the standard form of eigenvalue equation, $\mathbf{B}^{-1}\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$. However, this complicates the problem even more because the matrix $\mathbf{B}^{-1}\mathbf{A}$ is not Hermitian.

We may also choose to solve the equation for the **H** field, for which the differential operator is Hermitian. In this case, we expand the periodic part of the **H** field into a Fourier series

$$\mathbf{H}(\mathbf{r}) = \sum_{\mathbf{G}} \sum_{\lambda=1}^{2} h_{\mathbf{G},\lambda} \hat{\mathbf{e}}_{\mathbf{G},\lambda} e^{i(\mathbf{k}+\mathbf{G})\cdot\mathbf{r}} \tag{13}$$

Here $\hat{\mathbf{e}}_{\mathbf{G},1}$ and $\hat{\mathbf{e}}_{\mathbf{G},2}$ are the two polarization unit vectors that are chosen to be perpendicular to $\mathbf{k} + \mathbf{G}$ to satisfy the transversality condition. Note that imposing the transversality condition allows us to deal with a $2N \times 2N$ matrix instead of $3N \times 3N$, where N is the number of PWs used in the expansion per polarization. Substitution of Eq. (13) and a similar Fourier expansion of $\varepsilon(\mathbf{r})$ into Eq. (4) yields the following matrix equation

$$\sum_{\mathbf{G}'} |\mathbf{k}+\mathbf{G}||\mathbf{k}+\mathbf{G}'| \eta_{\mathbf{G},\mathbf{G}'} \begin{pmatrix} \hat{\mathbf{e}}_{\mathbf{G},2}\cdot\hat{\mathbf{e}}_{\mathbf{G}',2} & -\hat{\mathbf{e}}_{\mathbf{G},2}\cdot\hat{\mathbf{e}}_{\mathbf{G}',1} \\ -\hat{\mathbf{e}}_{\mathbf{G},1}\cdot\hat{\mathbf{e}}_{\mathbf{G}',2} & \hat{\mathbf{e}}_{\mathbf{G},1}\cdot\hat{\mathbf{e}}_{\mathbf{G}',1} \end{pmatrix} \begin{pmatrix} h_{\mathbf{G}',1} \\ h_{\mathbf{G}',2} \end{pmatrix} = \frac{\omega^2(\mathbf{k})}{c^2} \begin{pmatrix} h_{\mathbf{G},1} \\ h_{\mathbf{G},2} \end{pmatrix} \tag{14}$$

where $\eta_{\mathbf{G},\mathbf{G}'} = \eta(\mathbf{G} - \mathbf{G}')$ is the Fourier transform of $\varepsilon^{-1}(\mathbf{r})$. This form of matrix eigenvalue problem, derived first by Ho et al. [20], is the most convenient form, as the matrix is real and symmetric. Equation (14) would be exact if an infinite number of PWs were used in the expansion of $\mathbf{H}(\mathbf{r})$ and $\varepsilon(\mathbf{r})$, but in practice, the series must be terminated with a finite number of PWs to make the matrix finite, so that the eigenvalue equation may be solved numerically by the standard matrix diaginalization techniques. This truncation is an intrinsic problem of the PW method and will always introduce numerical errors, whether one chooses to solve the equation for **E** or for **H** fields. However, in the case of the **H** field equation, there is an additional problem concerning the evaluation of $\eta_{\mathbf{G},\mathbf{G}'}$. If an infinite number of PWs were used, we have

$$\sum_{\mathbf{G}''} \varepsilon_{\mathbf{G}\mathbf{G}''} \eta_{\mathbf{G}''\mathbf{G}'} = \delta_{\mathbf{G}\mathbf{G}'} \tag{15}$$

When the PW expansion is truncated at a finite number, however, Eq. (15) is no longer valid and $\eta^{-1}_{\mathbf{G}\mathbf{G}'}$ and $\varepsilon^{-1}_{\mathbf{G}\mathbf{G}'}$ can become quite different. One may choose to first invert the dielectric constant in real space, take the Fourier transform of $\varepsilon^{-1}(\mathbf{r})$, and then truncate it to obtain $\eta_{\mathbf{G}\mathbf{G}'}$ (direct method). Alternatively, we may first compute the Fourier transform of $\varepsilon(\mathbf{r})$, truncate it, and then invert the matrix to find $\eta_{\mathbf{G}\mathbf{G}'}$. This method is sometimes referred to as the HCS method after those who first suggested it [20]. The question is then which technique provides the more accurate treatment of dielectric function, which in many cases contains discontinuities. The convergence properties of these two methods have been investigated for a one-dimensional lattice that has a dielectric constant with a step function profile [21]. According to this study, the direct method was found to be inaccurate in reproducing the discontinuity of $\varepsilon(\mathbf{r})$. Dealing with large dielectric constant is particularly problematic because then $\varepsilon^{-1}(\mathbf{r})$ becomes very small and the Fourier expansion may become locally negative, producing singularities in dielectric function. Figure 2a shows the dielectric function reconstructed from the truncated matrix $\eta_{\mathbf{G}\mathbf{G}'}$ by the direct method. The dielectric function exhibits singularities, and the high dielectric constant part is also significantly overestimated. The HCS method, in contrast, was found to yield more accurate reconstruction of dielectric function without any singularities. The superior reproduction of dielectric function can be seen clearly in Fig. 2b. The HCS method was also shown to exhibit superior convergence for a two-dimensional lattice [22].

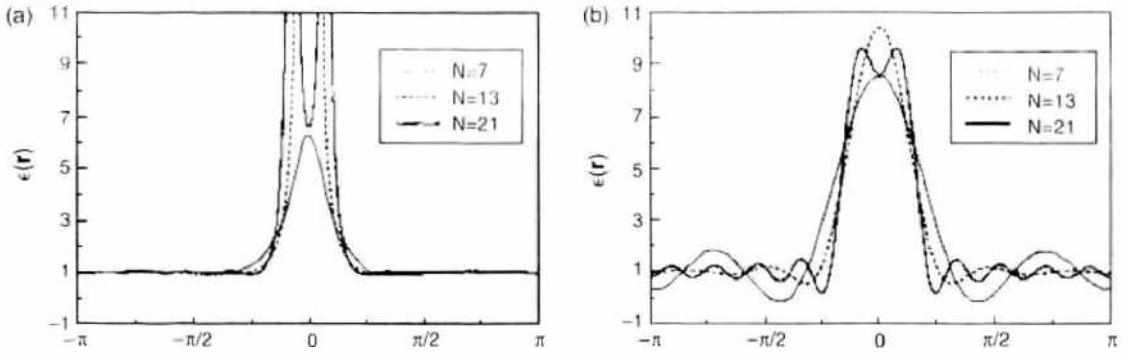


Figure 2. (a) Step profile dielectric function reproduced from the truncated η_{ka} obtained by the direct method with N plane waves. (b) Step profile dielectric function reproduced from the η_{ka} obtained by the HCS method, with N plane waves. Reprinted with permission from [21], P. R. Villeneuve and M. Piche, *Prog. Quantum Electron.* 18, 153 (1994). © 1994, Elsevier.

3.2. Finite-Difference Time-Domain Method

Another popular technique that is in many ways complementary to the PW method is the finite-difference time-domain (FDTD) method [23, 24]. Since the original proposal by Yee [25], FDTD has become one of the most widely used modeling techniques in optics and electromagnetics, and there exists a large database on numerical techniques and applications of this method (e.g., see Ref. [26]).

The FDTD method solves the discretized version of the original Maxwell's equations given in Eq. (2) rather than the wave equation. To solve the coupled equations for the **E** and **H** fields accurately, Yee proposed a unique discretization scheme now widely known as the Yee cell. As shown in Fig. 3, in the Yee cell, each **E** and **H** field component is surrounded by the loops of **H** and **E** fields, respectively, according to the Faraday's and Ampere's laws. This scheme is divergence free in nature, automatically satisfying the two Maxwell's divergence equations, and also allows us to express the spatial derivatives in the two Maxwell's curl equations by central differences that are accurate to the second order.

Once the difference equations are obtained according to the Yee cell, the time domain is discretized in such a way that the **E** field is first updated from the previous **H** field stored in

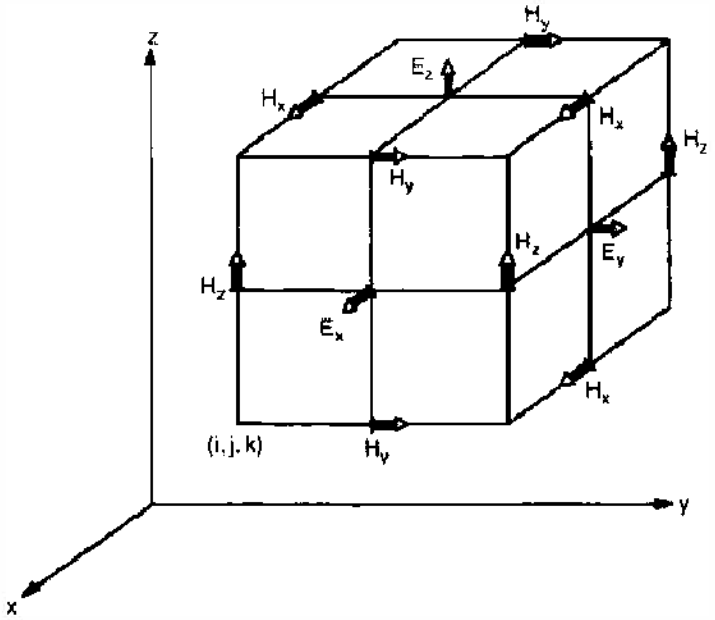


Figure 3. Positions of the electric and magnetic field vector components in the Yee cell. Reprinted with permission from [24], A. Taflove and S. C. Hagness, "Computational Electrodynamics: The Finite-Difference Time-Domain Method," 2nd ed. Artech House, Norwood, MA. Based on K. S. Yee, *IEEE Trans. Antennas Propag.* AP-14, 302 (1966). © 1966, Artech House and IEEE.

memory and then the **H** field is recalculated from the updated **E** field. The explicit forms of the discretized Maxwell's equations for B_x and D_x are

$$\frac{B_x^{n+1/2}(i, j+1/2, k+1/2) - B_x^{n-1/2}(i, j+1/2, k+1/2)}{\Delta t} = \frac{E_y^n(i, j+1/2, k+1) - E_y^n(i, j+1/2, k)}{\Delta z} - \frac{E_z^n(i, j+1, k+1/2) - E_z^n(i, j, k+1/2)}{\Delta y}$$

$$\frac{D_x^n(i+1/2, j, k) - D_x^{n-1}(i+1/2, j, k)}{\Delta t} = \frac{H_z^{n-1/2}(i+1/2, j+1/2, k) - H_z^{n-1/2}(i+1/2, j-1/2, k)}{\Delta y} - \frac{H_y^{n-1/2}(i+1/2, j, k+1/2) - H_y^{n-1/2}(i+1/2, j, k-1/2)}{\Delta z} + J_x^{n-1/2}(i+1/2, j, k) \quad (16)$$

where the superscripts represent the time step and the indices inside the parentheses are spatial grid points at which the field values are defined. The remaining discretized Maxwell's equations can easily be found by properly permuting the x-, y-, and z-components of the fields in Eq. (16), and they may also be found in Yee's original paper [25]. Each time **B** and **D** fields are updated according to Eq. (16), and they are converted to **H** and **E** fields by the constitutive relations given in Eq. (3).

The repeating cycle of updating fields, sometimes called the leapfrog time-stepping procedure, is carried out for a time period long enough to provide accurate information on the quantities one originally intended to obtain. Of fundamental importance in FDTD procedure is that the time-stepping procedure must remain stable. In other words, the field values must not diverge because of the accumulation of numerical errors. This requirement puts a restriction on the values for the space grid size and time increment, which is known as the Courant stability condition. The condition is derived from the numerical dispersion relation, where one must avoid the emergence of imaginary parts in angular frequency that can lead to exponential growth of fields [24]. For n-dimensional space uniformly discretized into a Cartesian grid, the stability condition is

$$\Delta t \leq \frac{\Delta x}{c\sqrt{n}} \quad (17)$$

where Δt and Δx are time and space grid resolutions, respectively, and c is the speed of light.

For photonic band structure calculation, the computational cell is usually the unit cell terminated by the periodic boundary conditions according to the Bloch theorem. In some cases such as two-dimensional slab PCs, one needs to simulate a uniform medium extending to infinity, which can be done by implementing an absorbing boundary condition. A simple and effective scheme to implement an absorbing boundary condition was proposed by Mur [27]. Mur's absorbing boundary condition uses the one-way wave equation to determine the field values at the boundaries that yield zero reflection at all incident angles. The one-way wave equation is the partial differential equation that allows propagation in only one direction and that can thus be used at FDTD grid boundaries to create conditions that absorb all incident waves, as was first proved by Engquist and Majda [28].

The analytical form of the one-way wave equation, however, cannot be directly implemented in the FDTD scheme but has to be approximated with a Taylor series. After retaining only one or two terms in the Taylor series, Mur transformed the resulting differential equation into the finite difference equations by taking central differences in both space and time. A more powerful scheme that gives much less reflection at all incident angles is the perfectly matched layer (PML) boundary condition, which was originally proposed by Berenger [29]. PML is an artificial medium perfectly impedance matched to the host medium so that PWs with any incident angle, polarization, and frequency would not experience any reflection at

the boundary. At the same time, the region is defined to be lossy so that the wave propagating inside the PML region dies out before it encounters the edge of the PML region. This is accomplished by making both ε and μ complex quantities.

To implement this concept for the FDTD algorithm, Berenger modified the Maxwell's equation by splitting up the field components so that he can separately define x-, y-, and z-components of the imaginary parts of ε and μ. This way, he could define PML layers that preferentially absorb waves propagating in certain directions. This is called the Berenger's split-field formalism. Since the original proposal by Berenger, numerous variations and implementation methods for the PML boundary condition have been proposed, and interested readers are referred to the literature available on this subject [26]. The PML boundary condition has been shown to exhibit superior performance to any other analytical absorbing boundary conditions. For example, Berenger's original paper reported a local reflection coefficient 3000 times smaller than that obtained by Mur's absorbing boundary condition.

Once the computational cell is defined with proper boundary condition, a certain initial field profile is assigned and then evolved in time by the leapfrog time-stepping procedure. At each time step, field values at certain low symmetry points are stored and the time series is later Fourier transformed to obtain allowed mode frequencies. This procedure is repeated for various $\mathbf{k}$ vectors to yield the complete band structure. The total time during which the fields are evolved must be kept long because it determines the frequency resolution in the subsequent Fourier transform. In addition, the initial field profile must be carefully chosen so that it contains nonzero projections to all allowed modes we wish to find while avoiding the unwanted dispersionless longitudinal modes. A convenient way to impose the transversality condition was proposed by Chan et al., who set up the initial $\mathbf{H}$ field as [30]

$$\mathbf{H} = (\mathbf{r}) = \sum_{\mathbf{G}} [\mathbf{v} \times (\mathbf{k} + \mathbf{G})] e^{i(\mathbf{k}+\mathbf{G})\cdot\mathbf{r}} \tag{18}$$

where $\mathbf{G}$'s are reciprocal lattice vectors and $\mathbf{v}$ is a vector that has roughly equal magnitudes for x, y, and z components. The $\mathbf{H}$ field defined in Eq. (18) obviously satisfies $\nabla \cdot \mathbf{H} = 0$ and will remain transverse throughout the time-stepping procedure. It should also be mentioned that the summation does not have to cover many reciprocal lattice vectors, and the accuracy of the simulation does not depend on the number of reciprocal lattice vectors in the sum.

Because the FDTD method works in the real space and time domain, it is an excellent technique for visualization of field profile and time evolution. It can handle impulsive or transient behaviors naturally. Also, like other real space techniques, choosing and imposing various different kinds of boundary conditions, as needed by the system of interest, is straightforward. This is in contrast to the PW method, in which the use of the periodic boundary condition is mandatory. The freedom in choosing boundary conditions is particularly advantageous for the modeling of nonperiodic structures such as photonic defects because in the FDTD method, one does not need to construct artificial supercells as in the PW method. Furthermore, FDTD is considered well suited for large and complex systems because the simulation load scales with N, the size of the system. This is again in contrast to the PW method, in which the computational load scales with N^3 because of the N^3-dependence of the matrix diagonalization procedure. Sakoda et al. compared the CPU time needed for simulations by PW and FDTD methods and, as shown in Fig. 4, the FDTD method was found more efficient for larger systems [31].

Calculating transmission and reflection spectra using FDTD is also straightforward. For this type of simulation, a finite size PC must be encoded in the computational cell, which is surrounded by absorbing boundary layers. Then an incident PW is specified as the initial field that is evolved in time by the FDTD algorithm. The transmission and reflection coefficients are calculated simply by recording field values at appropriate points and then comparing them with the incident field amplitude. This procedure can be repeated for various frequencies to obtain the full spectrum, or more efficiently, the initial PW can be made to be a Gaussian pulse in time, which yields a Gaussian pulse in frequency domain too. One then proceeds to record the field values at the appropriate detection points and to convert the times series into the frequency spectrum by Fourier transform. This way, transmission and reflection spectra for a range of frequency may be obtained with a single simulation run.

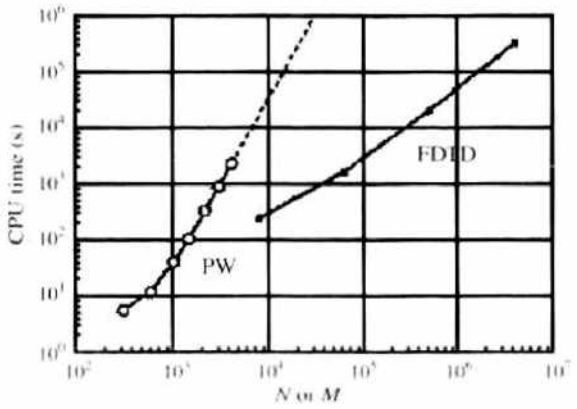


Figure 4. CPU time necessary for numerical calculation for one wave vector as a function of the number of mesh points N or the number of plane waves M. PW, plane waves; FDTD, finite-difference time-domain. Reprinted with permission from [31], K. Sakoda et al., *Phys. Rev. B* 64, 045116 (2001). © 2001, American Physical Society.

3.3. Transfer Matrix Method

Another useful real space method derived from low-energy electron diffraction theory has been proposed by Pendry and MacKinnon [32, 33]. Similar to the FDTD method, the transfer matrix method begins by discretizing the two Maxwell's curl equations and constructing the transfer matrix that relates **E** and **H** fields at grid points in two parallel planes. The discretization scheme is very similar to the Yee cell in the FDTD method, and each component of the **E** and **H** fields is surrounded by other components in a manner consistent with the Ampere's and Faraday's laws. For the time derivatives, however, explicit harmonic dependence is introduced instead of the direct discretization used in FDTD. This process yields six difference equations relating the **E** and **H** field components at one grid point with those at the neighboring grid points. One then proceeds to eliminate two components, usually chosen to be the z-components of the **E** and **H** fields to obtain a set of four equations relating the x- and y-components of **E** and **H** fields at one plane to those at an adjacent plane perpendicular to the z-axis. If a cubic mesh defined by vectors **a**, **b**, and c pointing x-, y-, and z-directions, respectively, and all with a length a, is used to discretize the Maxwell's equations, one finds through an algebraic procedure that is rather tedious but straightforward [32]

$$
\begin{aligned}
E_x(\mathbf{r}+\mathbf{c}) &= E_x(\mathbf{r}) + \frac{a^2\omega^2}{c^2}\mu(\mathbf{r})H'_y(\mathbf{r}) + \varepsilon^{-1}(\mathbf{r})[H'_y(\mathbf{r}-\mathbf{a}) - H'_y(\mathbf{r}) - H'_x(\mathbf{r}-\mathbf{b}) + H'_x(\mathbf{r})] \\
&\quad - \varepsilon^{-1}(\mathbf{r}+\mathbf{a})[H'_y(\mathbf{r}) - H'_y(\mathbf{r}+\mathbf{a}) - H'_x(\mathbf{r}+\mathbf{a}-\mathbf{b}) + H'_x(\mathbf{r}+\mathbf{a})] \\
E_y(\mathbf{r}+\mathbf{c}) &= E_y(\mathbf{r}) - \frac{a^2\omega^2}{c^2}\mu(\mathbf{r})H'_x(\mathbf{r}) + \varepsilon^{-1}(\mathbf{r})[H'_y(\mathbf{r}-\mathbf{a}) - H'_y(\mathbf{r}) - H'_x(\mathbf{r}-\mathbf{b}) + H'_x(\mathbf{r})] \\
&\quad - \varepsilon^{-1}(\mathbf{r}+\mathbf{b})[H'_y(\mathbf{r}-\mathbf{a}+\mathbf{b}) - H'_y(\mathbf{r}+\mathbf{b}) - H'_x(\mathbf{r}) + H'_x(\mathbf{r}+\mathbf{b})] \\
H'_x(\mathbf{r}+\mathbf{c}) &= H'_x(\mathbf{r}) + \varepsilon(\mathbf{r}+\mathbf{c})E_y(\mathbf{r}+\mathbf{c}) \\
&\quad - \frac{c^2}{a^2\omega^2}\mu^{-1}(\mathbf{r}-\mathbf{a}+\mathbf{c})[E_y(\mathbf{r}+\mathbf{c}) - E_y(\mathbf{r}-\mathbf{a}+\mathbf{c}) - E_x(\mathbf{r}-\mathbf{a}+\mathbf{b}+\mathbf{c}) + E_x(\mathbf{r}-\mathbf{a}+\mathbf{c})] \\
&\quad + \frac{c^2}{a^2\omega^2}\mu^{-1}(\mathbf{r}+\mathbf{c})[E_y(\mathbf{r}+\mathbf{a}+\mathbf{c}) - E_y(\mathbf{r}+\mathbf{c}) - E_x(\mathbf{r}+\mathbf{b}+\mathbf{c}) + E_x(\mathbf{r}+\mathbf{c})] \\
H'_y(\mathbf{r}+\mathbf{c}) &= H'_y(\mathbf{r}) - \varepsilon(\mathbf{r}+\mathbf{c})E_x(\mathbf{r}+\mathbf{c}) \\
&\quad - \frac{c^2}{a^2\omega^2}\mu^{-1}(\mathbf{r}-\mathbf{b}+\mathbf{c})[E_y(\mathbf{r}+\mathbf{a}-\mathbf{b}+\mathbf{c}) - E_y(\mathbf{r}-\mathbf{b}+\mathbf{c}) - E_x(\mathbf{r}+\mathbf{c}) + E_x(\mathbf{r}-\mathbf{b}+\mathbf{c})] \\
&\quad + \frac{c^2}{a^2\omega^2}\mu^{-1}(\mathbf{r}+\mathbf{c})[E_y(\mathbf{r}+\mathbf{a}+\mathbf{c}) - E_y(\mathbf{r}+\mathbf{c}) - E_x(\mathbf{r}+\mathbf{b}+\mathbf{c}) + E_x(\mathbf{r}+\mathbf{c})] \qquad (19)
\end{aligned}
$$

where $\mathbf{H}'$ is the reduced **H** defined as $\mathbf{H}' = i\mathbf{H}a\omega\varepsilon_0$. The first two transfer equations given in Eq. (19) express **E** in the next plane by **E** and **H** in the previous plane, whereas the last

two equations describe **H** in the next plane in terms of **H** in the previous plane and **E** in the same plane. It is therefore possible to find all field components in the next plane ($z = z' + a$) if the fields at the previous plane ($z = z'$) are known. Equation (19) may be recast into a matrix form to define the transfer matrix as follows

$$\mathbf{F}(z + a) = \mathbf{T}(z)\mathbf{F}(r) \tag{20}$$

where

$$\mathbf{F}(z) = \begin{pmatrix} E_x(\mathbf{r}) \\ E_y(\mathbf{r}) \\ H_y'(\mathbf{r}) \\ H_x'(\mathbf{r}) \end{pmatrix}$$

with the position vector **r** lying in the plane defined by z and where $\mathbf{T}(z)$ is the transfer matrix whose elements are specified by ω, $\varepsilon(\mathbf{r})$, and $\mu(\mathbf{r})$ according to Eq. (19).

The transfer matrix method provides a very convenient way to calculate transmission and reflection coefficients. For this purpose, one first needs to choose a basis set for which a set of PWs are frequently used. In the transfer matrix formalism, PWs are expressed as the eigenvectors of transfer matrix $\mathbf{T}^{(0)}$ for free space,

$$\mathbf{T}^{(0)}\mathbf{F}_{ij}^{(0)}(z) = \mathbf{F}_{ij}^{(0)}(z + a) = \exp(ik_i^{(0)}a)\mathbf{F}_{ij}^{(0)}(z) \tag{21}$$

Because the transfer matrix is not Hermitian, there exist distinct left eigenvectors

$$\mathbf{F}_{ij}^{(0)}(z)\mathbf{T}^{(0)} = F_{ij}^{(0)}(z + a) = \exp(ik_i^{(0)}a)\mathbf{F}_{ij}^{(0)}(z) \tag{22}$$

They are orthonormal to the right eigenvectors. These left and right eigenvectors then define a unitary transformation that may be used to transform a given transfer matrix to the PW basis set. Consider a layer of PC's whose transfer matrix is $\mathbf{T}(z)$. Using the PW basis set, the transfer matrix is expressed as

$$\langle \widetilde{\mathbf{T}}(z) \rangle_{ij} = \mathbf{F}_{ij}^{(0)}(z)\mathbf{T}(z)\mathbf{F}_{ij}^{(0)}(z) \tag{23}$$

If the matrix elements are arranged in such a way that the waves traveling to the right come first and are followed by the waves traveling to the left, the transfer matrix may be divided into four submatrices.

$$\widetilde{\mathbf{T}} = \begin{pmatrix} \mathbf{T}^{++} & \mathbf{T}^{+-} \\ \mathbf{T}^{-+} & \mathbf{T}^{--} \end{pmatrix} \tag{24}$$

where the superscripts + and − indicate right- and left-propagating waves, respectively. Consider now PWs, $\mathbf{F}_j^{(0)}$, incident on a PC slab from the left. The reflected and transmitted waves are related through the reflection and transmission matrices $\mathbf{t}^{-+}$ and $\mathbf{t}^{--}$. The total wave on the left is given by $\sum_j \mathbf{t}_{ij}^{-+}\mathbf{F}_j^{-(0)} + \mathbf{0} \cdot \mathbf{F}_j^{+(0)}$, whereas the wave on the right is expressed as $\sum_j \mathbf{t}_{ij}^{++}\mathbf{F}_j^{+(0)} + \mathbf{F}_i^{-(0)}$. These two waves are also related through the transfer matrix as

$$\begin{pmatrix} \mathbf{t}^{--} \\ \mathbf{I} \end{pmatrix} = \begin{pmatrix} \mathbf{T}^{++} & \mathbf{T}^{+-} \\ \mathbf{T}^{-+} & \mathbf{T}^{--} \end{pmatrix} \begin{pmatrix} \mathbf{0} \\ \mathbf{t}^{--} \end{pmatrix} \tag{25}$$

from which we find

$$\begin{aligned} \mathbf{T}^{++} &= \mathbf{t}^{+-}(\mathbf{t}^{--})^{-1} \\ \mathbf{T}^{-+} &= (\mathbf{t}^{--})^{-1} \end{aligned} \tag{26}$$

Similarly, the two other submatrices of the transfer matrix can be found by considering PWs incident on the PC slab from the right. One then arrives at the final expression of transfer matrix in terms of reflection and transmission matrices as follows

$$\begin{pmatrix} \mathbf{T}^{++} & \mathbf{T}^{+-} \\ \mathbf{T}^{-+} & \mathbf{T}^{--} \end{pmatrix} = \begin{bmatrix} \mathbf{t}^{++} - \mathbf{t}^{+-}(\mathbf{t}^{--})^{-1}\mathbf{t}^{-+} & \mathbf{t}^{+-}(\mathbf{t}^{--})^{-1} \\ -(\mathbf{t}^{--})^{-1}\mathbf{t}^{-+} & (\mathbf{t}^{--})^{-1} \end{bmatrix} \tag{27}$$

Thus, once the transfer matrix is determined in the PW basis set, transmission and reflection coefficients may be obtained using the above equation. Transmission and reflection for a thick slab can also be found by using the same equation if one uses the proper transfer matrix for the thick slab.

Photonic band structure may also be calculated using the transfer matrix. For a periodic system, translation by a primitive unit vector obeys the Bloch theorem.

$$\mathbf{F}(\mathbf{r} + \mathbf{u}_i) = \exp(ik_i a_i)\mathbf{F}(\mathbf{r}), \quad i = x, y, z \tag{28}$$

where $\mathbf{u}_i$ are the unit vectors for the crystal with lengths a_i. Translation along the z-direction by a unit vector is also related to the transfer matrix

$$\mathbf{F}(\mathbf{r} + \mathbf{u}_z) = \mathbf{T}\mathbf{F}(\mathbf{r}) = \exp(ik_z a_z)\mathbf{F}(\mathbf{r}) \tag{29}$$

When the frequency ω is given and k_x and k_y specified according to the periodic boundary condition, the transfer matrix $\mathbf{T}$ for translation by a unit cell is obtained, and then the eigenvalue equation in Eq. (29) is solved to find all allowed values of k_z. This procedure may be repeated for various frequencies to yield the complete band structure $k_z(\omega)$.

3.4. Kohn–Korringa–Rostocker Method

Adapted from the Kohn–Korringa–Rostocker (KKR) method widely used for electronic energy band structure studies, this technique uses expansion over vector spherical harmonics to solve the eigenvalue equation in Eq. (4) [34–37]. It turns out that for spherical scatterers, each vector spherical wave is scattered independent of others, and as a consequence this technique can handle discontinuities in dielectric constant across spherical boundaries with arbitrarily high accuracy. It is therefore extremely efficient in dealing with the PCs made of spherical building blocks, such as self-assembled opals. As pointed out by Busch [38], KKR and PW methods may be regarded as complementary to each other in dealing with systems with spherical symmetry. The KKR method allows an extremely accurate representation of the discontinuities in the dielectric function, but the lattice structure or the periodicity is incorporated approximately by the multicenter expansion technique. In contrast, the PW method provides an accurate representation of periodicity through the expansion over reciprocal lattice vectors but incompletely incorporates the discontinuities in the dielectric constant through the Fourier coefficients.

Following the original development by Stefanou et al. [35, 39], we begin with a general expression for spherical wave expansion.

$$\mathbf{E}(\mathbf{r}) = \sum_{l=1}^{\infty} \sum_{m=-l}^{l} \left[\frac{i}{k} a_{lm}^{E} \nabla \times z_l(kr)\mathbf{X}_{lm}(\hat{\mathbf{r}}) + a_{lm}^{H} \nabla \times z_l(kr)\mathbf{X}_{lm}(\hat{\mathbf{r}}) \right] \tag{30}$$

where z_l is a combination of spherical Bessel and Hankel functions and $\mathbf{X}_{lm}$ are the vector spherical harmonics. When a PW is incident on a single spherical scatterer, the coefficients in Eq. (30) for the scattered wave are given by the well-known Mie scattering solution (e.g., Ref. [40]). When the scattering is caused by a periodic array of spherical scatterers in a two-dimensional plane, the scattered wave can be expressed as a sum of the waves scattered by individual scatterers. Because of the periodicity, the wave scattered by a sphere at $\mathbf{r} = \mathbf{R}_n$ differs only by a simple phase factor $\exp(i\mathbf{k}_{\parallel} \cdot \mathbf{R}_n)$ from the wave scattered by a sphere at the origin. Here, $\mathbf{k}_{\parallel}$ represents the tangential component of $\mathbf{k}$ vector parallel to the plane of scatterers. The tangential $\mathbf{k}$ vector $\mathbf{k}$ can be redefined by using the reciprocal lattice

vectors of the two-dimensional lattice, $\mathbf{k}_{\text{tangential}} = \mathbf{k}_t + \mathbf{g}$, so that $\mathbf{k}_t$ now represents a vector within the unit cell of the two-dimensional reciprocal lattice, which is often called the surface Brillouin zone. Once the expression for the scattered wave in the form of Eq. (30) is found, one can construct a matrix equation relating the coefficients of the incident wave and the total scattered wave.

$$[E_{sa}]^{\nu}_{gi} = \sum_{g'i'} M^{\nu\lambda}_{gi;g'i'} [E_{in}]^{\lambda}_{g'i'} \tag{31}$$

The explicit expression for the matrix $\mathbf{M}$ may be found in Ref. [41]. To obtain the photonic band structure of a three-dimensional PC, we proceed to construct a three-dimensional system consisting of many layers of the two-dimensional structure and consider the multiple scatterings between layers in the same way as in the transfer matrix method. The field between the nth and $(n+1)$th layers is determined by the scattering of the two neighboring planes of scatterers. That is, the backward-propagating wave in the region between the nth and $(n+1)$th layer is the sum of the forward-propagating wave in the same region back scattered by $(n+1)$th layer and of the backward-propagating wave between the $(n+1)$th and $(n+2)$th layers transmitted by the $(n+1)$th layer. The forward-propagating wave is similarly defined. The transmission and reflection (or back scattering) matrices are obtained from the M matrix in Eq. (31) modified to account for the shift of origin. This way, one obtains equations relating fields at adjacent layers that must also satisfy the Bloch theorem because the system is periodic in the third dimension as well. Combining the multiple scattering equations and the Bloch theorem, one finds

$$\begin{pmatrix} \mathbf{Q}^{+-} & \mathbf{Q}^{++} \\ -[\mathbf{Q}^{--}]^{-1}\mathbf{Q}^{-+}\mathbf{Q}^{+-} & [\mathbf{Q}^{--}]^{-1}[\mathbf{1} - \mathbf{Q}^{-+}\mathbf{Q}^{+-}] \end{pmatrix} \begin{bmatrix} E^{-}(n) \\ E^{-}(n+1) \end{bmatrix} = \exp(i\mathbf{k}\cdot\mathbf{a}_3) \begin{bmatrix} E^{+}(n) \\ E^{-}(n+1) \end{bmatrix} \tag{32}$$

where $\mathbf{Q}$ are the scattering matrices defined similarly to the $\mathbf{t}$ matrices in Section 3.3, and their elements are determined from the $\mathbf{M}$ matrix in Eq. (31). Also, $\mathbf{a}_3$ is the primitive unit vector representing the periodicity among the two-dimensional layers. For given ω and $\mathbf{k}_{||}$ in the surface Brillouin zone, Eq. (32) can be solved to find k_z, the wave vector component in the direction perpendicular to the two-dimensional layers. In practice, only a finite number of $\mathbf{g}$ vectors are included in evaluating Eq. (32), which means the matrix equation would be $6N$ dimensional, where N is the number of $\mathbf{g}$ vectors. The eigenvalue equation would then result in $6N$ eigenvalues for $k_z(\omega, \mathbf{k}_{||})$. Most of the solutions are complex and thus represent evanescent waves. Only a few of the solutions are real, corresponding to propagating waves. By scanning the frequency region of interest and the various $\mathbf{k}_{||}$'s in the surface Brillouin zone, one can obtain complete photonic band structure. Note that the multiple scattering formalism is very similar to the transfer matrix and thus naturally provides transmission and reflection coefficients. In addition to having finite $\mathbf{g}$ vectors, the spherical wave expansion in Eq. (30) must also be truncated at a certain value of $l_{\max}$ in practice, similar to the PW method where the PW expansion has to be truncated. This truncation consequently introduces numerical errors. However, it was found that the KKR method generally requires far fewer basis functions than the PW method. For example, Wang et al. reported that using $l_{\max} = 3$ resulted in the same level of accuracy as the PW method [34]. Typically, $l_{\max} = 6$ is enough to yield good convergence, in which case the dimensions of the secular determinant are less than 100. This is in contrast to the PW method, which typically handles secular determinants with dimensions of ~500.

4. PHOTONIC BAND STRUCTURE

4.1. Photonic Band Gap

Since Yablonovitch first suggested the prospect of completely inhibiting spontaneous emission by PBG [1], the search for structures exhibiting complete PBG has been one of the most strong driving forces in PC research. PBG refers to a range of frequency in which there

are no allowed photonic states, and thus propagation and generation of light is completely prohibited. Realization of complete PBG then provides opportunities to suppress unwanted radiative transitions or light transmission and also to create strongly localized states within the PBG by deliberately introducing defects. Because of the strong confinement mechanism provided by PBG, the photonic defect structures can be made to exhibit extraordinary properties, such as near-perfect transmission around a sharp bend and an extremely high confinement factor, as will be discussed in Section 5.

4.1.1. Bragg Reflector

The formation of PBG is best illustrated in the simplest example of one-dimensionally periodic multilayer film, which is usually referred to as a Bragg reflector and is widely used in various photonic devices. A Bragg reflector consists of repeating pairs of two dielectric layers with distinct refractive indices, n_1 and n_2. The interfaces between materials n_1 and n_2 generate a series of reflected and transmitted waves, which interfere with one another to produce the system response. If we denote the individual layer thicknesses as a and b, a pair of two dielectric layers may be considered a unit cell, with a lattice constant $\Lambda = a + b$. Then, the condition for constructive interference between all reflected waves is given by

$$K\Lambda = m\pi; \quad m = \text{integer} \tag{33}$$

where K is the wave vector component along the direction perpendicular to the film (x-direction in the following discussion). This is the optical analog of the well-known Bragg condition for x-ray diffraction in crystals, from which the name Bragg reflector is derived. To proceed farther, let us construct a transfer matrix relating the amplitudes of forward- and backward-propagating waves in two adjacent unit cells. Noting that the periodicity affects only the propagation along the x-direction, we will omit, for simplicity, the in-plane propagation in the following discussion, but the extension for arbitrary propagation direction is straightforward. Following the development by Yeh and Yariv [42], we write

$$\begin{pmatrix} a_{n-1} \\ b_{n-1} \end{pmatrix} = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} a_n \\ b_n \end{pmatrix} \tag{34}$$

where a and b are the amplitudes of electric field for forward- and backward-propagating waves, respectively, and the subscripts indicate the unit cell in which the fields are defined. The matrix elements A, B, C, and D are obtained by using the standard technique, where we account for the propagation through the individual layers with the wave vector and layer thickness and for the reflections and transmissions at the interfaces with the Fresnel coefficients. For normal incidence, one finds,

$$\begin{aligned} A &= e^{ik_1 a}\left[\cos k_2 b + \frac{1}{2}i\left(\frac{n_2}{n_1} + \frac{n_1}{n_2}\right)\sin k_2 b\right] \\ B &= e^{-ik_1 a}\left[\frac{1}{2}i\left(\frac{n_2}{n_1} - \frac{n_1}{n_2}\right)\sin k_2 b\right] \\ C &= e^{ik_1 a}\left[-\frac{1}{2}i\left(\frac{n_2}{n_1} - \frac{n_1}{n_2}\right)\sin k_2 b\right] \\ D &= e^{-ik_1 a}\left[\cos k_2 b - \frac{1}{2}i\left(\frac{n_2}{n_1} + \frac{n_1}{n_2}\right)\sin k_2 b\right] \end{aligned} \tag{35}$$

where $k_1 = n_1\omega/c$ and $k_2 = n_2\omega/c$. Since the system is periodic, we can also invoke Bloch's theorem, which asserts that the fields must be described by a PW modulated by a periodic function with the same periodicity as the multilayer film

$$E(x) = E_K(x)e^{-iKx} \text{ where } E_K(x + \Lambda) = E_K(x) \tag{36}$$

or

$$E(x + \Lambda) = E(x)e^{iK\Lambda}$$

Combining Eqs. (34) and (36) yields

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} a_n \\ b_n \end{pmatrix} = e^{jK\Lambda} \begin{pmatrix} a_n \\ b_n \end{pmatrix} \tag{37}$$

The Block wave vector, **K**, is therefore determined from the eigenvalue of the transfer matrix describing the light propagation over a unit cell. Solving the secular equation, we find

$$e^{jK\Lambda} = \frac{A+D}{2} \pm \left[\left(\frac{A+D}{2} \right)^2 - 1 \right]^{1/2} \tag{38}$$

It should be noted that this formalism is the one-dimensional version of the transfer matrix method described in Section 3.3. Because the transfer matrix is unimodular, the two eigenvalues are reciprocal to each other, and thus one finds two eigenvalues $e^{\pm jK\Lambda}$ for a given frequency. They represent forward- and backward-propagating Block waves, which are expected to appear together for an infinitely periodic system. Eq. (38) provides the dispersion relation between K and ω,

$$K = \frac{1}{\Lambda} \cos^{-1} \left[\frac{A+D}{2} \right] \tag{39}$$

If $(A+D)/2 < 1$, then K is real and we have propagating Bloch waves. If, however, $(A+D)/2 > 1$, then K becomes complex, leading to exponentially decaying evanescent waves. This region is called the forbidden band in which no propagating waves are allowed. Notice that $(A+D)/2$ has a maximum at which the Bragg condition is met. Thus, the Bragg condition corresponds to the middle of a forbidden band. It is noted that the evanescent waves are indeed solutions to the eigenvalue Eq. (37), but they do not satisfy the periodic boundary conditions that are routinely imposed in photonic band calculations. Therefore, they do not constitute proper solutions for a perfect, infinite PC structure and are relevant only for a finite PC or a PC containing defects.

Rewriting Eq. (37) explicitly using Eq. (35) gives,

$$\cos K\Lambda = \cos k_1 a \cos k_2 b - \frac{1}{2} \left(\frac{n_2}{n_1} + \frac{n_1}{n_2} \right) \sin k_1 a \sin k_2 b \tag{40}$$

Figure 5 shows the dispersion curve obtained from Eq. (40). Forbidden gaps open when the Bloch wave vector **K** satisfies the Bragg condition. For an infinitely periodic system, all K's related by integer multiples of $2\pi/\Lambda$ are equivalent because of Bloch's theorem. We may therefore represent the whole dispersion curve within a finite region of K,

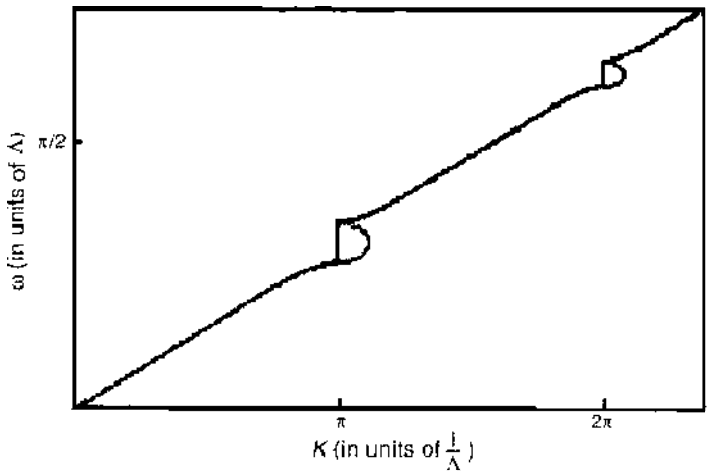


Figure 5. Dispersion relation for a Bragg reflector for normal incidence. From A. Yariv and P. Yeh, "Optical Waves in Crystals," p. 173. Wiley, New York, 1984. © 1984, Wiley.

say $-\pi/\Lambda < K < \pi/\Lambda$, which is equivalent to the well-known reduced zone scheme in condensed matter physics.

We will conclude this section by estimating the width of the forbidden band. For this analysis, let us consider a quarter-wave stack for which $k_1 a = k_2 b = \pi/2$ when $\omega = \omega_0$, the central frequency of a forbidden gap. The band edges are defined by setting the left-hand side equal to 1 in Eq. (40), which leads to

$$1 = \cos\left(\frac{\omega_0 - \omega'}{c} n_1 a\right) \cos\left(\frac{\omega_0 - \omega'}{c} n_2 b\right) - \frac{1}{2}\left(\frac{n_2}{n_1} + \frac{n_1}{n_2}\right) \sin\left(\frac{\omega_0 - \omega'}{c} n_1 a\right) \sin\left(\frac{\omega_0 - \omega'}{c} n_2 b\right) \tag{41}$$

where we introduced a new variable $\omega' = \omega_0 - \omega$ that represents the deviation from the middle of the forbidden gap. In a quarter-wave stack, $n_1 a = n_2 b = \pi c/2\omega_0$, so Eq. (41) gets reduced to

$$1 = \sin^2\left(\frac{\pi}{2}\frac{\omega'}{\omega_0}\right) - \frac{1}{2}\left(\frac{n_2}{n_1} + \frac{n_1}{n_2}\right) \cos^2\left(\frac{\pi}{2}\frac{\omega'}{\omega_0}\right) \tag{42}$$

Solving for ω' yields,

$$\omega' = \pm\frac{2\omega_0}{\pi} \sin^{-1}\left|\frac{n_2 - n_1}{n_2 + n_1}\right| \tag{43}$$

from which we find the forbidden gap width $\Delta\omega$ as

$$\Delta\omega = \frac{4\omega_0}{\pi} \sin^{-1}\left|\frac{n_2 - n_1}{n_2 + n_1}\right| \tag{44}$$

Equation (44) shows that the forbidden gap width is generally bigger for a larger $\Delta n = |n_1 - n_2|$ and is linearly proportional to Δn when Δn is small. This principle holds for two-dimensional and three-dimensional systems, as we will see later, and thus one of the fundamental requirements for the realization of PBG is to have a high refractive index contrast.

4.1.2. Two-Dimensional Photonic Crystals

The concept discussed in the previous section can be extended to a system that possesses two-dimensional periodicity. In this case, the fields in the plane of two-dimensional periodicity is described by a Bloch wave

$$\mathbf{E}(\mathbf{r}) = e^{i\mathbf{k}_z} e^{i\mathbf{k}_{||}\cdot\boldsymbol{\rho}} u(\boldsymbol{\rho}) \tag{45}$$

where $\mathbf{k}_z$ and $\mathbf{k}_{||}$ are the wave vectors for propagation in the out-of-plane and in-plane directions, respectively, and $u(\boldsymbol{\rho})$ is a function with the same periodicity as the two-dimensional PC lattice. The two-dimensional periodicity restricts the in-plane wave vector $\mathbf{k}_{||}$ within the first Brillouin zone, whereas $\mathbf{k}_z$ is unrestricted. For in-plane propagation, that is, $k_z = 0$, the two-dimensional periodicity may give rise to Bragg reflections for certain frequencies, producing a two-dimensional PBG in which the light propagation in all in-plane directions is forbidden. For the ideal two-dimensional systems that are invariant in the out-of-plane direction, there exist two independent polarizations, TE (**E**-field in the two-dimensional plane) and TM (**H**-field in the two-dimensional plane), and therefore a complete PBG requires overlapping band gaps for both polarizations. This was found to be realized in a triangular array of air holes in a dielectric material, which has naturally been the most widely investigated system. Figure 6 shows the photonic band diagram for a triangular lattice of air holes in Si, which has a dielectric constant $\varepsilon = 12$. The air hole radius was set to be $0.45a$, where a is the lattice constant (i.e., the center-to-center distance between two adjacent holes). The reciprocal lattice of a triangular lattice is also a triangular lattice, and thus the first Brillouin zone is a hexagon, as shown in the inset of Fig. 6, in which two high-symmetry points, K and M, as well as the zone center, Γ, are also indicated. Because of the sixfold rotational

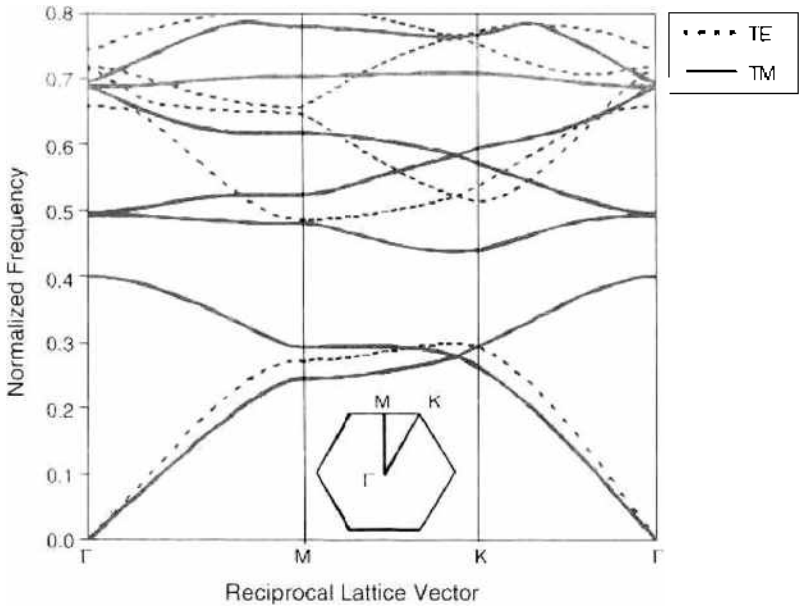


Figure 6. Photonic band structure for a triangular array of air holes in a dielectric material with $\varepsilon = 12$. The solid line is for TM modes, and the dashed line is for TE modes. The inset shows the hexagonal Brillouin zone of the triangular crystal structure.

symmetry of the triangular lattice, there are six equivalent K and M points. It is noted that the Γ–K direction points along the nearest-neighbor direction in the real space triangular lattice, as the real space triangular lattice and the reciprocal lattice are rotated by 30° with respect to each other. Figure 6 shows a complete PBG opening up for both TE and TM polarizations for normalized frequencies ($\omega a/2\pi c$) from 0.40 to 0.44. Early theoretical studies have found that a system with isolated high dielectric material tends to support TM band gaps, whereas a system with interconnected high dielectric material favors TE band gaps [43]. Consistent with this general observation, the TE band gap shrinks rapidly as the air hole radius is further increased and completely disappears at approximately $r = 0.5a$, at which the air holes begin to overlap, forming isolated dielectric columns. In contrast, the TM band gap tends to decrease for smaller air holes and closes up at around $r = 0.4a$. Therefore, the complete PBG is supported in a relatively narrow region in which the triangular lattice exhibits the characteristics of both interconnected and isolated dielectric systems.

The emergence of PBG is closely related to the symmetry of the system. Quantum mechanics tells us that degeneracy originates from symmetry, and thus the lowering or breaking of symmetry can lift degeneracy, possibly opening up complete PBG. Symmetry lowering can be achieved either by using a multiple atom basis or by making the atoms themselves nonsymmetric. The first scheme was realized first in the three-dimensional system in which a large complete PBG was predicted in a diamond lattice, whereas the simple face centered cubic (fcc) lattice was found not to support a complete PBG [20]. The same approach is certainly possible for two-dimensional PC structures, as well. A square lattice of circular air holes in a dielectric material ($\varepsilon = 11.4$) supports TE band gaps for small air hole radii and TM band gaps for large air-hole radii, as shown in Fig. 7a. A complete PBG for both polarizations opens up in a very narrow range of air hole radius around $r = 0.5a$. The maximum width of the PBG is also very small, $\Delta\omega = 0.0188$. Attempts to increase the PBG width by simply changing the air hole radius are not effective because of the conflicting tendencies of the TE and TM band gaps. After close examinations of the band structures, Anderson and Giapis recognized that the TM band gap closes at large air hole radii because the second and third bands tend to be degenerate at the M point in the first Brillouin zone [44]. They attempted to lift the degeneracy by inserting an additional small air hole at the center of each square unit cell. The resulting new structure is still a square lattice, but with a two-atom basis. The subsequent changes in photonic band structure depend on the size of the second air hole. It was found that the largest increase in PBG width is realized when the radius ratio of

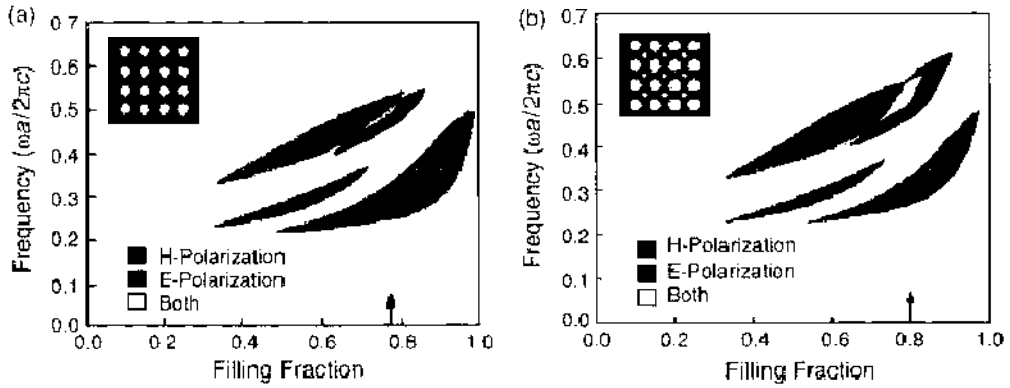


Figure 7. Gap map for (a) simple square lattice and (b) reduced-symmetry square lattice of air holes in a background dielectric material with $\varepsilon = 11.4$. The reduced symmetry increases the second TM gap, resulting in much increased complete PBG. Arrow indicates the occurrence of maximum PBG. Reprinted with permission from [44], C. M. Anderson and K. P. Giapis, *Phys. Rev. Lett.* 77, 2949 (1996). © 1996, American Physical Society.

the two air holes $\beta = 0.1 \sim 0.2$. Figure 7b shows the gap map for the reduced-symmetry square lattice with $\beta = 0.16$. The increase of the second TM gap is clearly visible, resulting in a corresponding increase in the complete PBG. The largest PBG width was $\Delta\omega = 0.0548$, almost three times greater than the largest gap width possible in the simple square lattice. The same scheme can also be applied to the triangular lattice, which may be transformed to honeycomb lattice by incorporating a two-atom basis. The honeycomb lattice was found to exhibit wide, complete PBGs in a wide range of air-hole radii, demonstrating the effect of symmetry lowering [43]. The PBG width may be further increased by inserting small air holes at the center of the hexagonal cells. Anderson and Giapis reported about 10% increase in PBG width when $\beta = 0.11$ [44].

One may also attempt to lower symmetry by incorporating noncircular scatterers. However, early studies indicated that elliptical scatterers in square and triangular lattices did not yield larger PBGs than the circular scatterers, nor did they create new PBGs [45]. It was later recognized that reorienting the asymmetric scatterers off the symmetry directions of the lattice is essential in creating and widening PBGs. Wang et al. investigated a square lattice of square air holes in a dielectric background material [46]. When the refractive index of background material, n_b, is 3.4 and the air filling fraction is 0.5, this structure does not exhibit a complete PBG.

As shown in Fig. 8a, this interconnected dielectric structure exhibits a large gap for TE polarization but not for TM polarization, which is consistent with the general rule of thumb. In fact, this structure was found to not support complete PBG at any filling fraction, indicating that the mere change of scatterer shape is not enough to significantly alter photonic band structures. However, because of the noncircular shape of the individual scatterers, further reduction in symmetry is possible by rotating the scatterers so that they would be oriented off the symmetry axis of the lattice. This reorientation of noncircular scatterers was found to have profound effects on photonic band structures. When the square holes are oriented 45° off the symmetry axis of the square lattice, the degeneracy of the third and fourth TM bands at Γ point is lifted, and the second and third TE bands are also further separated, opening a complete PBG, as shown in Fig. 8b.

Comprehensive numerical modeling study for various reorientation angles and filling fractions revealed that the largest complete PBG is obtained for $\theta = 30°$ and $f = 0.68$. The resultant PBG width was $\Delta\omega = 0.063$, or in terms of the gap width to midgap frequency ratio, $\Delta\omega/\omega_c = 14.2\%$, which is much greater than the largest gap width, $\Delta\omega/\omega_c = 4\%$, achievable in the triangular lattice of circular air holes. It should also be noted that the smallest feature size, which is the width of the dielectric material between two adjacent square holes in this structure is $\sim 0.05a$ for the structure with maximum PBG width. This is much greater than that for a triangular lattice of circular holes, $\sim 0.003a$, for the structure exhibiting maximum PBG width. The larger minimum feature size, as well as the elimination of the need for patterning circular holes, makes the fabrication process more tractable.

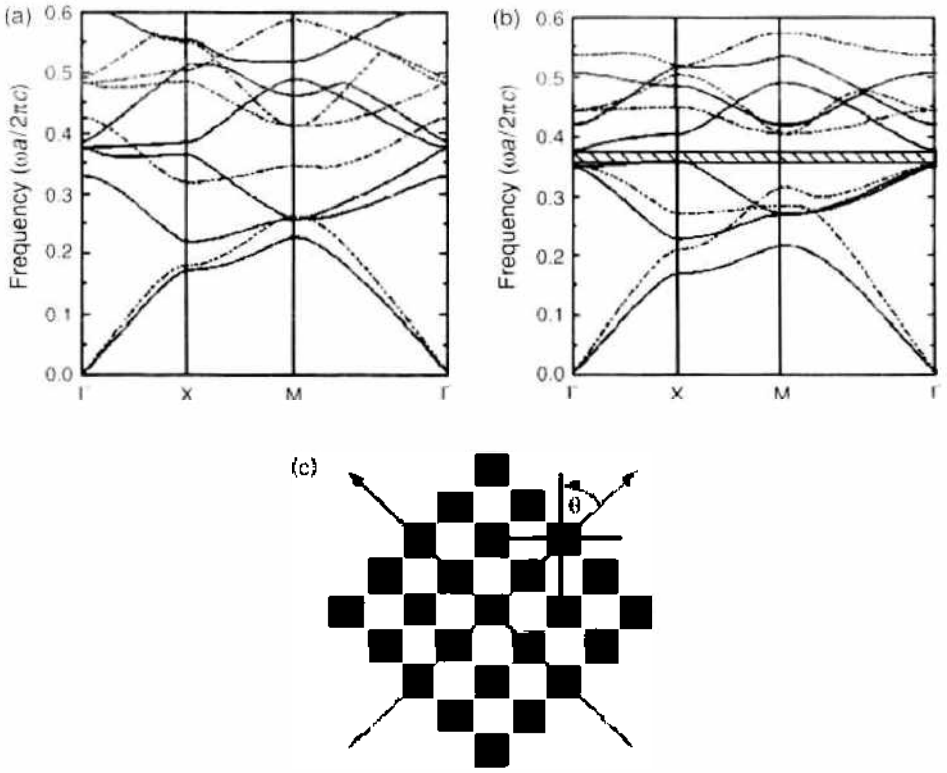


Figure 8. Photonic band structures for a square lattice of square air-rod structures with background index $n_b = 3.4$ and filling fraction $f = 0.5$: (a) $\theta = 0°$, and (b) $\theta = 45°$. The solid and dotted lines correspond to the TM and TE polarization modes, respectively. The reorientation angle θ is defined in (c). Reprinted with permission from [46], X. H. Wang et al., *Phys. Rev. B* 60, 11417 (1999). © 1999, American Physical Society.

A comprehensive survey was later made on the dependence of two-dimensional PBG on the types of lattice and shapes of the scatterer [47]. In the triangular lattice formed by air rods in dielectric background, the hexagonal rods tilted by 24° off the lattice symmetry axis produced the largest complete PBG for which the gap width to midgap frequency ratio amounted to 23.6% for a filling fraction $f = 0.805$. For circular rods, the maximum PBG width was $\Delta\omega/\omega_o = 19.5\%$ for a filling fraction $f = 0.843$. In contrast, square- and rectangular-shaped rods yielded smaller PBG, even with reorientation. Honeycomb lattice exhibited a similar behavior, with hexagonal and circular rods producing larger PBGs and with square and rectangular rods showing smaller PBGs.

Unlike the triangular lattice, however, the honeycomb lattices with reoriented hexagonal rods were found to yield about the same PBG width as the circular rods. This could be explained by the fact that in the honeycomb lattice the symmetry is already lowered from the original triangular lattice by the two-atom basis, and thus the further symmetry reduction by the shape of atoms has smaller effect on the PBG. In the square lattice of air rods in a dielectric medium, square rods exhibited the largest gap width. The maximum gap width to midgap ratio was $\Delta\omega/\omega_o = 14.9\%$ when $f = 0.68$ and the reorientation angle $\theta = 30°$. Circular and hexagonal rods yielded much smaller gaps with maximum gap width to midgap ratios of only 5.1% and 3.1%, respectively. It is interesting to note that the behavior of triangular and square lattices are almost opposite and that the scatterers that have the same symmetry as the lattice tend to give the maximum PBG width. The results on other types of lattices and scatterer shapes are listed in Table 2. Summarizing the studies on two-dimensional PCs, a variety of two-dimensional structures, some of which seem quite attractive from the fabrication standpoint, exhibit complete PBGs, and the maximum gap width was found in the triangular lattice of misoriented hexagonal scatterers.

So far, we only discussed pure two-dimensional structures that are assumed to be invariant in the third dimension. This type of structure can be made on Si or Al_2O_3 by an electrochemical process in which ordered pores are created with an extremely high aspect ratio [48, 49]. The use of prepatterned substrates improves high crystallinity and also allows the creation

Table 2. Dependence of photonic band gaps on the lattice types and scatterer shapes.

Lattice type	Shape of Brillouin zone	Scatter	f	θ	ω_R	$(\Delta\omega)_{max}$	ω_R	$\Delta\omega_R$
Triangular[a]	hexagon	hexagon	0.805	24°	0.449	0.106	23.6%	2.1%
		circle	0.843	—	0.481	0.094	19.5%	—
		square	0.843	30°	0.476	0.059	12.3%	0.6%
		rectangle	0.678	0°	0.378	0.034	8.9%	1.3%
Honeycomb[b]	hexagon	hexagon	0.143	30°	0.544	0.058	10.6%	0.0%
		circle	0.150	—	0.539	0.055	10.3%	—
		square	0.405	41°	0.200	0.011	5.4%	0.1%
		rectangle	0.383	58°	0.203	0.013	6.3%	0.0%
Square[c]	square	square	0.680	30°	0.425	0.063	14.9%	2.8%
		rectangle[d]	0.585	45°	0.386	0.030	7.6%	0.6%
		circle	0.770	—	0.475	0.024	5.1%	—
		hexagon	0.780	15°	0.488	0.015	3.1%	0.0%
Square[b]	square	square	0.395	2°	0.648	0.037	5.8%	2.5%
		rectangle[e]	0.495	39°	0.356	0.015	4.1%	6.0%
		ellipse[e]	0.355	41°	0.493	0.013	2.6%	13.5%
Rectangular[bc]	rectangle	rectangle[f]	0.688	28°	0.480	0.053	11.0%	2.7%
		ellipse[f]	0.710	40°	0.728	0.031	4.3%	1.2%
		hexagon	0.620	30°	0.430	0.012	2.8%	0.1%
		square	0.555	22°	0.419	0.007	1.8%	0.6%

[a] Air rods in dielectric media.
[b] Dielectric rods in air.
[c] Lengths of lattice vectors satisfy $l_y/l_x = 0.8$, frequencies are in units of $2\pi c/l_x$.
[d] $L_b/L_a = 0.84$.
[e] $R_b/R_a = 0.44$.
[f] $R_b/R_a = 0.64$.
Source: Reprinted with permission from [47], R. Wang et al., *J. Appl. Phys.* 90, 4307 (2001). © 2001, American Institute of Physics.

of waveguide or cavity structures. However, despite the recent progress in this area, the two-dimensional structures with finite thickness that can be fabricated by the conventional lithography-based techniques still bear the highest practical importance. Naturally, extensive research has been carried out on two-dimensional slab PCs with finite thickness. A slab PC typically consists of a thin layer of high-index material that is periodically patterned. The slab may lie on a low-index substrate or be suspended in air.

In either case, the electromagnetic fields are scattered by the periodic structure in the plane of the dielectric slab, and at the same time they are confined within the slab because of the index profile along the direction perpendicular to the slab. Therefore, the confinement of light within the slab is imperfect and only the modes that satisfy the total internal reflection condition survive; the rest radiate into the surrounding media with short lifetimes. This imperfect light confinement consequently imposes limitations on the functionality of slab PCs. However, it has been discovered that many of the desired properties of pure two-dimensional structures are retained by the guided modes in two-dimensional slab PCs.

From the modeling standpoint, the photonic band structure of two-dimensional slab PC may be calculated by the PW or FDTD methods. In the PW method, where the periodic boundary condition must be used, it is necessary to set up an artificial supercell that contains the slab in the middle sandwiched by thick layers of background medium, typically air. It is important to ensure that the supercell size is large enough that the coupling between the slabs in the adjacent supercells is negligible. In this case, the guided modes whose field amplitudes decay exponentially outside the slab are not affected by the artificial periodicity along the out-of-plane direction, and thus the bands within the guiding regime are accurately modeled. The radiation modes, however, may possess substantial field amplitudes even at a large distance away from the slab. Thus, these modes are always affected by the supercell and cannot be faithfully modeled by the PW method.

When accurate band structures are needed for both guided and radiation modes, it is desirable to use the FDTD technique, in which an absorbing boundary condition can be used to accurately simulate a medium with infinite extent. Thus, for the modeling of a slab PC structure, the FDTD computation cell is typically terminated by periodic boundary conditions for

the directions of two-dimensional periodicity and by absorbing boundary conditions in the direction perpendicular to the slab. In any case, the most important characteristic of the band structure of a two-dimensional slab PC that distinguishes itself from that of a pure two-dimensional PC is the presence of light line that forms the boundary between the guided and radiation regimes. When the slab is located in a uniform dielectric medium, the light line is simply given by the dispersion relation of the background medium. In addition, the system is no longer invariant along the direction perpendicular to the slab.

As a consequence, the modes in a slab PC can no longer be classified into two noninteracting TE and TM polarizations. Instead, the slab possesses mirror symmetry (i.e., the system remains invariant on reflection in the plane lying in the middle of the slab, if the slab is surrounded by the same dielectric material on both sides). Therefore, the modes in the two-dimensional slab PC are either even or odd, depending on how they transform on reflection in the mirror plane. The even and odd modes in a slab PC bear close resemblance to TE and TM modes in the pure two-dimensional PC. In fact, in the mirror plane, even and odd modes do have the same polarization as the TE and TM modes, respectively. For this reason, the even and odd modes are sometimes referred to as TE-like and TM-like modes.

Figure 9 shows the photonic band structures for two-dimensional slab PCs calculated by the PW method [50]. The shaded regions represent the radiation regime, in which modes are unbounded and thus can couple out of the slab. The modes under the light line, in contrast, are guided and are thus confined within the slab. Because of the presence of a continuum of radiation modes, it is clear that there is no complete PBG in these systems. However, as indicated in Fig. 9, there are regions of frequency in which there exist no guided modes. These regions are frequently called PBGs, in which it is implicitly understood that only the guided modes are concerned. The square lattice of finite dielectric rods exhibits a PBG for odd modes, whereas the triangular lattice of air holes in a dielectric slab shows a PBG for even modes. This is consistent with the fact that the corresponding pure two-dimensional structures exhibited TM and TE gaps, respectively, confirming the similarity between the even and odd modes in slab PCs and the TE and TM modes in pure two-dimensional structures. Furthermore, the lower modes were found to be tightly confined within the slab, and their field profiles also closely resemble those of corresponding modes of the pure two-dimensional structures. In the slab PC structure, in addition to the usual parameters such as the type of lattice, shapes of the scatterers, and the filling fraction, the slab thickness critically affects the photonic band structure. If the slab is very thick, it will support many guided modes, which would naturally decrease the chance of opening a PBG. If the slab is too thin, however, it will support only weakly guided modes that will exist close to the light line. Thus, it is conceivable that there will exist an optimum thickness leading to the maximum PBG width. Johnson et al. reported that the maximum PBG width was

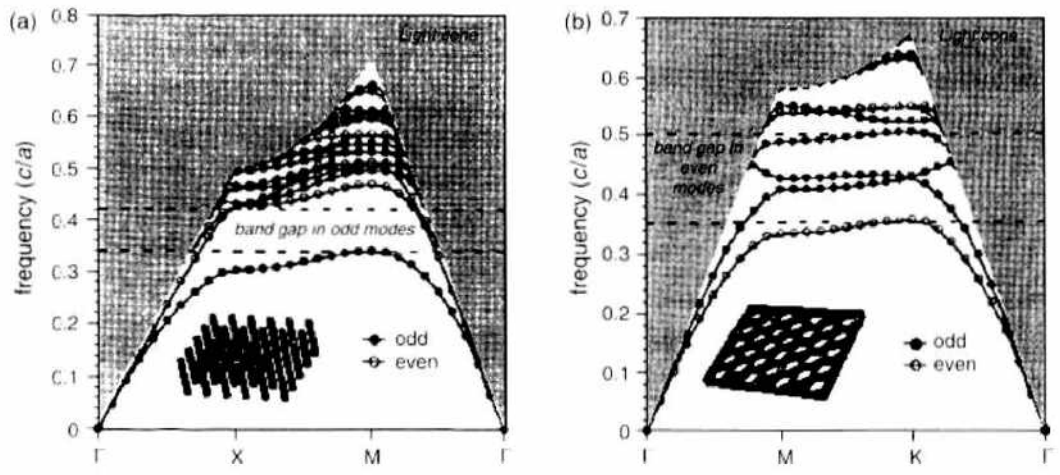


Figure 9. Photonic band structures of two two-dimensional slab photonic crystals. (a) square lattice of dielectric rods with finite height, and (b) triangular lattice of air rods in a dielectric slab. Reprinted with permission from [50], S. G. Johnson et al., *Phys. Rev. B* 60, 5751 (1999). © 1999, American Physical Society.

found at thicknesses of $2.3a$ and $0.6a$ for dielectric rod and air hole lattices, respectively [50]. The optimum thickness for maximum PBG width will, of course, change with the exact dimensions of the two-dimensional PC structures, but from the large difference in optimum thickness between dielectric rods and air holes systems, one notices that the polarization plays an important role in determining the optimum thickness.

Although the previous analyses showed that two-dimensional slab PCs support guided modes that can propagate losslessly within the slab, when the perfect periodicity is disturbed either intentionally or unintentionally, the guided Bloch modes can couple with the radiation modes and leak out of the slab. Accurate estimation of radiation losses is therefore very important for practical applications based on two-dimensional slab PC structures. For a quantitative description of radiation losses, Benisty et al. considered a two-dimensional PC of a triangular array of air holes in a dielectric slab and developed a phenomenological model in which the losses arise from the radiations by effective dipoles located inside the air holes [51]. The strength of the oscillating dipole is determined by the average field inside the hole. The two-dimensional periodicity would result in Bragg scattering that can strongly modify the radiation losses, but this effect was assumed to be small, and the dipoles were assumed to radiate independent of one another. It was further assumed that the radiation is isotropic from normal direction to the critical angle—below which the light will be confined within the slab by total internal reflection. As a result, the model predicts that the radiation loss be proportional to $(\Delta\varepsilon)^2$, where $\Delta\varepsilon$ is the difference of dielectric constant between the slab and cladding materials.

To incorporate this loss term into a pure two-dimensional modeling, the holes are treated as a dissipative medium with an imaginary dielectric constant that gives rise to a power dissipation equal to the radiation loss. This fictitious imaginary dielectric constant allows the use of simpler and faster two-dimensional modeling tools for quantitative estimation of radiation loss in two-dimensional slab PCs. Figure 10 shows the transmission coefficients calculated by the transfer matrix method along with the experimental ones measured from a GaAs-based two-dimensional slab PC. The theoretical curves calculated with no loss term show sharp, well-defined edges at both the low and high frequency boundaries of the stop bands where the transmission coefficient abruptly rises from zero to close to 1. However, the

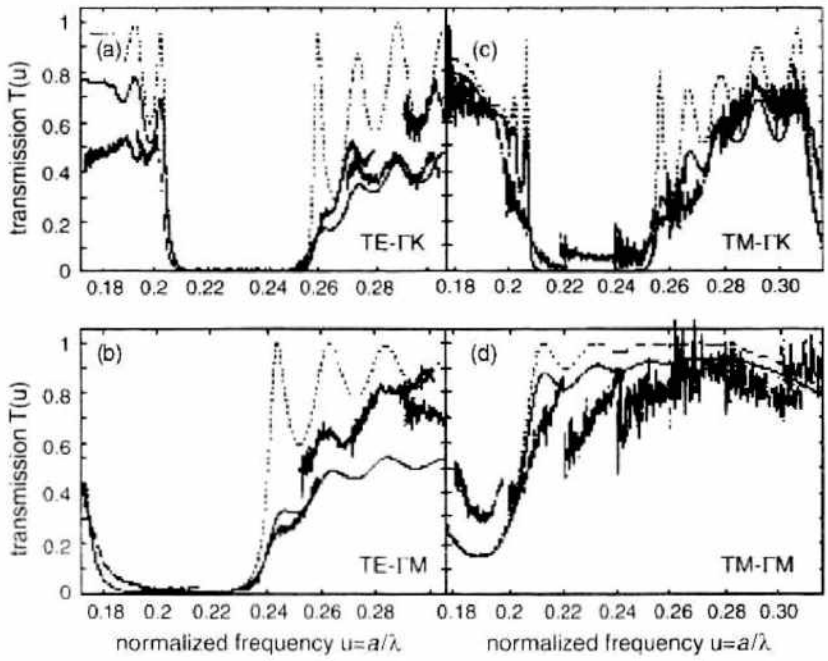


Figure 10. Comparison of experimental transmission of eight-rows-thick two-dimensional photonic crystals (thin line) with calculations with losses $\varepsilon'' = 0.18$ (bold line) and without losses (dashed line) as a function of the normalized frequency $u = a/\lambda$. Orientations (ΓK on top and ΓM on bottom) and polarizations (TE to the left and TM to the right) are indicated. Reprinted with permission from [51], H. Benisty et al., *Appl. Phys. Lett.* 76, 532 (2000). © 2000, American Institute of Physics.

experimental curves indicate that the edges of the stop bands are smeared out and do not show sharp edges, particularly on the high-frequency side. The theoretical curves calculated with nonzero imaginary dielectric constant better represent the experimental results. The strong decrease in transmission on the high-frequency side of the stop band is readily understood if one considers the field profiles. The photonic band below the stop band has most of its fields in the dielectric region (dielectric band), whereas the band above the stop band has its field concentrated in the holes (air band). This naturally leads to a higher loss for the air band, which is correctly reflected in the calculated spectra.

The above model is somewhat simplistic because it disregards the guiding condition for the slab and treats the holes as independent radiators irrespective of the frequency. Numerical simulations by Bogaerts et al. later found that that the photonic bands of a slab PC that lie below the light line do not exhibit radiation loss in the bulk, as expected [52]. The only radiation sources in this case are defects, which include intentional and unintentional defects and also the interfaces that inevitably exist in a finite-size PC structure. It was also discovered that the radiation loss is proportional to $(\Delta\varepsilon)^2$ only for small $\Delta\varepsilon$ and tends to level off for high $\Delta\varepsilon$. Thus, it is advantageous to keep $\Delta\varepsilon$ as low as possible when $\Delta\varepsilon$ is relatively low because a small decrease in $\Delta\varepsilon$ would result in a large decrease in radiation loss. When $\Delta\varepsilon$ is large, however, the radiation loss becomes insensitive to $\Delta\varepsilon$.

4.1.3. Three-Dimensional Photonic Crystals

Three-dimensional PC is needed to produce complete three-dimensional PBG, in which light propagation or generation is absolutely forbidden. Unlike the two-dimensional slab PC in which one needs to worry about guidance condition and out-of-plane losses, a three-dimensional PC exhibiting complete three-dimensional PBG can serve as a platform where light is controlled entirely by the photonic band structure. Three-dimensional PCs hold high promises for novel applications such as three-dimensional optical interconnects and nanocavities with extremely high confinement factors. The most widely studied three-dimensional system is fcc structure. In the original proposal of the PC concept, it was suggested that an fcc crystal would produce the most round-shaped Brillouin zone in the reciprocal space, thereby increasing the chance that a PBG would encompass the entire Brillouin zone surface [1]. Soon after, Yablonovitch et al. manufactured fcc crystals consisting of dielectric spheres, air spheres and elongated air spheres, and experimentally demonstrated in the microwave region that complete three-dimensional PBG is possible in some of these fcc crystals [53, 54].

These pioneering works were followed by extensive experimental research on creating fcc crystals with atoms small enough for optical applications. In particular, the rapid progress in nanoparticle self-assembly technique has provided a fast and economical means of producing high-quality fcc crystals [55, 56]. Early modeling studies on fcc crystals were primarily based on scalar-wave approximation [57, 58]. Ignoring the vector nature of the electromagnetic fields, this approach incorrectly suggested that simple-cubic, body-centered-cubic, face-centered-cubic, and diamond structures all exhibit complete three-dimensional PBG.

It was soon realized that the vector nature of the electromagnetic field critically affects the photonic band structure. Thus followed more rigorous theoretical investigations, taking the vector nature fully into account, which determined that a simple fcc structure did not support a complete PBG [20]. In clear contrast to the scalar-wave approximation, the vector calculation showed the second and third photonic bands are degenerate at W and U points of the Brillouin zone, preventing the complete PBG. An alternative structure was thus sought, having lower symmetry that can lift the degeneracy at W and U points but at the same time preserve the round shape of the Brillouin zone, which is believed to be favorable for complete PBG. On the basis of this reasoning, a diamond lattice made of two-atom basis on fcc lattice was modeled and a wide complete PBG was observed. Note the analogy with the two-dimensional systems, for which we observed much larger PBG by going from a simple triangular lattice to a honeycomb lattice made of a two-atom basis on triangular lattice.

Figure 11 shows the photonic band structure for a diamond lattice, comprising dielectric spheres with a refractive index of 3.6 and filling ratio of 0.34. A complete PBG opens between

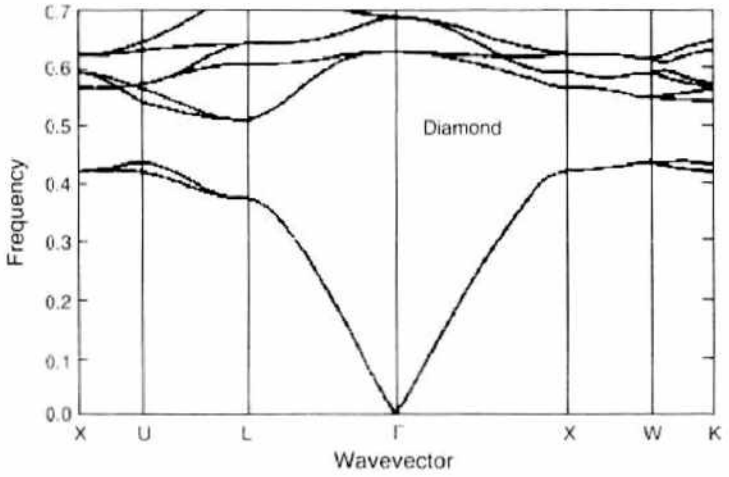


Figure 11. Photonic band structure of a diamond lattice consisting of dielectric spheres with refractive index 3.6 in an air background. Reprinted with permission from [20], K. M. Ho et al., *Phys. Rev. Lett.* 65, 3152 (1990). © 1990, American Physical Society.

the second and third photonic bands and encompasses the entire Brillouin zone including the W and U points. Further studies discovered that the diamond lattice supports a complete PBG if the refractive index contrast exceeds 2, and that the lattices with air spheres tend to provide wider PBGs. The largest gap width was $\Delta\omega/\omega_o = 28.8\%$ for air spheres with filling fraction 0.81.

Another method to reduce symmetry and lift the degeneracy is to use nonspherical atoms, as discussed in Section 4.1.2 for two-dimensional systems. This new structure, often called the Yablonovite after its inventor, is made by drilling three sets of holes 35.26° off the vertical axis into a dielectric slab [54]. The three drilling axes are 120° apart from each other on the azimuth. This process results in a network of holes with roughly cylindrical shape that forms a fcc lattice. This structure was found to exhibit complete PBG if the background index is greater than 2.1 and, for a background index of 3.6, the gap width to mid gap ratio was 20%.

Although these modified fcc structures with reduced symmetry exhibit wide and robust PBG, simple fcc structures with spherical atoms are still the most favorable system from the fabrication standpoint. In this respect, it was a significant discovery that a simple fcc structure composed of spherical air atoms in a dielectric background also possesses a complete PBG between the eighth and ninth bands [59], which happened to fall in the high-frequency region not examined in the earliest modeling studies. This structure can be readily produced by the self-assembly technique in which highly monodispersed nanospheres (most often SiO_2 or polystyrene) self-assemble themselves into a fcc close-packed structure. The self-assembled opal then serves as a template into which high-index dielectric material may be infiltrated to create an inverted opal structure comprising air spheres in a dielectric background. As the self-assembly process makes rapid progress, further theoretical investigations on self-assembled opal and related structures followed.

Busch and John determined that the simple fcc structures made by mono-dispersed SiO_2 and TiO_2 nanospheres do not support a complete PBG and that the total photonic DOS changed little from the free-space DOS [60]. Infiltration of dielectric material and subsequent removal of SiO_2 particles, however, can produce a complete PBG if the infiltrate possesses a high enough refractive index. Figure 12a shows the photonic band structure of a Si inverted opal structure that is a fcc lattice of air spheres in a Si background. It can be seen clearly that a complete PBG extends across the entire Brillouin zone between the eighth and ninth bands. The gap width to midgap ratio was $\Delta\omega/\omega = 4.25\%$. The PBG is also visible in the photonic DOS plot in Fig. 12b, in which the total photonic DOS falls to zero in the PBG.

The DOS plot also reveals that there exists a pseudogap between the fourth and fifth bands, where the normalized frequency $\omega \sim 0.524$. In the inverted opal structure, complete

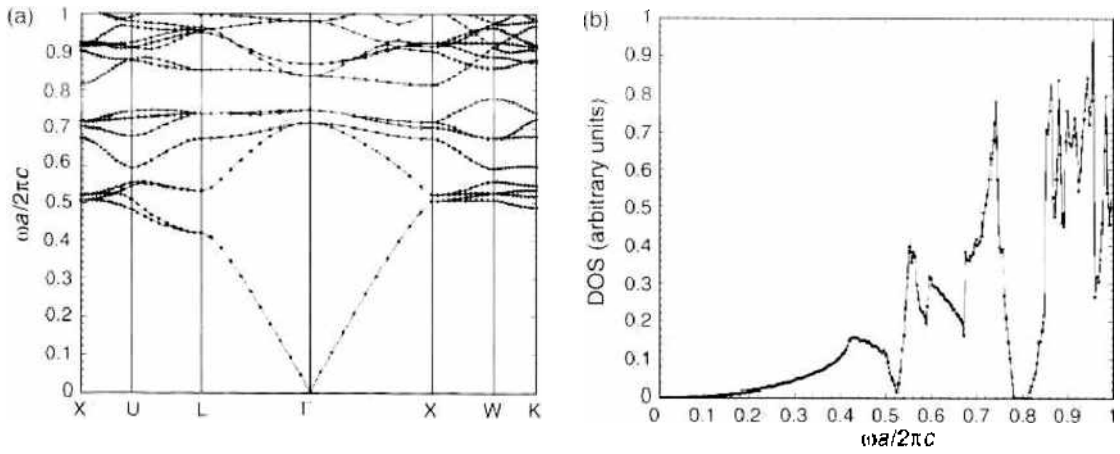


Figure 12. (a) Photonic band structure and (b) photonic density of states for a Si-inverted opal structure; that is, fcc structure of air spheres in a silicon background ($\varepsilon = 11.9$). Reprinted with permission from [60]. K. Busch and S. John, *Phys. Rev. E* 58, 3896 (1998), © 1998, American Physical Society.

PBG begins to open up when the background refractive index exceeds 2.8 and the gap width monotonically increases with an increasing background index. For Ge infiltration ($n = 4.0$), the gap width amounts to $\Delta\omega/\omega = 7.35\%$. It was also found that the hexagonal close-packed (hcp) structure made of air spheres in Si has a pseudogap and a complete PBG at roughly the same frequencies as its fcc counterpart. The width of the PBG, which opens between the sixteenth and seventeenth bands in this case, was somewhat smaller, with $\Delta\omega/\omega = 2.8\%$. During the inverted opal manufacturing process, it is customary to sinter the self-assembled opal so that the nanoparticles bond together, providing enough structural integrity to withstand further processing. After infiltration, the bonded particles provide passage for the chemical etchant to flow through, enabling complete removal of original template. Therefore, the real inverted opals are not truly close-packed structures of touching air spheres but are better modeled by a fcc structure of slightly overlapping air spheres. The photonic band properties will then be affected by the degree of overlapping.

Busch and John investigated the dependence of PBG width in a Si-inverted opal on the size of the opening created by the bonded particles [60]. The PBG width was found to increase as the opening size was increased. A maximum width of $\Delta\omega/\omega \approx 6.6\%$ was obtained when the radius of the opening was $0.133a$, where a is the lattice constant of the fcc lattice. Considering that the sintering improves the connectivity of air, this result indicates that the formation of PBG is promoted by the presence of true network topology.

The same authors also investigated the effect of incomplete infiltration, another strong possibility in inverted opal fabrication. By assuming that the air spheres are coated with Si rather than immersed in a completely filled Si background, it was discovered that the PBG width could be almost doubled from the full infiltration case when the Si volume fraction was 21%. When the infiltration becomes more incomplete, the PBG is eventually closed because there is not enough high-index material in the system. However, as the infiltration becomes more and more complete, the gap width approaches the value for the Si-inverted opal.

Another three-dimensional structure that has received much attention is the woodpile structure first proposed by Ho et al. [61] and fabricated by Lin et al. [62]. As shown in Fig. 13a, this structure is derived from the diamond structure by replacing the ⟨110⟩ chains of atoms with the rods and has a layered design so that it can be fabricated by the lithography technique. Each layer consists of parallel rectangular rods whose orientation is rotated by 90° as they are stacked on top of one another. Furthermore, the rods are shifted by half the spacing between every other layer. Therefore, four adjacent layers are combined to form a unit stacking sequence. This structure in general forms a face-centered-tetragonal (fct) lattice and in the special case in which $c/d = 1.414$ (c and d are indicated in Fig. 13a), it forms an fcc lattice. The photonic DOS calculation for the woodpile structure made with Si shows a wide, complete PBG opening between frequencies 0.46 and 0.56, resulting in $\Delta\omega/\omega \approx 20\%$.

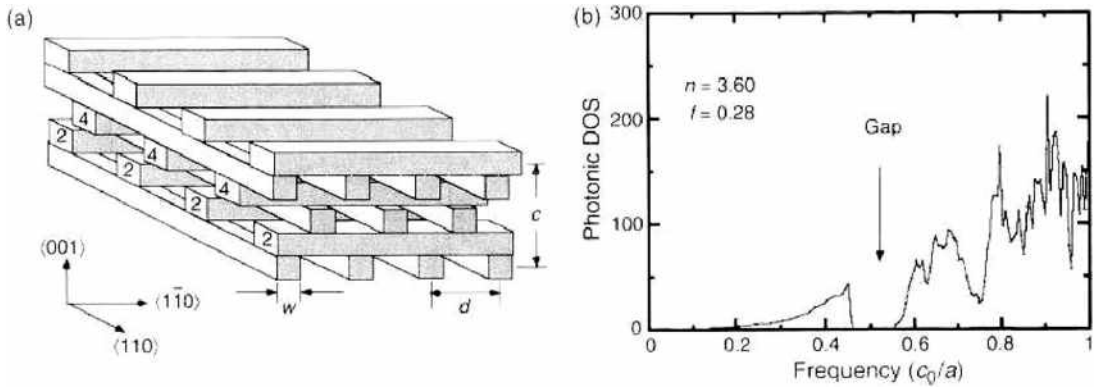


Figure 13. (a) Schematic diagram of the woodpile structure. (b) calculated photonic density of states for rod index $n = 3.6$, $f = w/d = 0.28$ and $c/d = 1.414$. Reprinted with permission from [62]. S. Y. Lin et al., *Nature* 394, 251 (1998). © 1998, Nature Publishing Group.

The presence of PBG was also confirmed by optical transmission measurements, which showed a strong decrease in transmission in the midinfrared region.

Other structures that were also found to support complete PBG involve atoms with a cylindrical shape. Chan et al. discovered that a whole class of structures with rhombohedral symmetry, known as the A7 structure, supports wide three-dimensional PBGs and that many structures known to possess complete three-dimensional PBG are in fact a subset of this broader class of materials [63]. The A7 structure is a rhombohedral lattice with a two-atom basis. The structure is characterized by the shear angle, the angle between two primitive translation vectors, and the displacement of two basis atoms. A system in which circular cylinders connect the lattice points, thereby producing interconnected topology for both high and low dielectric materials was studied in Ref. [63]. The structures created this way were found to exhibit wide PBG for a variety of values of shear angle and displacement of basis atoms, except when the combination of the two parameters accidentally produced a high-symmetry structure. The maximum PBG width was found to be as large as $\Delta\omega/\omega = 30\%$.

Another structure designed for lithographic fabrication is composed of multiple layers of dielectric slabs patterned with a triangular array of air holes [64]. These slabs are separated by a triangular array of dielectric columns, and the positions of the air holes in the adjacent slabs are shifted according to the stacking sequence of a fcc lattice along the (111) direction. The resulting structure is essentially a fcc lattice with nonsymmetric atoms. The photonic band structure exhibited a complete PBG of over 21% for a Si–air system and the PBG was found to persist down to a refractive index contrast of 2:1. Furthermore, when intentional defects were introduced for waveguides or cavities, this structure was expected to support modes similar to those of a two-dimensional slab PC because it comprises stacks of two-dimensional slab PCs. This would allow one to use the extensive research carried out on two-dimensional slab PCs, as discussed in the previous section.

It has recently been recognized that wide and robust three-dimensional PBG may be realized by using metals that are generally strong light scatterers. Periodic metallic or metal–dielectric structures in both two-dimensional and three-dimensional geometry have been widely investigated for microwave or millimeter-wave regions in which most metals behave as perfect conductors. At optical frequencies, however, metals become strongly absorbing and dispersive, hindering the development of metal-based PCs for photonic applications. It was recently discovered that absorption loss in metallic PC can be made very small [65]. Transfer matrix calculations on simple cubic structures showed that structures with interconnected metal exhibit higher loss than those with isolated metals. This is because the long-range conduction current that can flow in the interconnected geometry leads to higher absorption. Among the metal species studied, Cu exhibited the lowest loss, and Au and Ag were found acceptable, though the loss was slightly higher. Al showed the highest absorption loss and is thus not recommended for optical applications.

To properly model metallic PC, the issue of dispersion must be addressed. Metals generally possess a strongly frequency-dependent complex dielectric constant that is frequently described by the Drude model.

$$\varepsilon(\omega) = 1 - \frac{\omega_p^2}{\omega(\omega + i\gamma)} \tag{46}$$

where ω_p is the plasma frequency and γ is the damping parameter, which gives rise to absorption. This poses a serious problem for the PW method. A study of two-dimensional metallic PCs showed that the eigenvalue equation by the PW expansion coefficients is no longer linear in ω^2/c^2, and thus must be linearized by taking a larger basis set [66]. It was found that three or four times larger basis sets are needed, depending on the polarization, resulting in a 27–64 times heavier computational load. Thus, extension of the PW method to three-dimensional metallic or metal-dielectric systems seems difficult. In contrast, the dispersion does not bring any new problem to the KKR method, as it calculates allowed **k** vectors for given ω. Thus, KKR method is one of the most widely used modeling techniques for metallic systems. As pointed out in Section 3.4, the KKR method can handle spherically symmetric scatterers very efficiently. This naturally directs much attention to the opaline structure that may be produced from the colloidal suspension of metallic nanospheres.

Moroz pointed out that a fcc lattice constructed with nonabsorbing metal spheres can induce a wide PBG because $\varepsilon(\omega_p) = 0$, which subsequently results in a very high dielectric constant contrast at frequencies near ω_p [67]. His KKR calculations showed that the metal fcc structures indeed possess complete three-dimensional PBGs. As shown in Fig. 14, the photonic band structure of the metallic system is similar to that of a purely dielectric system, but the complete PBG between the eighth and ninth bands is now much greater than the inverted opal structure. In addition, there is a second PBG opening between the fifth and sixth bands. The dependence on the metal fill fraction revealed that the largest PBG opens in the close-packed structure with $f = 0.74$. In this case, $r_s n_h/\lambda_p > 0.9$ is required, where r_s is the metal sphere radius, n_h is the refractive index of the background dielectric material, and λ_p is the plasma wavelength. The metal lattice in air ($n_h = 1$) represents the worst case in which the dielectric contrast is lowest. Even in this case, however, a complete PBG as wide as $\Delta\omega/\omega = 9\%$ opens when $r_s n_h/\lambda_p = 1.013$. This PBG is almost twice as large as that in the Si-inverted opal structure. It was later found that complete PBG opens in any periodic system containing metal spheres or metal-coated dielectric spheres.

Approximately treating metal with a large negative dielectric constant that is frequency independent, Zhang et al. found complete PBGs in all simple-cubic, fcc, body-centered-cubic,

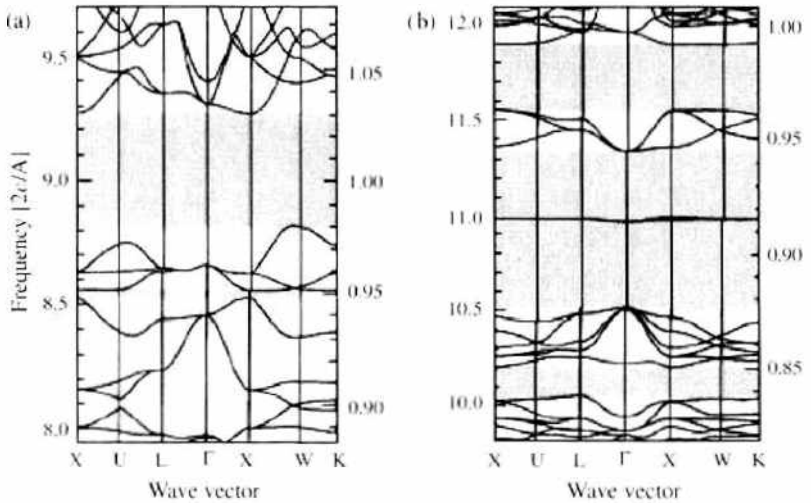


Figure 14. Photonic band structures for fcc lattices of metal spheres. The background dielectric material was assumed to be air. (a) $\omega_p = 9$ and $r_s/\lambda_p = 1.013$ and (b) $\omega_p = 12$ and $r_s/\lambda_p = 1.35$. Reprinted with permission from [67]. A. Moroz, *Phys. Rev. Lett.* 83, 5274 (1999). © 1999, American Physical Society.

body-centered-tetragonal, hexagonal-close-packed, diamond, and hexagonal diamond structures made of metallic cores with dielectric coating whose thickness is 5% of the radius [68]. The midgap frequency and the gap width are shown in Fig. 15. In the inset, the gap width to midgap frequency ratio is plotted as a function of the fill fraction of the metal core. Complete PBG begins to open at $f = 0.53$ and is seen to monotonically increase with an increasing metal fill fraction. The gap widths are much larger than those predicted by Moroz [67]. This disparity seems to arise primarily from the difference in frequencies. Moroz investigated the frequency region near the plasma frequency, but Zhang et al. [68] primarily looked at frequencies much smaller than the plasma frequency by assuming a large negative dielectric constant.

The same technique was extended to deal with the frequency dependence of the dielectric constant and was applied to fcc structures made with Ag, Cu, and Ni [69]. Despite the differences in plasma frequency, all three metals exhibited complete PBG in the visible and near-infrared region, when the metal nanosphere radius was 160 nm. As before, the PBG width was seen to increase with increasing metal fill fraction and reached $\Delta\omega/\omega = 40\%$ for case of close-packed structure of Ag nanospheres. Despite the different plasma frequencies, the minimum fill fraction of metal required to open complete PBG was the same for all three metal species. The position of the PBG, however, was seen to follow the same order as the plasma frequency. Ag, which has the largest plasma frequency, has the gap situated at the highest energy, and Ni and Cu followed in the same order as their plasma frequencies.

These modeling studies, however, assumed purely real dielectric constant and thus did not include the effect of absorption. To see the effect of absorption by metals, the same authors performed transfer matrix calculation, taking the experimentally determined complex dielectric constant for each metal species [69]. Figure 16 shows the transmission, reflection, and absorption spectra calculated for four-layer-thick (111) oriented fcc slabs of Ag, Cu, and Ni. The spectra for nonabsorbing Ag plotted for the sake of comparisons shows well-defined stop band and PBG at energies 0.92–1.54 eV and 2.04–2.40 eV, respectively. The presence of absorption strongly distorts the transmission and reflection spectra. For Ni, which has the smallest plasma frequency among the three, the effect of absorption is so strong in the visible and near-infrared, that the characteristic features of the PC are almost completed washed out. In contrast, Ag, which has the highest plasma frequency, exhibited only small changes from the ideal case, preserving the stop band and PBG. The Cu structure represents the intermediate case, in which the stop band is well preserved but the PBG is destroyed. This result once again shows that the choice of proper metal species is very

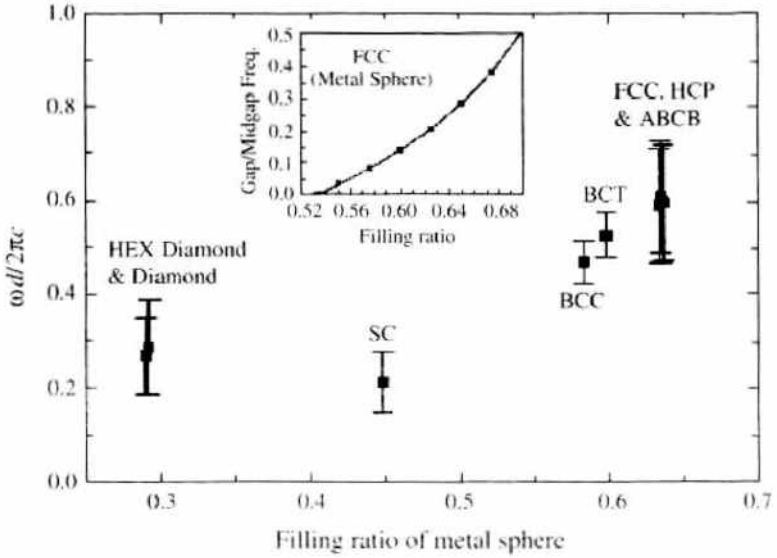


Figure 15. Positions and widths of complete three-dimensional PBGs in various structures made of metal spheres with a thin dielectric coating ($\varepsilon = 12$). The inset shows the gap width-to-midgap frequency ratio for various filling fractions of metal in the fcc lattice. Reprinted with permission from [68], W. Y. Zhang et al., *Phys. Rev. Lett.* 84, 2853 (2000). © 2000, American Physical Society.

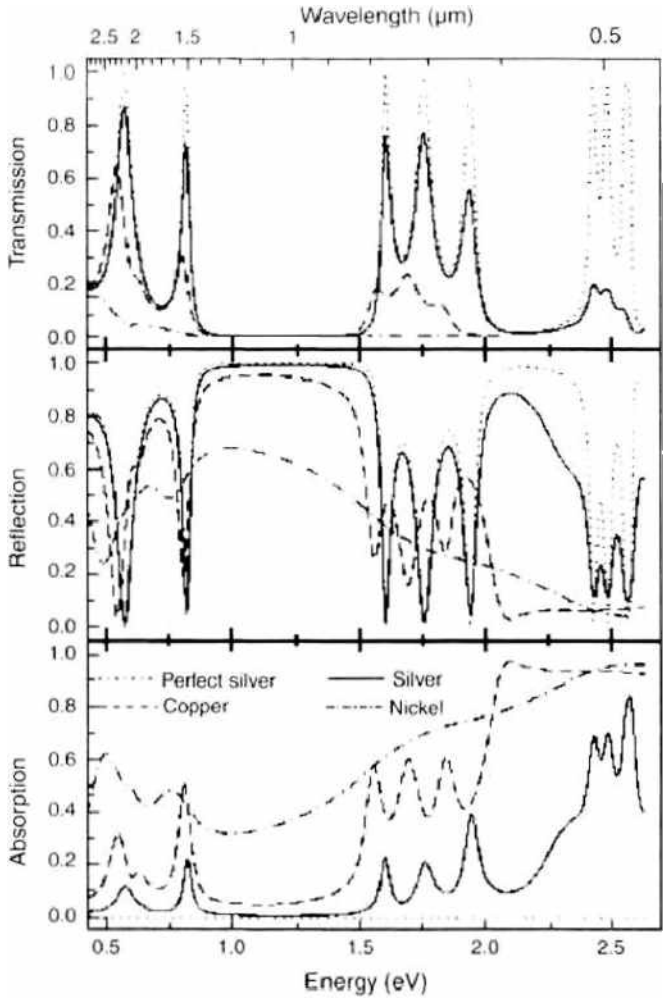


Figure 16. Calculated transmission, reflection, and absorption spectra for four layers of (111)-oriented fcc lattice formed by spherical Ag, Cu, and Ni particles with a radius of 160 nm. Nonabsorbing Ag is also shown for comparison. Reprinted with permission from [69], Z. Wang et al., *Phys. Rev. B* 64, 113108 (2001). © 2001, American Physical Society.

important in metallic or metal-dielectric PCs and that metals with high plasma frequency are desirable.

It is also possible to model metallic or metal-dielectric PCs with the FDTD method. No major change from the purely dielectric system is necessary if one chooses to model the metal with a large negative dielectric constant that is independent of frequency. Assignment of initial field profile does need be changed, however, because the initial field given in Eq. (18) tends to result in incomplete bands. Defining, instead, an oscillating dipole source at a low-symmetry position in the unit cell works much better. The dipole source may be weighted with a Gaussian envelope in time, so that it would excite modes in a range of frequency. A minor concern is that most FDTD codes incorporate the Cartesian coordinate system, and thus it is most convenient to use the conventional cubic unit cell when modeling an fcc lattice. However, because the conventional cubic unit cell is not the primitive unit cell, the corresponding unit cell in the reciprocal lattice is smaller than the first Brillouin zone and thus the resulting band structure will be a folded version. As Fan et al. [70] suggested, to obtain the proper band structure, one needs to set up four dipole sources that are displaced from one another by a primitive unit vector and also their phases satisfy the Bloch theorem. Sakoda et al. later developed an FDTD algorithm that can effectively deal with frequency-dependent dielectric constant [31]. In this method, to be able to handle frequency-dependent dielectric constant, a generalized constitutive equation is used

$$\mathbf{D}(\mathbf{r}, t) = \varepsilon_0 \int_{-\infty}^{\infty} dt' \Phi(\mathbf{r}, t - t')\mathbf{E}(\mathbf{r}, t') \tag{47}$$

where $\Phi(\mathbf{r},t)$ is the dielectric response function given by the Fourier transform of dielectric constant, $\varepsilon(\mathbf{r}, \omega)$

$$\Phi(\mathbf{r}, t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} d\omega\varepsilon(\mathbf{r}, \omega)\exp(-i\omega t) \tag{48}$$

Using the dielectric function, $\varepsilon(\mathbf{r}, \omega)$, given by the Drude model, one finds

$$\Phi(\mathbf{r}, t) = \varepsilon_\infty\delta(t) + \frac{\varepsilon_\infty\omega_p^2}{\gamma}[1 - \exp(-\gamma t)]\theta(t) \tag{49}$$

where θ is the unit step function needed to ensure $\Phi(\mathbf{r}, t) = 0$ for $t < 0$, as required by causality condition. Combining Eqs. (47) and (49), one obtains

$$\frac{1}{\varepsilon_0}\frac{\partial}{\partial t}\mathbf{D}(\mathbf{r}, t) = \varepsilon_\infty\frac{\partial}{\partial t}\mathbf{E}(\mathbf{r}, t) + \varepsilon_\infty\omega_p^2\int_0^\infty dt' e^{-\gamma t'}\mathbf{E}(\mathbf{r}, t - t') \tag{50}$$

for the metal region. For the dielectric region, the usual constitutive Eq. (3), is used. In actual computation, the integral in Eq. (50) is evaluated by using the following recursive equation

$$\mathbf{F}^{(p+1)}(\mathbf{r}) = \Delta t\mathbf{E}^{(p+1)}(\mathbf{r}) + e^{-\gamma t'}\mathbf{F}^{(p)}(\mathbf{r}) \tag{51}$$

where $\mathbf{F}^{(p)}$ represents the integral in Eq. (50) being evaluated by a discrete sum up to the pth time step. Evaluation of $\mathbf{F}$ requires additional storage and computation time, but the additional load was found to be not very serious, amounting to only a few tens of percent increase from the purely dielectric case. The remaining steps are the same as in the regular FDTD method, and the periodic boundary condition extracts only the eigenmodes.

Thanks to the large database on metallic colloids, most metal-dielectric PCs are being made by the self-assembly of metallic or metal–dielectric composite colloidal particles. Recently however, a three-dimensional metallic PC with a woodpile structure was fabricated using a lithography technique [71]. The PC was produced by first creating a SiO_2/Si woodpile structure by layer-by-layer lithography process. Si was then removed from the mold by selective etching, and tungsten was back-filled by chemical vapor deposition. This structure, as explained before, forms a fcc lattice and the incorporation of tungsten proved to open a wide PBG in the infrared region.

As shown in Fig. 17a, the transfer matrix calculation predicted that reflectance would reach almost 100% at wavelengths between 8 and 25 μm, indicating the formation of PBG. Also observed are an absorption peak at 6–7 μm and an allowed band at ~5 μm at which the transmission reaches ~80% in a four-layer system. For a six-layer structure, the transmission peak height increased to ~95%, which is consistent with the photonic band effect. To further

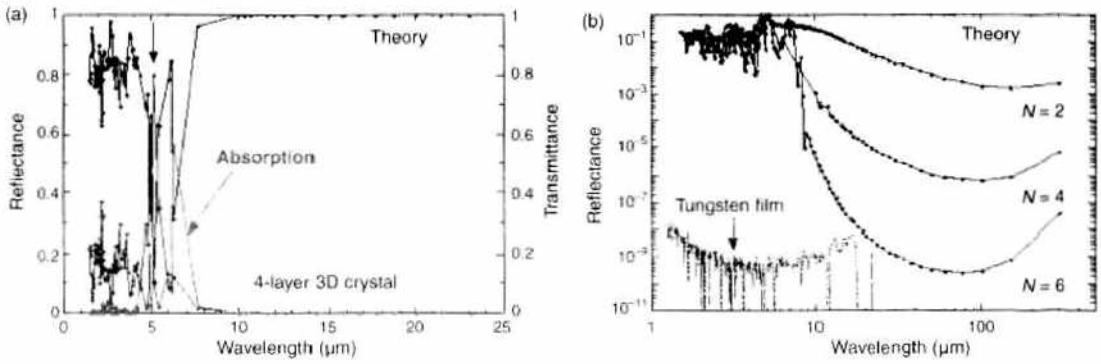


Figure 17. Reflectance and transmission/absorption spectra of three-dimensional tungsten photonic crystal calculated by the transfer matrix method. (a) Calculated reflectance (black), transmittance (blue), and absorptance (red) spectra for a four-layer three-dimensional tungsten photonic crystal. (b) Calculated transmission spectra for three-dimensional tungsten photonic crystals with different numbers of layers: $N = 2$, 4, and 6. Reprinted with permission from [71], J. G. Fleming et al., *Nature* 417, 52 (2002). © 2002, Nature Publishing Group.

confirm that the observed optical properties are the result of the PC structure, and not the absorption by metal, the transmission spectrum is calculated for various number of layers, as shown in Fig. 17b. The transmission spectrum for a tungsten thin film, which is also shown in Fig. 17b for comparison, exhibited very low transmission at $\lambda < 6$ μm, consistent with the small skin depth (300–500 Å) in the infrared region.

In contrast, the tungsten PC showed much higher transmission at the same spectral region, indicating the transmission is dominantly determined by the photonic band effect. In addition, the attenuation observed at $\lambda > 8$ μm scaled with the number of layers and not with the skip depth of tungsten. This also indicates that the attenuation in this region is the result of the photonic band effect. These theoretical calculations were confirmed by Fourier-transform infrared (FTIR) spectroscopy measurements, and the reflectance measured for various tilt angles also suggested the presence of complete PBG at $\lambda > 8$ μm. This strongly modified transmission and absorption can lead to the development of a unique lighting device in which the blackbody radiation in the infrared region is suppressed by the PBG, thereby recycling energy into the visible spectrum. The photonic band enhanced absorption band at the band edge can then act as an emission channel. It can also be used for a thermal-photovoltaic device, in which case the efficiency was predicted to be more than 4 times higher than the blackbody emitter.

4.2. Beyond the Photonic Band Gap

Identifying conditions for a large PBG is important because the PC can then support linear waveguides or cavities produced by intentionally introducing linear or point defects, as will be discussed later. However, it is increasingly being recognized that the optical properties of a PC outside the band gap could be just as intriguing and technologically important as the band gap itself. An immediate benefit of a device that does not require a PBG is the significant relaxation of material requirements, which would make fabrication of PC structures much easier. Among the interesting properties of the photonic band structures of PCs are the strong anisotropy and nonlinearity that some structures may exhibit. These properties can subsequently lead to nonclassical phenomena such as giant refraction, superprism, and negative refraction.

This was first recognized by Lin et al., who designed and experimentally demonstrated a PBG prism in the millimeter wave region [12]. A two-dimensional PC structure exhibits, in general, a mostly linear dispersion curve at low frequencies but strongly nonlinear dispersion near the PBG, where the bands become flat. At low frequencies corresponding to the lowest photonic band, the dispersion surface remains fairly isotropic, and thus the PC may be treated as a homogeneous dielectric material with an effective refractive index. In this case, the slope of the dispersion curve is inversely proportional to the effective refractive index. Thus, the effective index is expected to increase near the band gap. By measuring the deviation angle of millimeter wave by a triangular PC of alumina rods, Lin et al. [12] determined the effective index of the PC as a function of frequency. The dependence of the deviation angle on the incident angle was found to be described well by an effective index. Also, as shown in Fig. 18, the effective index was found to increase by up to ~20% as the millimeter-wave frequency approached the band edge, which was consistent with the band structure. If the same structure is made for visible light, a strongly dispersive prism can be made in a very compact size, ~20 μm, which can be integrated with other optical components such as ridge waveguides, forming highly integrated photonic devices.

At higher frequencies, the photonic bands can become highly anisotropic, causing a strong angular dispersion. This was first demonstrated by Kosaka et al., who used a three-dimensional PC structure to enable an angular dispersion ~500 times greater than in the conventional prism [14]. The PC structure used in this study consists of alternating layers of amorphous Si and SiO_2 deposited by self-replicating sputtering technique. The substrate was prepatterned with a triangular array of holes so that the subsequent deposition creates alternating layers of triangular lattice of Si and SiO_2.

The resulting three-dimensional structure shown in Fig. 19 is then a simple hexagonal graphite structure belonging to the space group P6/mmm. Because the structure is uniaxial, the allowed modes may be classified as TE-like and TM-like modes. Figure 20 shows the

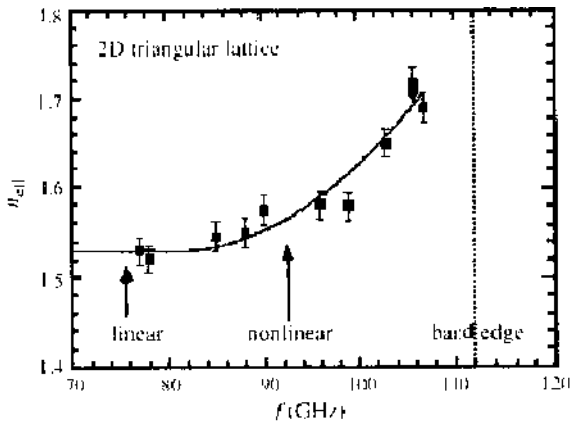


Figure 18. Measured effective refractive index as a function of frequency for a two-dimensional (2D) triangular lattice structure with lattice constant $a_0 = 0.81$ mm and rod diameter $d = 0.305$ mm. As expected, effective index increases rapidly as the frequency approaches the band edge of the fundamental TM band gap. Reprinted with permission from [12], S.-Y. Lin et al., *Opt. Lett.* 21, 1771 (1996). © 1996, Optical Society of America.

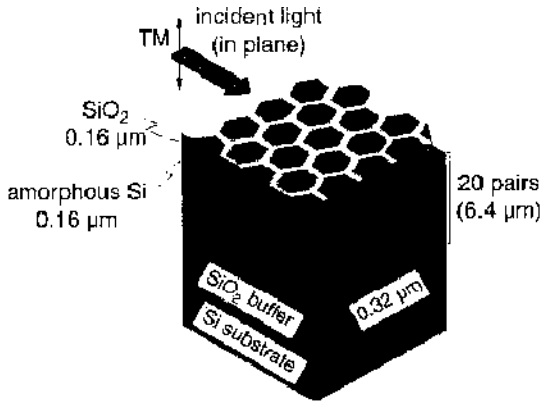


Figure 19. Schematic diagram of the Si/SiO_2 three-dimensional PC structure prepared by a self-replicating sputtering technique. Reprinted with permission from [14], H. Kosaka et al., *Appl. Phys. Lett.* 74, 1370 (1999). © 1999, American Institute of Physics.

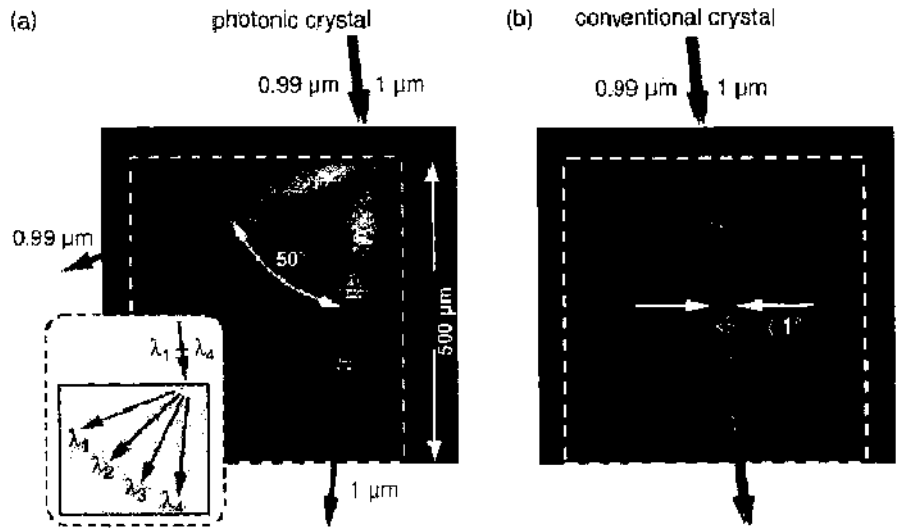


Figure 20. Photographs showing light paths inside (a) a photonic crystal and (b) a silicon wafer for incident lights with two different wavelengths, 0.99 and 1.0 μm. The TM polarized laser lights were incident at a tilting angle of 15° onto the edge measured from normal to the edge. Reprinted with permission from [14], H. Kosaka et al., *Appl. Phys. Lett.* 74, 1370 (1999). © 1999, American Institute of Physics.

propagation of a TM-polarized laser beam directed into the crystal with an incident angle of 15°. Two beams whose wavelengths differ only by 1% were found to propagate along two completely different directions. The angle separation was 50°, which is in sharp contrast with the less than 1° separation expected in a homogeneous dielectric material. The vastly increased angular dispersion is the result of the strongly anisotropic and frequency-dependent dispersion surface that causes the group velocity, and particularly its direction, to be strongly dependent on incident angle and frequency. The photonic band structure calculated by the PW method showed that the third photonic band possesses strong anisotropy.

As shown in Fig. 21a, the third photonic band exhibits very different slopes for the Γ–X and Γ–M directions. As a result, the constant frequency surface is strongly deformed from the circular shape of a homogeneous material and becomes a starlike shape, as shown in Fig. 21b. Recalling that the group velocity is defined by the gradient of dispersion surface in **k** space,

$$\mathbf{v}_g = \nabla_k \omega(\mathbf{k}) \tag{52}$$

the group velocity points to the direction normal to the constant frequency dispersion curve. Therefore, the large curvature near the tips of the dispersion curve shown in Fig. 21b means that a small change in incident angle will result in a very large variation in the group velocity direction. Indeed, a laser beam directed into the crystal with an incident angle of 7° was observed to be deviated with a refraction angle of 70°. The observed refraction angle was consistent with the group velocity direction predicted from the dispersion diagram.

The two structures discussed above are either pure two-dimensional crystal with long rod structures or a complicated three-dimensional crystal, both of which are rather difficult to fabricate in the nanoscale required for operations at optical frequencies. In this respect, two-dimensional PC slabs previously discussed in Section 4.1.2 represent more realistic structures that can be fabricated by the conventional thin-film deposition and lithography techniques commonly used in the electronics industry. Recently, Park and Summers investigated the refraction and dispersion properties of two-dimensional slab PCs and showed that the novel properties that have been predicted for two-dimensional and three-dimensional PCs are also realizable in slab PCs [72]. In this study, FDTD method was used to analyze a thin slab of Si ($\varepsilon = 12$) patterned with a triangular array of air ($\varepsilon = 1$) holes. The computation cell consisted of three layers: a dielectric slab with an air hole, an air layer, and an absorbing boundary layer. Only the upper half of the dielectric slab was defined, with a mirror boundary condition imposed at the computation cell boundary.

This system simulates a symmetric slab surrounded by air on both sides of the slab, in which case all modes are either even or odd under the reflection in the symmetry plane.

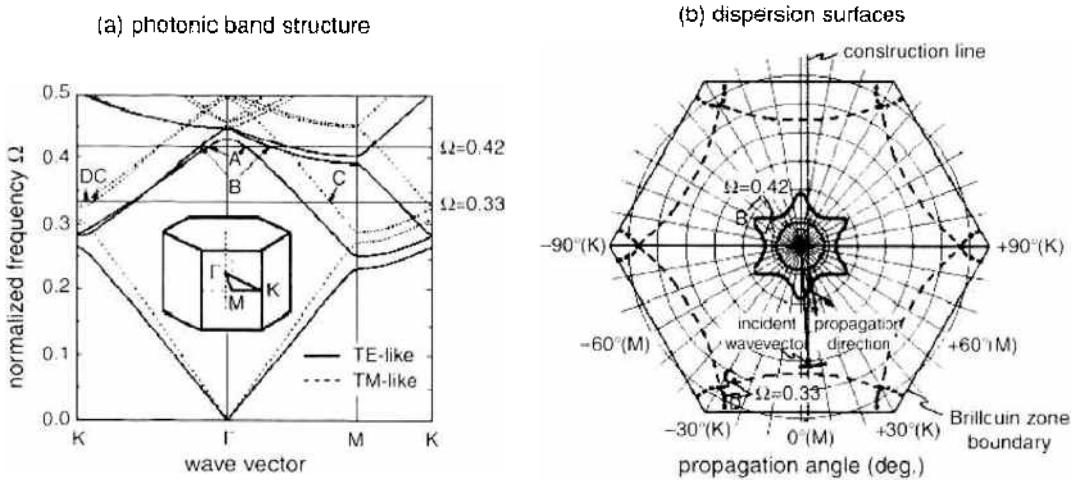


Figure 21. (a) Photonic band structure and (b) the constant-frequency dispersion surfaces calculated by the plane wave method. The solid and dashed curves denote the TE and TM modes, respectively. Reprinted with permission from [13], H. Kosaka et al., *Phys. Rev. B* 58, 10096 (1998). © 1998, American Physical Society.

By imposing the mirror boundary condition, the computation load was reduced by half, and it is also possible to select only even or odd modes. Within the plane of the slab, the periodic boundary condition according to Bloch's theorem is used to simulate the triangular lattice. The thickness of the air layer was varied between $2a$ and $4a$, where a is the lattice constant; that is, the distance between two adjacent air holes. The absorbing boundary layer was used to prevent unphysical reflections from the computation cell boundary and to simulate infinitely large air layers. For this purpose, the Berenger-type PML boundary condition discussed in Section 3.2 was used.

More specifically, the authors used an implementation of PML similar to that originally developed by Zhao and Cangellaris [73] but modified to work properly with the nonorthogonal coordinate system used for the triangular periodicity in the slab. The photonic band structure for a Si slab PC with an air-hole diameter of $0.4a$ and a slab thickness of $0.5a$ was found to exhibit two band gaps between normalized frequencies of 0.25–0.27 and 0.37–0.39 for the even modes. The band gap here refers to the region in which no guided modes exist, in accordance with the previous discussion in Section 4.1.2. Despite the drastic changes in the band structure resulting from the guiding condition, the slab PC proved to exhibit anisotropic dispersion surfaces similar to what was found in a pure two-dimensional PC.

The inset of Fig. 22a shows the dispersion diagram calculated at a normalized frequency of 0.357. The circle is the cross-section of the light cone, inside which no guided modes may exist. Although partially cut away by the light cone, a significant portion of the dispersion diagram, which is highly anisotropic and exhibits sharp inflection points along the Γ–M directions, still lies within the guiding regime. Because the group velocity is defined by the gradient of the dispersion surface, the light propagation direction may be predicted from the shape of the dispersion diagram. As shown in Fig. 22a, the refraction angle calculated from the group velocity direction changed from 0° to over 70° as the incident angle, measured in reference to the Γ–M direction, was varied from 0° to 7°. It should be pointed out that the cutoff is reached at an incident angle of 11°, as indicated in Fig. 22a, and thus the entire variation of 70° in the refraction angle is achieved within the guiding regime. The giant refraction properties were also strongly frequency dependent, leading to an extraordinary dispersion (superprism) effect.

As shown in Fig. 22a, for a range of normalized frequencies between 0.357 and 0.364, the refraction angle rapidly decreases as the frequency is increased. For a fixed incident angle of 2°, the refraction angle increased from 35° to 56° as the frequency was decreased

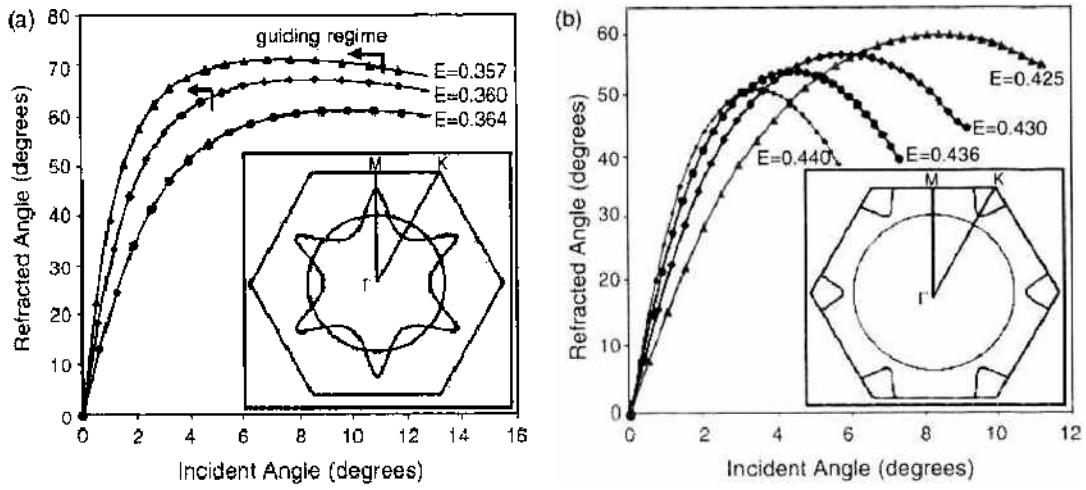


Figure 22. (a) Calculated refraction angles for normalized frequencies between 0.357 and 0.364. The incident angle is measured in the Γ–M direction. Inset: dispersion diagram calculated at a normalized frequency of 0.357. The circle represents the cutoff line, outside of which is the guiding regime. (b) Calculated refraction angles for normalized frequencies between 0.425 and 0.440. The incident angle is measured from the Γ–K direction. Inset: dispersion diagram calculated at a normalized frequency of 0.430. The circle represents the cutoff line, outside of which is the guiding regime. Reprinted with permission from [72], W. Park and C. J. Summers, *Opt. Lett.* 27, 1387 (2002). © 2002, Optical Society of America.

from 0.364 to 0.357, yielding over 21° separation in the refraction angle for a 2% change in frequency. This corresponds to approximately 8 nm in the visible spectrum and to about 0.03 μm near the communication wavelength of 1.54 μm. This strong dispersion effect is caused by the rapid contraction of the dispersion curve with the increasing frequency.

This process unfortunately moves a greater portion of the dispersion diagram inside the cutoff circle, where the dispersion surfaces become radiation modes. Therefore, as indicated in Fig. 22a, the cutoff occurs at a smaller incident angle for a higher normalized frequency, limiting the range of refraction angles attainable with guided modes. This problem, however, is not intrinsic to the structure. At a different frequency range, it was found that the same structure exhibits a similar giant refraction and superprism effect entirely within the guiding regime.

The inset of Fig. 22b shows the dispersion diagram calculated for a normalized frequency of 0.430. As shown, the dispersion curve is anisotropic and shows a sharp negative curvature along the Γ–K direction. Figure 22b shows the calculated refraction angle as a function of incident angle, which is now measured with respect to the Γ–K direction. At a normalized frequency of 0.430, for example, the refraction angles were found to change from 0° to nearly 60° as the incident angle was varied from 0° to 6°. Strong frequency dependence was also observed in this frequency region, as shown in Fig. 22b. At small incident angles, the refraction angle increased with increasing frequency, opposite to what was observed in Fig. 22a. For an incident angle of 2°, the refraction angle was increased from 29° to 44° as the normalized frequency was increased from 0.425 to 0.440, yielding a 15° separation for a 3% change in frequency. This phenomenon has a similar origin to the case in Fig. 22a, that is, the dispersion curves shifts rapidly as the frequency is changed. However, it should be emphasized that in this case the entire dispersion curve lies outside the cutoff circle at all frequencies, and thus the observed giant refraction and superprism effects are achieved with guided modes only.

A significant extension of the concept discussed in the preceding paragraph is incorporating electro-optic material and thereby achieving tunability. Incorporating liquid crystal (LC) in a PC structure to tune the photonic band structure was first proposed by Busch and John [74], which was soon followed by experimental demonstrations by Yoshino et al. [75] and Leonard et al. [76], who showed how the PBG could be tuned by using the temperature dependence of the refractive index of LC that was infiltrated into a silica opal and a two-dimensional anodic Si PC, respectively. Applying this scheme to the giant refraction in two-dimensional slab PC, Park and Summers performed three-dimensional FDTD simulations on LC-infiltrated Si slab PC structures and found that, with an LC index change of $\Delta n = 0.2$, the incident optical beams could be steered over ~5° [77]. This value was much smaller than what have been predicted by pure two-dimensional modeling studies carried out for lead lanthanum zirconate titanate incorporated PC structure in which an Δn of 0.12 and gigahertz switching speed could be achieved [78, 79].

When a high-performance LC with $\Delta n = 0.4$ was used, the range of steering angle could be increased to over 20° [80]. However, the effect of dynamic changes in refractive index can be dramatically increased by selectively biasing the optically active material so that the PC structure is transformed into a superlattice structure. This scheme significantly enhances the achievable tunability because in addition to the refractive index it also modifies the periodicity. The proposed superlattice two-dimensional PC structure is shown in Fig. 23. In addition to the refractive index modulation by the triangularly periodic holes patterned in a dielectric slab, an additional modulation is achieved by selectively addressing alternate rows of the LC-filled holes. This modulation creates a superlattice PC in which the additional periodicity superimposed by the selective biasing scheme modifies both the reciprocal lattice space and consequently the photonic band diagram, with profound effects on the optical behavior.

In the simplest implementation of this concept shown in Fig. 23, the top electrodes are linearly patterned to align along one side of the triangular lattice. In the reciprocal space, this direction corresponds to the Γ–K direction, as shown in Fig. 24. The bottom electrode, however, is unpatterned and covers the entire slab. The top electrodes are used to separately bias rows of holes along the Γ–K direction. When biased in an alternating fashion,

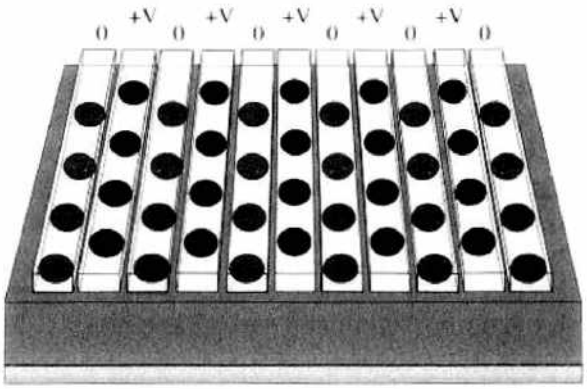


Figure 23. Schematic diagram of a superlattice photonic crystal structure produced by alternating biasing of liquid crystal–filled holes.

an additional periodicity in the refractive index arises along the direction perpendicular to the top electrodes (which corresponds to the Γ–M direction in the reciprocal lattice) and thereby creates a superlattice PC. The effect of the superlattice is to make some of the **k** vectors in the reciprocal space equivalent to one another, and thus to reduce the size of the first Brillouin zone, which is known as the Brillouin zone folding. For the superlattice structure shown in Fig. 23, the original hexagonal Brillouin zone of the triangular PC is folded into a rectangular Brillouin zone, as shown in Fig. 24. Because of the symmetry lowering induced by the superlattice, only four of the six M points in the original hexagonal Brillouin zone remain equivalent. The other two M points that are no longer equivalent to the rest are folded onto the Γ point in the new rectangular Brillouin zone. The K points are similarly affected, and only two (along the vertical direction in Fig. 24) remain as high-symmetry directions. The new high-symmetry points are labeled X and X', as shown in Fig. 24.

Figure 24 shows two dispersion curves calculated at a normalized frequency ($\omega a/2\pi c$) of 0.36, the first for a triangular PC in which all the holes are infiltrated with LC but are unbiased, and the second for the superlattice created by biasing every other electrode. The two dispersion curves exhibit distinctly different shapes as a consequence of the change in refractive index and the Brillouin zone folding. In interpreting this data, attention must be paid to the major difference between a two-dimensional PC slab and a pure two-dimensional or three-dimensional PC, which is the imposition of the guiding condition represented by the cutoff circle. Only the modes located outside the circle meet the guiding condition and

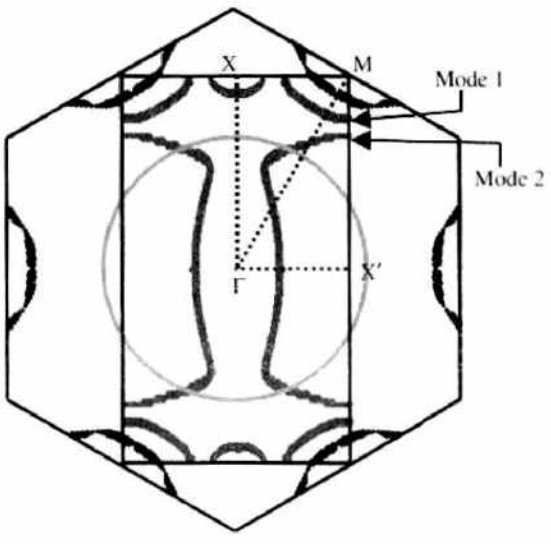


Figure 24. Dispersion diagrams for the liquid crystal–infiltrated two-dimensional slab photonic crystal calculated at a normalized frequency of 0.36. Blue curves are for the triangular PC, with all liquid crystals unbiased. Red curves correspond to the superlattice created by biasing every other electrode.

are confined or guided within the slab, whereas those inside the circle radiate into the surrounding media. Examination of the dispersion curves shows that the optical properties become critically dependent on the direction of light propagation direction.

For light incident along the Γ–X' direction (the direction perpendicular to the top electrodes shown in Fig. 23), the dispersion curves for the triangular PC and superlattice do not exhibit a significant difference in their curvatures, and thus the refraction properties remain similar to the case of uniform biasing. However, there is a major difference, in that the Brillouin zone folding transfers the originally guided modes of a triangular PC into the radiation regime in the superlattice. Thus, along the Γ–X' direction, one achieves switching between guided and radiation modes with modest changes in refraction angles. In contrast, along the Γ–X direction, which corresponds to light propagating parallel to the top electrode, the triangular PC with all electrodes unbiased exhibits a stop band and thus does not support any modes. However, when the electrodes are switched on in an alternating manner to create a superlattice PC, the Brillouin zone folding brings in allowed modes along the Γ–X direction, as shown in Fig. 24. Thus, along the Γ–X direction, one may switch between two states in which light is allowed to propagate with modest refraction or in which all of the incident light is back reflected.

A more complicated behavior is expected along the Γ–M directions, which make an angle of 30° with the top electrodes. As shown in Fig. 24, the superlattice possesses three allowed modes along the Γ–M direction, but only two are guided modes lying outside the cutoff circle. The outermost mode (mode 1) exhibits a curvature similar to that of the triangular PC. Thus, for this branch, we expect only modest changes in the refraction angle.

Figure 25 shows the refraction angles calculated by numerically evaluating the curvatures. As shown, the outermost branch exhibits a refraction angle of ~10° for very small incident angles ($\theta_i < 2°$) and reaches a refraction angle of almost 50° for $\theta_i = 12°$, beyond which the mode ceases to exist. The other guided mode (mode 2) exhibits a fundamentally different curvature. At small incident angles, the refraction angle has its maximum of 47° and decreases with increasing incident angle. This behavior is completely opposite to the triangular PC, which exhibits small refraction effects at low incident angles. Therefore, by preferentially coupling to this branch, light incident along the Γ–M direction may be electronically, or optically, scanned over very large angles up to 47°. Furthermore, for incident angles greater than 12°, the outermost branch no longer exists, resulting in a single-mode regime in which the superlattice exhibits refraction angles smaller than the triangular PC. In this single-mode

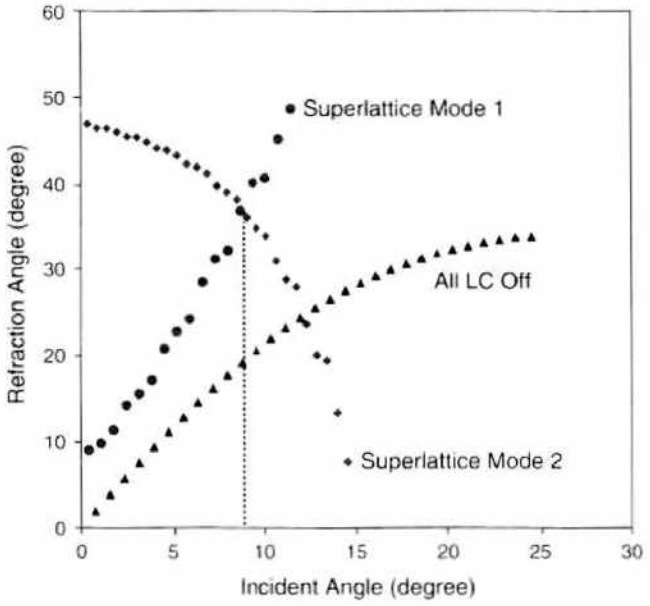


Figure 25. Calculated refraction angles for the triangular photonic crystal with all liquid crystals (LCs) unbiased and for the superlattice structure created by biasing every other electrode. The incident angle is measured from the Γ–M direction.

operation, the largest achievable difference in refraction angle was approximately 20° at an incident angle of 15°.

As the anomalous light propagation in PC structures receives increasingly strong attention, Notomi recently provided the theoretical framework within which the unconventional light propagation in PC structures is to be interpreted [81]. According to this theory, the light propagation in a weakly modulated PC is described in a similar way to the conventional diffraction grating. In the limit of vanishingly small index modulation, the dispersion curve at a constant frequency consists of many circles displaced by various reciprocal lattice vectors, as shown in Fig. 26a. Unlike the diffraction grating in which the periodic modulation is typically one-dimensional, however, the reciprocal lattice vectors in a PC structure have two-dimensional or three-dimensional periodicity.

This model predicts that the light could be decomposed into more than one beam in the PC; one corresponding to the refracted wave and the other corresponding to the diffracted wave, as indicated by arrows A and B in Fig. 26a. The propagation direction of each beam is normal to the curve, as determined by Eq. (52). An interesting behavior is then predicted near the intersections of the dispersion curves, where a large variation of propagation direction is expected even for a slight change in incident angle. This giant refraction and superprism effect has been discussed in the previous paragraphs.

As the strength of index modulation is increased, however, an additional effect arises at the intersections. The different modes approaching together at the intersection begin to repel each other and create discontinuities. As a result, the dispersion curve at a constant frequency now consists of a set of disjointed starlike curves, as shown in Fig. 26b. As the modulation strength remains weak, the overall shape of the curves is much the same as the vanishingly small modulation case. Because of the gaps near the intersections, however, some of the modes will not be excited. Figure 26b shows the case in which the parallel component of the incident **k**-vector falls in one of the gaps. As a result, the refracted wave becomes evanescent, and the only propagating wave is the diffracted wave, as indicated in Fig. 26c.

This situation may seem as if the optical beam is refracted in an unconventional way, but it merely represents the special condition under which the refracted beam is suppressed. The overall picture still remains the same, and the mode inside the PC is characterized by either a diffracted wave corresponding to a reciprocal lattice vector **G**, or a refracted wave

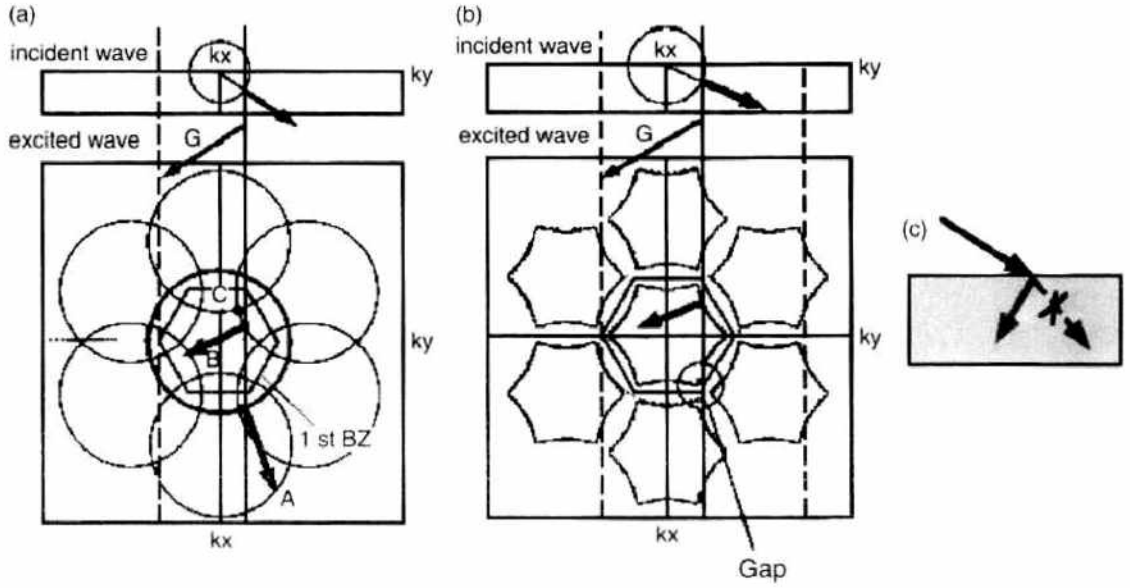


Figure 26. Dispersion curves at a constant frequency (a) for a hexagonal two-dimensional photonic crystal with a vanishingly small index modulation and (b) for a hexagonal two-dimensional photonic crystal with finite index modulation. The first Brillouin zone is shown as a hexagon. (c) Schematic of anomalous diffraction near the singular point. Reprinted with permission from [81], M. Notomi, *Phys. Rev. B* 62, 10696 (2000). © 2000, American Physical Society.

(corresponding to $\mathbf{G} = 0$). Furthermore, because this behavior occurs only for the limited special cases in which the tangential component of incident $\mathbf{k}$ vector falls in the gap between dispersion curves, one cannot define a meaningful effective refractive index that describes the light propagation for all possible incident angles.

This is the situation for PC structures with low refractive index contrast or for frequencies far from PBG. If, however, the index modulation is large, the constant-frequency dispersion curve is now dominated by the effect of mode mixing and the resultant gap openings. This consequently leads to the breakdown of the grating picture used for the case of weak modulation. The mode inside the PC can no longer be characterized by a single dominant reciprocal lattice vector but by a linear combination of many Bloch waves corresponding to various different reciprocal lattice vectors. As a consequence, the dispersion curve at a constant frequency strongly deviates from a simple ensemble of circles.

Near the PBG, the dispersion curves are found to approach a circular shape, as shown in Fig. 27a. This is a general phenomenon also observed in Fermi surfaces of electrons in crystals. Therefore, at frequencies near the PBG, the refraction angles determined by Eq. (52) could be fitted well by Snell's law, from which an effective refractive index may be deduced. Figure 27b shows the refraction angles calculated by Eq. (52), and Fig. 27c shows the effective refractive index obtained by fitting the refraction angle curves. It was found that the effective index is not sensitive to the incident $\mathbf{k}$ vector for frequencies between 0.59 and 0.645. This defines the range of frequencies in which the effective index is well defined. Furthermore, we notice that the effective refractive index is negative for frequencies below the band gap, which is centered around 0.635 in the example shown in Fig. 27. This means that the light is refracted at negative angles for all incident angles.

The difference is clear from the previously discussed case of weak modulation, in which negative-refraction-like behavior is observed for a limited range of incident angles. The negative refraction provides ample opportunities to observe novel phenomena. An example is readily seen in Fig. 27b, which shows a cutoff in refraction at an incident angle of $\sim 30^\circ$ for $\omega = 0.61$. In other words, no refraction occurs at incident angles larger than this cutoff angle given by $\sin^{-1} |n_{\text{eff}}/n_o|$, which represents the onset of total internal refraction. Clearly, this is not possible in the conventional dielectric system, in which the light is incident in the lower-index air to an interface with a dielectric material. Also, recognizing that the effective index is related to the curvature of the band, Notomi pointed out an interesting analogy between the electronic and photonic systems in periodic structures. It is well known that electrons near the band edges in a crystal behave in much the same way as the free electrons and their motion is characterized by their effective masses which are determined from the curvature of the energy bands. The conduction band electrons possess positive effective mass, whereas

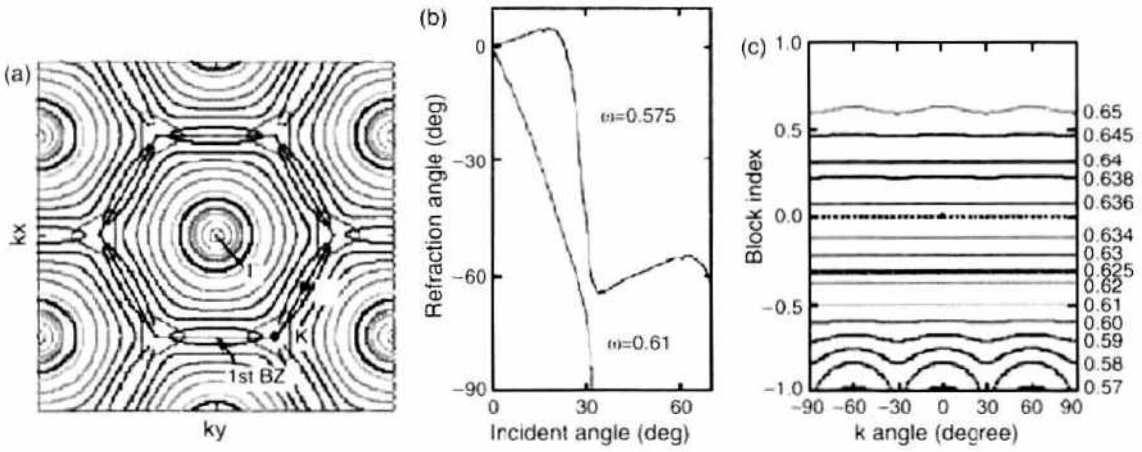


Figure 27. (a) Dispersion curves for TE modes in a two-dimensional GaAs pillar hexagonal photonic crystal ($n_1 = 3.6, n_2 = 1, 2r = 0.7a$) at $\omega = 0.56$ to 0.635 (from outer to inner). The colors represent frequency, as indicated in (c). (b) Refraction angle versus incident angle at $\omega = 0.575$ and 0.61. (c) Effective refractive index as a function of the angle of the $\mathbf{k}$ vector at various frequencies. Reprinted with permission from [81], M. Notomi, *Phys. Rev. B* 62, 10696 (2000). © 2000, American Physical Society.

the valence band represents negative effective mass states. Just like this electronic system, the photonic states described in Fig. 27 represents two bands above and below the PBG that are characterized by positive and negative effective refractive indices, respectively.

Luo et al. later discovered that negative refraction does not necessarily require negative effective index but is also possible in a system exhibiting positive effective index [82]. By calculating the constant-frequency dispersion curves for a two-dimensional square lattice of air holes in a Si slab, the authors observed that the curvature (photonic effective mass) of the first photonic band becomes negative near the M point of the Brillouin zone. This subsequently results in negative refraction, as shown in Fig. 28a. For this negative refraction to occur at all incident angles, the dispersion curve around the M point must be larger than the circle representing the dispersion curve in air. Also, to ensure that there is no mode branching inside the PC, the frequency must be restricted to $\omega < \frac{1}{2}\frac{2\pi c}{a_x}$, where a_x is the periodicity along the interface. Using these criteria, the authors discovered that all-angle negative refraction occurs at frequencies between 0.186 and 0.198 for the Si–air two-dimensional structure. This all-angle negative refraction was then used to demonstrate the strong focusing effect anticipated from a slab of negatively refracting material. Figure 28b shows the result of FDTD simulation for a continuous-wave source placed $0.35a$ apart from the surface. A point image with a transverse size of 0.67λ appears on the opposite side of the thin slab of PC. In this calculation, the slab thickness and symmetry were adjusted to achieve resonant maximum in transmission. Also, to focus a distant object, the PC slab should be thick enough that the rays negatively refracted at the front surface crosses one another within the slab. This ensures that the subsequent negative refraction at the back surface brings the light back into focus.

5. PHOTONIC DEVICES BASED ON PHOTONIC CRYSTAL

5.1. Photonic Crystal Waveguides

The presence of PBG presents exciting possibilities of creating active optical elements in which light is strongly confined by the surrounding PBG matrix. When a line defect is introduced in an otherwise perfect PC structure exhibiting a PBG in the frequency region of interest, light is forced to stay within the defect structure, which therefore acts as a waveguide. In the conventional dielectric waveguides, the light confinement relies on the index difference between the guide and the surrounding materials. The confinement is therefore

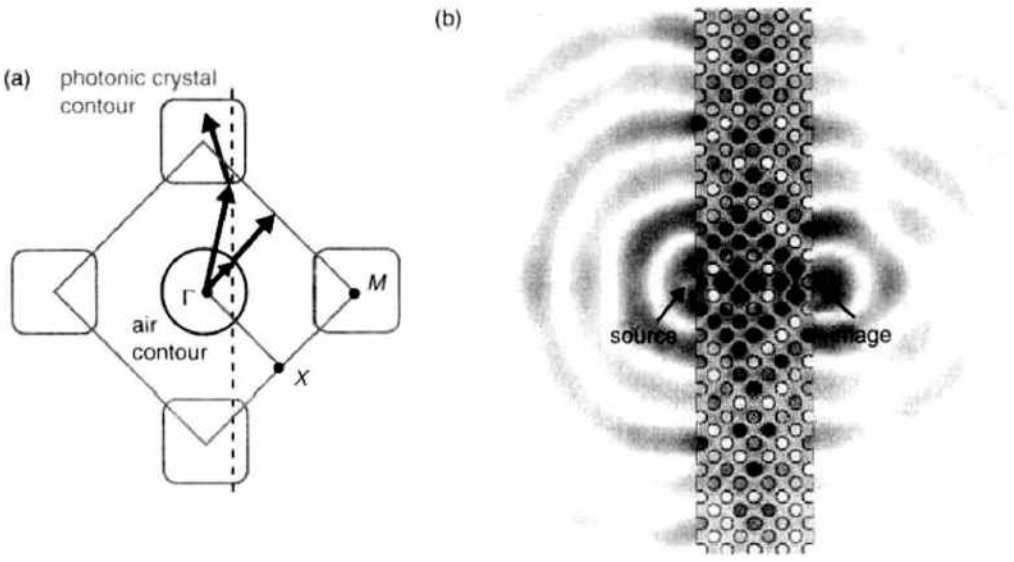


Figure 28. (a) Schematic of constant-frequency dispersion curve exhibiting all-angle negative refraction. (b) H_z field of a point source and its image across a photonic crystal slab (yellow), exhibiting all-angle negative refraction. Blue, white, and red correspond, respectively, to negative, zero, and positive H_z. Reprinted with permission from [82], C. Luo et al., *Phys. Rev. B* 65, 201104(R) (2002). © 2002, American Physical Society.

weak, and a large loss is expected whenever a nonregularity is encountered. For this reason, to steer light around a corner, the radius of curvature must be made much larger than the light wavelength to minimize radiation loss, which consequently limits the miniaturization of the photonic device containing such waveguides. In a waveguide constructed inside a PC structure, however, the guidance is provided by the PBG, which can lead to low loss guidance around a very sharp corner. This unique features of the PBG-based waveguides make the PC an ideal platform for highly integrated photonic devices that may be considered the photonic equivalent of very large scale integrated circuits in electronics.

In a two-dimensional PC, a linear waveguide is usually produced by modifying the dielectric constant of a row of cylinders or holes that comprise the PC structure. This type of linear waveguide is usually modeled by constructing a supercell containing the modified element in the center. Along the waveguide direction, the supercell contains only one unit cell and is terminated by the periodic boundary condition as the structure still remains periodic in that direction. Along the direction perpendicular to the waveguide, however, the supercell must contain many unit cells so that it would properly simulate the surrounding PC matrix of infinite extent. Because the guided modes decrease exponentially as one moves away from the waveguide, either periodic or absorbing boundary conditions may be used for the computational cell boundary in this direction. When the periodic boundary condition is used, the numerical results are actually for a system of parallel waveguides, but as long as they are far apart, the guided modes in separate waveguides would have negligible coupling and therefore provide proper simulation of the single-waveguide structure.

Mekis et al. used the PW method to calculate the guided modes in a linear waveguide created in a square lattice of GaAs rods [83]. This structure exhibits a TM band gap, which reaches a maximum of $\Delta\omega/\omega = 38\%$ when the rod radius is $r = 0.18a$. When a row of rods along the (10) direction is removed, new allowed modes are created within the band gap.

In Fig. 29a, the result of supercell calculations for various k values along the waveguide direction is overlaid on the band structure of the perfect PC. The shaded region indicates the presence of extended modes that are the eigenstates of the perfect crystal. The supercell calculation showed that the waveguide supports one mode at most of the k values along the waveguide direction but ceases to support guided modes for wave vectors close to the Brillouin zone boundary. The electric field profile indicates that this mode has even symmetry with respect to the mirror plane along the middle of the waveguide. Similar to the conventional dielectric waveguide, the field was found to decay exponentially outside the waveguide while propagating along the guide.

Figure 29b shows the dispersion curves for a waveguide made by removing three rows of GaAs rods along the (11) direction. In this case, three guided modes were found to exist. The mode with the lowest frequency was again an even mode and was followed by odd and even modes at higher frequencies. It is easily understandable that a wide waveguide supports more modes because it provides more room for the fields to develop nodes within the

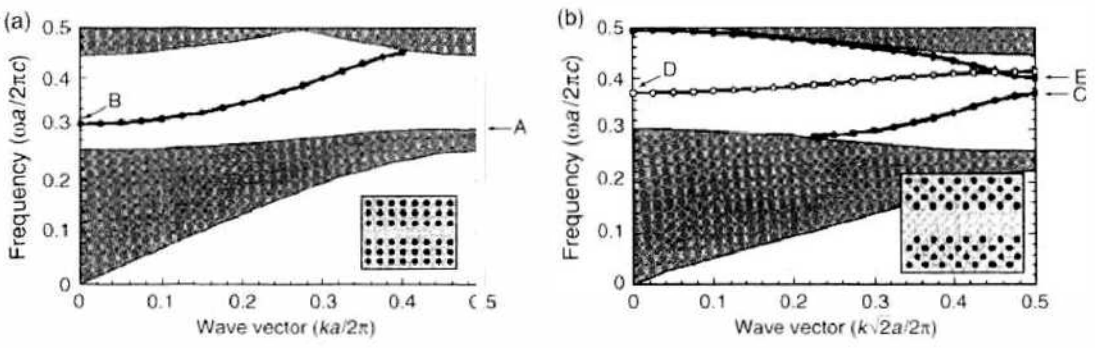


Figure 29. Dispersion relations for the two PC waveguides. The geometry of the waveguides is shown in the insets. The gray areas are the projected band structure of the perfect crystal. The filled circles correspond to even modes, and the open circles correspond to odd modes. The frequencies at the points indicated are (in units of $2\pi c/a$) A = 0.302, B = 0.312, C = 0.371, D = 0.373, and E = 0.400. Reprinted with permission from [83], A. Mekis et al., *Phys. Rev. B* 58, 4809 (1998). © 1998, American Physical Society.

waveguide, creating higher-frequency modes. The strong guiding mechanism provided by the PBG forces the light confined inside the waveguide even in the presence of a sharp bend. Mekis et al. used the FDTD method to calculate the transmission around a 90° bend introduced in a linear waveguide in a square lattice of GaAs rods [3]. The waveguide direction corresponds to (10) direction and thus forms a single mode guide as shown in Fig. 29a.

The condition for single-mode operation is important because irregularities such as sharp bends can cause mode mixing in a multimode waveguide. Input light pulses were created from a point dipole source placed at the entrance of the waveguide, and a Gaussian envelop in time was imposed to produce a Gaussian frequency response, as shown in Fig. 30a. The transmission coefficients remain close to 100% at frequencies above the cutoff. At high frequencies, reflection from the bend becomes larger, reaching ~10%. This reflection could be mitigated by slightly modifying the shape of the bend. The field profile shown in Fig. 30b clearly indicates that the fields remain completely confined within the guide despite the existence of a sharp bend. It should also be noted that the light is guided in the region of lower effective index, which is not possible in conventional dielectric waveguides. This phenomenon is uniquely observed in PC-based structures only.

As mentioned in Section 4.1.2, from the fabrication standpoint, the two-dimensional slab PC with a finite thickness is more desirable than the pure two-dimensional PC, which requires very long structures with high aspect ratios. Because the two-dimensional slab PCs can exhibit a photonic band gap for guided modes, it should be possible to introduce a line defect and create localized modes that are confined by the photonic band structure in the plane of the slab and by the index difference in the direction perpendicular to the slab. Johnson et al. used the PW method to calculate the waveguide modes in various types of line defects in two-dimensional slab PCs [84]. Figure 31 shows the dispersion curves for a linear waveguide formed by a row of rods with reduced radius in a two-dimensional square array of rods with a radius and height of $0.2a$ and $2.0a$, respectively. The line defect of reduced rods is formed along the (10) direction, as shown in the inset of Fig. 31, and therefore the dispersion curve was plotted along the corresponding Γ–X direction in the Brillouin zone. The two shaded regions in Fig. 31 represent the light cone, where the light can propagate out of the slab into the surrounding medium, and the slab bands, for which the light can freely propagate into the bulk of the two-dimensional slab PC. The calculations for various

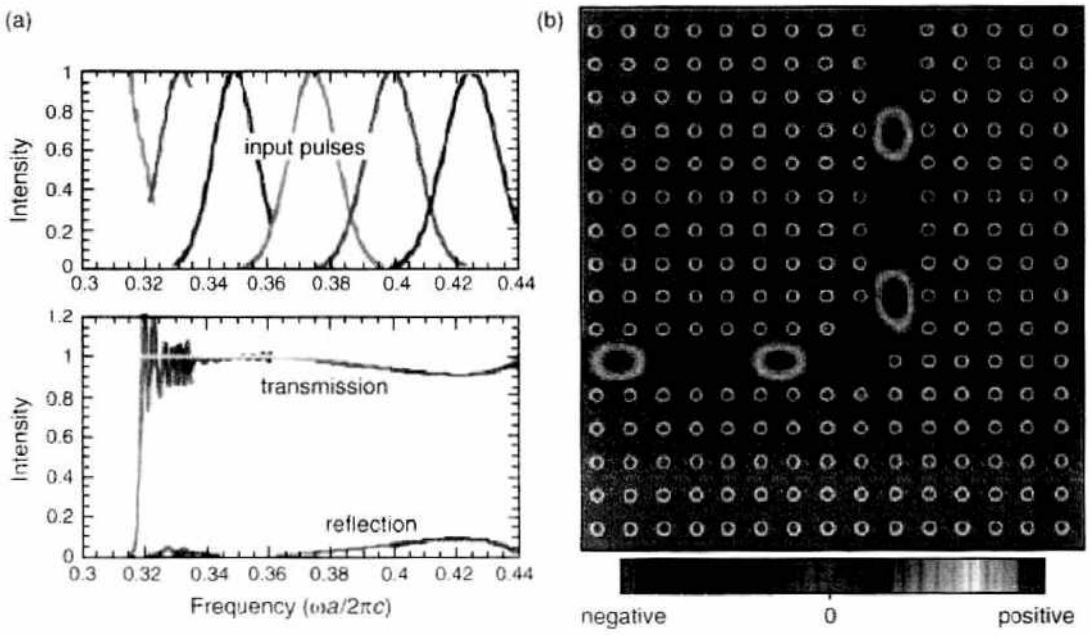


Figure 30. (a) Spectral profile of six input pulses and computed transmission and reflection coefficients for each input pulse; (b) electric field pattern in the vicinity of the bend for frequency $\omega = 0.353 \times 2\pi c/a$. The electric field is polarized along the axis of the dielectric columns. Reprinted with permission from [3], A. Mekis et al., *Phys. Rev. Lett.* 77, 3787 (1996). © 1996, American Physical Society.

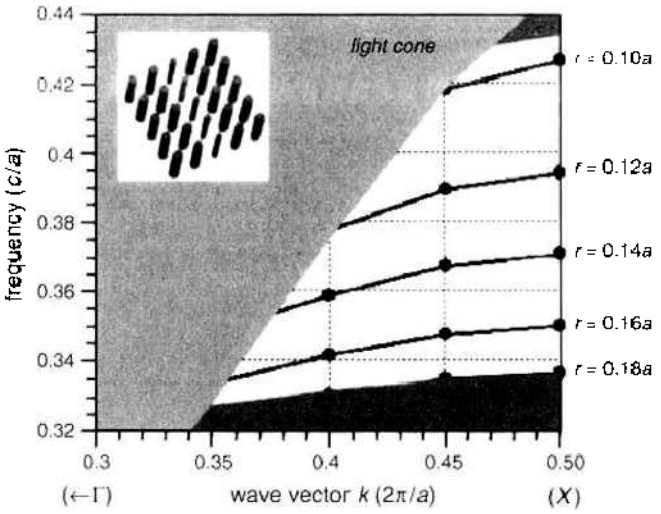


Figure 31. Projected band structure for the reduced-index-rod slab waveguide showing the odd-symmetry-guided bands for various defect rod radii. The bulk radius was set to be $0.20a$. Reprinted with permission from [84], S. G. Johnson et al., *Phys. Rev. B* 62, 8212 (2000). © 2000, American Physical Society.

radii for the defect rods showed that, for radii between $0.10a$ and $0.18a$, the linear defect gives rise to a single mode that lies outside of both the light cone and the slab bands.

As a consequence, these modes cannot couple to the radiation modes leaking out of the slab nor to the slab modes propagating throughout the two-dimensional slab. The fields are therefore strongly confined within the line defect and decay exponentially away from it, forming linear waveguide modes. When the defect rod radius is decreased to $0.08a$, the guided mode disappears completely into the light cone. This contrasts with the pure two-dimensional case, in which the line defect of zero radius (i.e., the case in which the defect waveguide is formed by entirely removing a row of rods), still supports waveguide modes. This shows the restrictions that can be imposed by the additional condition of guidance within the slab that the modes in a two-dimensional slab PC must satisfy.

In general, it is difficult for a waveguide in a slab PC to support modes that exist mainly in air because of the difficulty in vertical confinement. Investigations for other types of waveguide structures led to similar conclusions that the two-dimensional slab PC waveguides do support waveguide modes but that the parameter values that support waveguide modes are restricted by the slab guiding condition. For the waveguide consisting of a solid dielectric strip that can be formed by, for example, removing a row of holes in a PC structure of periodic holes in a dielectric slab, new modes arise at frequencies below the continuum of slab bands. These modes sense the surrounding PC structure as a homogeneous medium with an average refractive index, which is lower than the index of the dielectric strip. Thus, in this case, the light is guided by the index difference in both the vertical and the lateral direction. The field profile indicates that the index guided modes are the fundamental modes with no nodes or with the lowest-order modes when there is more than one index-guided mode. However, the modes that fall in the PBG region and that are therefore guided by the photonic band structure tend to be higher-order exhibiting nodes in either the vertical or lateral directions.

Although the PW method using a large supercell provides accurate descriptions of guided modes, the FDTD technique incorporating absorbing boundary conditions is sometimes more desirable because it can correctly describe even the radiation modes. One still needs to construct a supercell, but the supercell boundaries are terminated with absorbing boundary conditions, except for the waveguide direction, for which periodic boundary condition is used. Depending on the symmetry of the given structure, computational cell size can be reduced by properly applying mirror boundary conditions.

Figure 32 shows the dispersion diagrams for a linear waveguide in a slab PC consisting of a square array of holes in a Si slab [85]. As shown in the inset, the waveguide is formed along the symmetry direction, corresponding to the Γ–X direction in the reciprocal lattice.

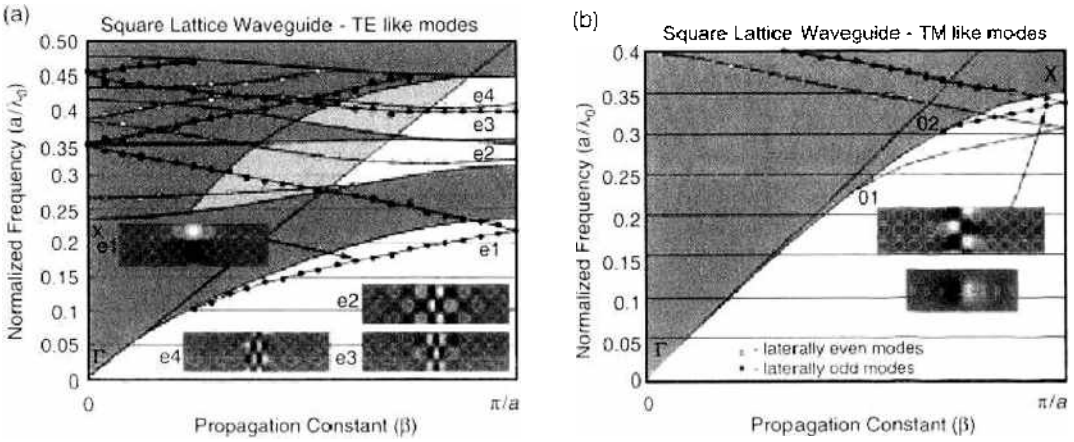


Figure 32. Dispersion diagram for (a) vertically even (TE-like) and (b) vertically odd (TM-like) guided modes in the waveguide made as a single-line defect in a square PC lattice. Solid line from (0, 0) to (π/a, 0.5) represents the light line. Insets show field patterns of (a) B_z and (b) E_z components in the middle of the slab (*z*-slice) for different guided modes. In the case of TM-like guided modes (o1 and o2), the distribution of the E_z component along the cross section of the waveguide is also shown. Reprinted with permission from [85], M. Loncar et al., *J. Lightwave Technol.* 18, 1402 (2000). © 2000, IEEE.

The index-guided modes lying below the continuum of slab bands are present for both polarizations. For the TM-like polarization, they are the only waveguide modes, as the photonic band structure does not exhibit a band gap. For TE-like polarization, however, a wide PBG opens and three waveguide modes arise within the PBG. The vertical field profiles indicate that all these modes are fundamental and exhibit no nodes.

The lateral field profiles show, however, that higher-frequency modes exhibit more nodes. The lateral profiles shown in the inset of Fig. 32 were calculated for a **k** vector close to the Brillouin zone boundary where the lowest PBG guided mode approaches the slab band. As a result, the field pattern for this mode extends deeper into the PC matrix. Ideally, a waveguide should support only a single mode, so that it would not exhibit mode mixing or cross-talk even in the presence of bends or interfaces. This could be achieved by properly modifying the waveguide structure, which consequently affects the waveguide mode frequencies and field profiles. Furthermore, this technique can also be used for minimization of radiation loss and efficient coupling with fiber or other integrated optical elements.

One of the simplest ways to change the waveguide mode frequency is to modify the width of the waveguide. Loncar et al. investigated the effect of modifying the width of a waveguide structure defined along the Γ–K direction in a triangular PC structure [86]. First, the waveguide width was reduced by translating the two surrounding PC structures along the direction normal to the waveguide axis (Γ–M direction), as shown in Fig. 33a and 33b. As the waveguide width is decreased, the mode frequencies begin to increase, pushing mode 1, for example, into the midgap, as shown in Fig. 34a. The mode located in the midgap is better confined by the photonic band structure and thus exhibits less loss through the coupling with the slab bands. This type of waveguide, however, cannot be used to create a sharp bend of the same width, as illustrated in Fig. 33b, because the original triangular symmetry is broken by the translation along the Γ–M direction.

A better scheme is to translate the two PCs along the nearest-neighbor direction (Γ–K), as shown in Fig. 33c and 33d. This way, the waveguide width can be adjusted freely and a 60° bend can be introduced without an abrupt change in waveguide width at the corner. The dispersion curves for this type of waveguide are shown in Fig. 34b. Similar to the previous case, the waveguide modes are pushed upward in frequency as the width is decreased. An advantage of this type of waveguide is that the band labeled mode 2 in Fig. 34b can be pushed up all the way into the air band. This mode has a more or less flat dispersion curve extending into the light cone and thus can couple efficiently with the radiation modes, leaking light out of the slab. By pushing this mode out of the PBG, it is possible to have a reduced radiation loss and more efficient in-coupling.

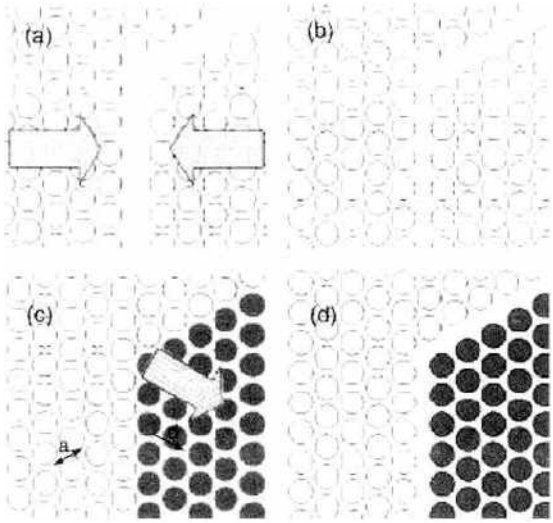


Figure 33. Different types of waveguides. By moving planar photonic crystal mirors that surround (a) a single-line defect waveguide along the direction indicated by arrows (Γ–X direction), we can form (b) a type 1 waveguide. (c) Unperturbed planar photonic crystal lattice. If we offset the gray holes along the Γ–K direction by d, we can form (d) a new type of waveguide (type 2). Reprinted with permission from [86], M. Loncar et al., *J. Opt. Soc. Am. B* 18, 1362 (2001). © 2001, Optical Society of America.

It is, however, difficult to create a 120° bend with this type of waveguide, and also the coupling with a nanocavity is not straightforward. It is noted that new modes, labeled as "accept. mode" in Fig. 34, appear in both types of reduced-width waveguide. These modes are acceptor-type modes pulled up from the slab band located below the band gap. This slab band has the fields concentrated in the dielectric region, and thus referred to as the dielectric band. When a waveguide is created by removing the dielectric material (i.e., by removing a row of dielectric rods in a two-dimensional array of rods) then the dielectric band is primarily perturbed and the waveguide modes arise from the dielectric band pulled up into the band gap. This type of waveguide mode is usually called the *acceptor-type mode*. The slab band lying above the band gap, however, exhibits a field profile concentrated in the air holes and thus called the air band.

If a waveguide is introduced by removing air holes, the air band is mainly disturbed and pulled down into the band gap, giving rise to the waveguide modes. This type of waveguide mode is called the *donor-type mode*. In this structure, the waveguide is created by removing

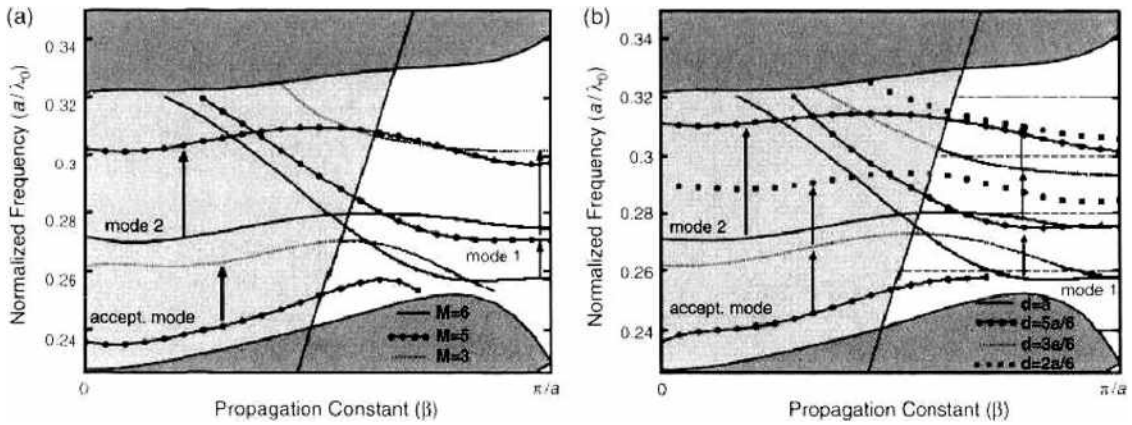


Figure 34. Dispersion relations for the guided modes of the type 1 and type 2 waveguides shown in Fig. 33 for different values of controlling parameters, M and d, respectively. (a) Type 1 waveguide. The width of the waveguide (center-to-center distance between two holes adjacent to the waveguide) is defined as $a\sqrt{3}/2(1 + M/6)$; $M = 6$ yields a single-line defect waveguide. (b) Type 2 waveguide. Position of the modes as a function of the parameter d, offset along the Γ–K direction. Reprinted with permission from [86], M. Loncar et al., *J. Opt. Soc. Am. B* 18, 1362 (2001). © 2001, Optical Society of America.

rows of air holes, and thus only donor-type modes are expected. In the case of the single-line defect waveguide created by simply removing a row of air holes, this is indeed the case, and we only see the donor-type modes. When the waveguide width is reduced, however, some regions that were originally occupied by the dielectric material in the perfect PC structure are now replaced by air holes, and other regions that were originally air holes are now dielectric. Therefore, both the dielectric and air bands are perturbed, creating both the donor- and acceptor-type modes.

One way to change the waveguide width without physically translating air holes is to modify the size of the air holes neighboring the waveguide. As the size of the air holes adjacent to the waveguide is increased, the effective width of the waveguide decreases, and thus the waveguide modes are shifted to higher frequencies. At the same time, acceptor modes appear because the enlarged air holes perturb the dielectric band. In general, this type of waveguide exhibits a complicated band structure, as three rows of air holes are modified.

Yet another scheme to control the waveguide modes is to change the size of a row of air holes instead of completely removing them. A row of reduced air holes forms a waveguide supporting donor-type modes. As the hole size is reduced, it acts as if the waveguide width is increased, and thus the waveguide modes are shifted to lower frequencies. In contract, when the defect air holes are larger than the regular air holes, the waveguide supports only acceptor-type modes. Increasing hole size once again pushes the mode frequencies upward. As shown in Fig. 35, this type of waveguide can produce a truly single-mode waveguide, and the guided mode can be made to exist near the midgap by adjusting the defect hole size. This would lead to strongly confined, low-loss propagation through the waveguide. It is pointed out once again that this type of waveguide is uniquely achieved in the PC only because the waveguide region has a lower effective index than the surrounding medium.

5.2. Photonic Crystal–Based Nanocavity

Spontaneous emission of an atom has long been believed to be an intrinsic property that cannot be altered by external perturbations. However, in the correct quantum electrodynamical description, spontaneous emission is an interaction between atom and vacuum. The apparent irreversibility of spontaneous emission is in fact caused by the infinitely many vacuum states available for the emitted photon. Therefore, the vacuum acts as a large reservoir in which the original atomic excitation decays. If, however, the vacuum states are significantly modified, it can result in strong suppression or enhancement of spontaneous emission. One way to strongly modify the photonic density of states is to use an optical cavity. The cavity design then allows tailoring the photonic density of states and controlling the interaction between atom and vacuum.

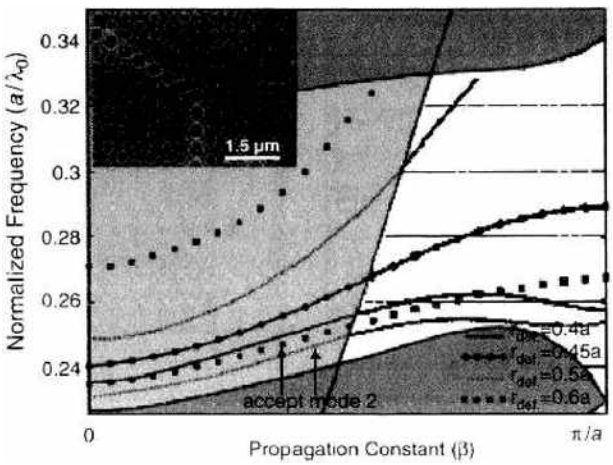


Figure 35. Dispersion relations for the modes guided in the waveguide shown in the inset as a function of the radius of the defect holes (r_{defect}). Reprinted with permission from [86]. M. Loncar et al., *J. Opt. Soc. Am. B* 18, 1362 (2001). © 2001. Optical Society of America.

This exciting possibility has naturally attracted much attention and led to the establishment of cavity quantum electrodynamics (CQED) [8]. The most highly publicized effect predicted by the CQED is the possibility of strongly enhancing the spontaneous emission rate. The probability Γ_s of spontaneous emission per unit time, widely known as the Einstein A coefficient, is proportional to the square of the Rabi frequency, the characteristic frequency that describes atom-field interaction, and also to the mode density $\rho_0(\omega)$, which represents the number of available photon modes at the emission frequency [8].

$$\Gamma_s = 2\pi\Omega_{ef}^2 \frac{\rho_0(\omega)}{3} = \frac{\omega^3}{3\pi\hbar c^3} \frac{|D_{ef}|^2}{\varepsilon_0} \tag{53}$$

where Ω_{ef} is the Rabi frequency, ρ_0 is the free-space photon density of states, and D_{ef} is the electric dipole matrix element between the two levels of the atom. The above equation describes the familiar exponential decay law for spontaneous emission. When the atom is placed in a cavity, however, the consequent modification of the photon density of states leads to large variations in spontaneous emission rates. A simple and well-known example is the planar cavity constructed by two parallel mirrors. This system is known to exhibit a cutoff at $\lambda = 2d$ (d is the spacing between the two mirrors), below which there exist no allowed TE modes (electric fields parallel to the mirrors). This subsequently leads to complete suppression of spontaneous emission resulting from a parallel dipole. This problem is illustrative of the spontaneous emission modification effect in a cavity and is also very useful as a reference because it has analytical solutions.

The spontaneous emission rate in this cavity may be found by using an infinite set of image charges [87]. As shown in Fig. 36, the emission rate by a parallel dipole is zero when $d/2\lambda < 1.0$ and shows sudden increases whenever $d/2\lambda$ = odd integers. These discontinuities occur because at those values of cavity spacing, new modes begin to propagate inside the cavity. New modes are added when $d/2\lambda$ = even integers, too. However, these modes do not have electric field components parallel to the mirrors at the center of the cavity where the dipole is located, and therefore do not affect the radiation rate of the parallel dipole. This is why we do not see any discontinuities at $d/2\lambda$ = even integers.

For a vertical dipole polarized perpendicular to the mirrors, there is no cutoff because it can always couple to the TEM mode, for which the electric field is uniform and normal to the mirrors. For small values of $d/2\lambda$, the radiation rate diverges as $1/d$ because of the divergence of the normalized field distribution. The discontinuities appear at $d/2\lambda$ = even integers at which new modes with nonzero coupling to the vertical dipole are introduced.

For both dipoles, the radiation rate approaches the free-space value as d is increased because in a large cavity there exist a large number of modes available irrespective of the

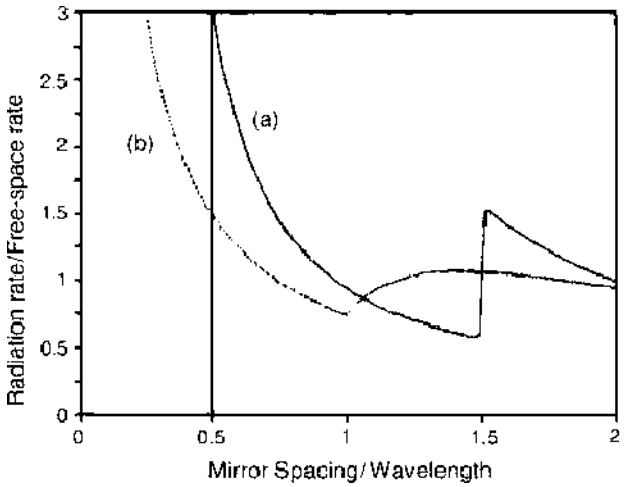


Figure 36. Radiation rates of dipole sources polarized (a) parallel and (b) perpendicular to the mirrors of a parallel plate cavity. The dipole is assumed to be located at the center of the cavity, and the radiation rates are normalized to the free-space value. Reprinted with permission from [86], E. A. Hinds, in "Cavity Quantum Electrodynamics," (P. R. Berman, Ed.), Academic, New York, 1994. © 1994, Elsevier.

polarization. More general cases need full quantum mechanical treatment, in which one must solve the Schrödinger equation for the atom-field system. For a two-level atom located in a cavity with a quality factor Q, one finds two regimes of atom-field system evolution [88]. If the cavity has a very high Q value such that $\omega/Q < 4\Omega$ where ω is the photon frequency and Ω is the Rabi frequency, the atom-field states exhibit coupled oscillations with a frequency 2Ω, and the oscillation is damped at a rate of $\omega/2Q$, as shown in Fig. 37a. This behavior represents a periodic exchange of energy between atom and field and is called the *Rabi oscillation*. This effect, dramatically showing the reversibility of spontaneous emission, is caused by the spontaneously emitted photon staying inside the cavity long enough to be reabsorbed by the atom. The damping term obviously represents the rate at which the energy inside the cavity decays away. In the opposite regime, where $\omega/Q > 4\Omega$, the cavity damping rate is high and, thus the atom-field states exhibit irreversible decay, as shown in Fig. 37b.

The decay is quasi-exponential with a rate

$$\Gamma_{\text{cav}} = 4\Omega^2 \frac{Q}{\omega} = \eta_{\text{cav}} \Gamma_0 \quad \text{where } \eta_{\text{cav}} = \frac{3}{4\pi^2} \frac{Q\lambda^3}{V_{\text{eff}}} \tag{54}$$

Here Γ_0 is the free-space decay rate and V_{eff} is the effective cavity volume. Equation (54) shows that the spontaneous emission rate of an atom inside the cavity is enhanced by a factor of η_{cav}, often called the Purcell factor, compared to the free space rate. If the effective cavity volume is similar to λ^3, then the enhancement factor is on the order of the cavity Q. This phenomenon was first theoretically predicted by Purcell more than 50 years ago and is thus called the Purcell effect [89].

Although the Purcell effect in a planar cavity has been extensively investigated both theoretically and experimentally, cavities providing three-dimensional confinement are much more desirable because they can support small effective cavity volume and also independent control of cavity volume and quality factor. Many different ways to create three-dimensional cavities have been investigated during the past decade, including microspheres, pillars, disks, wires, and PCs [90]. Micropillars attracted much attention because they have the same structure as the vertical cavity surface emitting laser. A micropillar is typically composed of a λ cavity sandwiched between a pair of distributed Bragg reflectors, which provide vertical confinement of light. The lateral confinement is achieved by the large refractive index difference between the semiconductor and air.

Figure 38 shows a GaAs/AlAs micropillar structure with a diameter of 1 μm [91]. Using the linear combination of guided modes in a GaAs cylinder, the effective cavity volume was estimated to be $\sim 5(\lambda_c/n)^3$. The resonant modes of the cavity are usually studied by investigating the photoluminescence from the quantum dots embedded inside the cavity. The resultant photoluminescence spectrum consists of a series of sharp lines whose positions

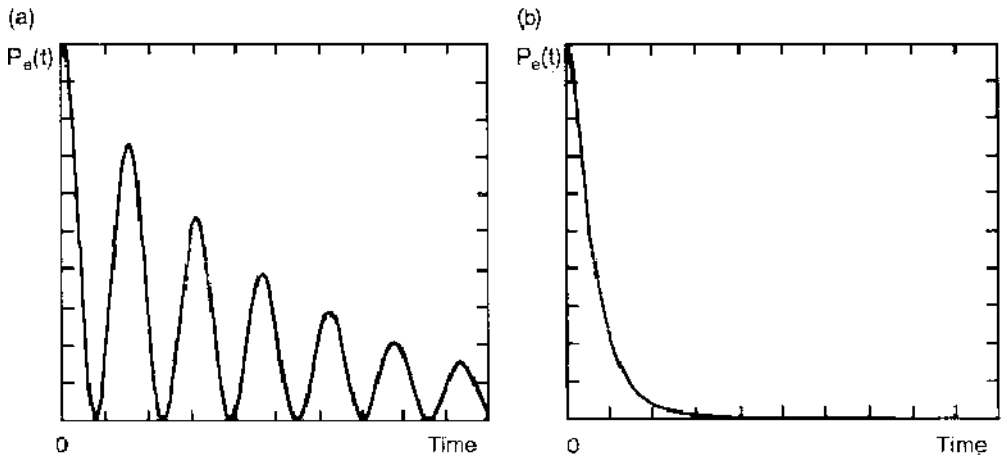


Figure 37. Decay of a two-level atom in a cavity. (a) oscillatory regime, $\omega/Q = 0.2\Omega$; (b) overdamped regime, $\omega/Q = 5\Omega$. Reprinted with permission from [88], S. Haroche and J. M. Raimond, in "Advances in Atomic and Molecular Physics" (D. Bates and B. Bederson, Ed.), Vol. 20, Academic, Orlando, FL, 1985. © 1985, Elsevier.

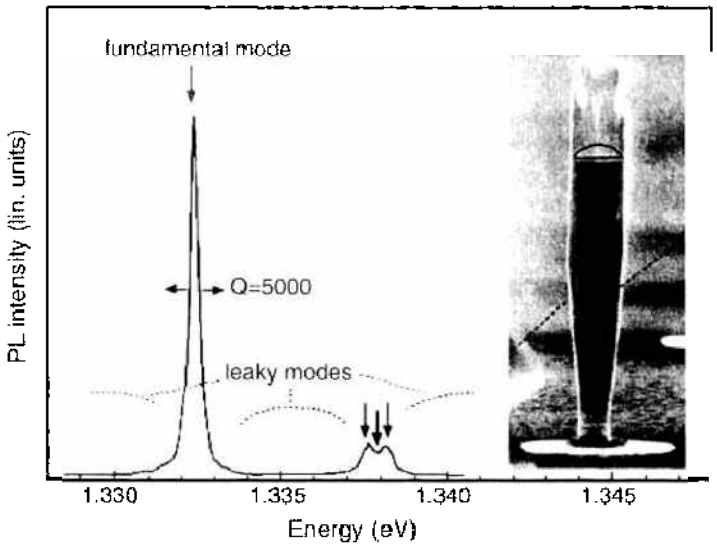


Figure 38. Photoluminescence spectrum for a 3-μm-diameter GaAs/AlAs micropillar containing InAs quantum dots. Inset: scanning electron micrograph of a 1-μm-diameter pillar structure. Reprinted with permission from [91], J.-M. Gerard and B. Gayral, in "Confined Photon Systems: Fundamentals and Applications," H. Benisty et al., Springer, Berlin, 1998, © 1998, Springer-Verlag.

and widths provide cavity mode frequencies and Q factors. The Q factor was found to be close to the planar cavity value for large pillar diameters and to decrease as the pillar diameter was reduced. This was because of the increased scattering loss by the cavity walls. Despite the decrease in Q, the Purcell factor exhibited a sharp increase for small pillar diameters because the decrease in effective cavity volume overcame the degradation in Q. For a cavity with a 1 μm diameter, the Q factor was approximately 1000 and the Purcell factor was found to exceed 30 [91]. Other structures that can exhibit high Q and small-cavity volume are microdisks, microspheres, and microhemispheres, which support whispering gallery modes. In the whispering gallery modes, light is guided by total internal reflection along the curved boundary of the cavity. When an extremely smooth surface is achieved, these modes exhibit extraordinarily high Q values, reaching as high as 10^8 in a near-spherical liquid droplet. As a consequence, a very high Purcell factor, ~120, has been observed for dye molecules in this cavity [90].

All of the above-mentioned conventional cavities rely on refractive-index difference for light confinement. As the cavity volume is reduced, it becomes more and more difficult to achieve high Q factors, and one needs a stronger light confinement mechanism than the refractive index difference. Soon after the introduction of the photonic crystal concept, it was recognized that the photonic band gap could be used to create strong light confinement in a very small volume. Just as a linear defect acts as a waveguide, a point defect in a PC structure can act as an optical cavity. If the PC matrix possesses a complete PBG, light would be strongly confined within the defect, which can lead to a very high quality factor and small mode volume. This defect naturally commanded much attention for its potential for realizing thresholdless lasers. The strong light confinement by the PC structure was first demonstrated in a one-dimensional periodic structure fabricated in a ridge waveguide [92–94].

The structure schematically shown in Fig. 39a is composed of a series of periodic holes patterned on a single-mode Si ridge waveguide. The periodic air holes form a one-dimensional PC structure and produce a one-dimensional PBG (stop band resulting from Bragg reflection), prohibiting light propagation along the waveguide in a certain wavelength range. The dimensions and spacing of the air holes were chosen such that the strop band extends from 1.3 to 1.7 μm. Then a defect is introduced by deliberately increasing the hole spacing in the middle of the structure. The defect produces an allowed mode within the stop band, as shown in Fig. 39b. Numerical calculations predicted that the resonant mode should appear at $\lambda = 1.547$ μm. The resonant mode was also found to be strongly confined within the cavity. The confinement was achieved by the index difference in the direction perpendicular

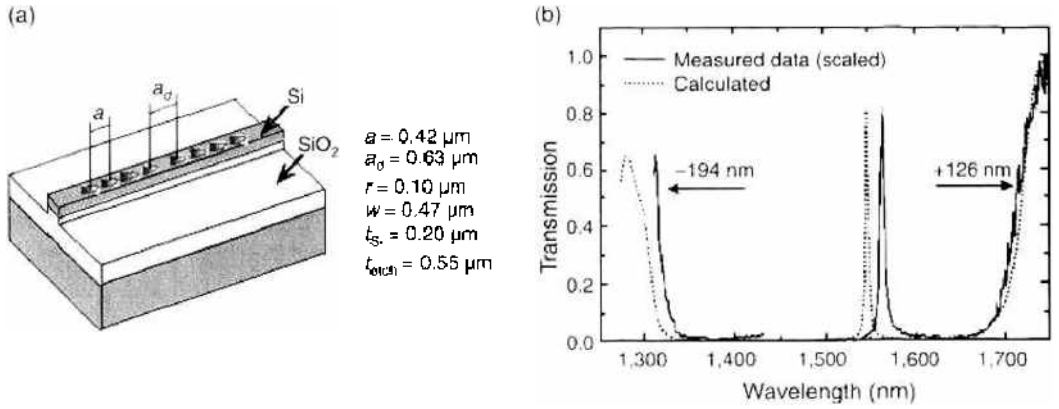


Figure 39. (a) Schematic diagram and dimensions of a photonic crystal cavity formed in a Si ridge waveguide. (b) Theoretical and experimental transmission spectra for the photonic crystal cavity shown in (a). Reprinted with permission from [94]. J. S. Foresi et al., *Nature* 390, 143 (1997). © 1997, Nature Publishing Group.

to the ridge waveguide and by the periodic structure in the direction of the waveguide. The Q factor was estimated as 280 and the modal volume, defined as

$$V = \frac{1}{P_{\max}} \int P(r)\, d^3r \tag{55}$$

where $P(r)$ is energy density and $P_{\max}$ is the peak value of $P(r)$, was found to be 0.055 μm^3. An experimental transmission spectrum, shown also in Fig. 39b, confirmed the theoretical estimations, showing a sharp transmission band at $\lambda = 1.560$ μm and a Q factor of 265. The slight discrepancy of the resonant mode wavelength was well within the experimental error margin of the scanning electron microscope used to determine the feature sizes. The modal volume corresponded to $\sim 5(\lambda_c/2n)^3$ and represented the smallest volume ever achieved at that time.

The success in the one-dimensional PC structure naturally directed attention to the extension of the concept to two-dimensional and three-dimensional systems in which light is confined by PC structures in two or more directions. In the two-dimensional system, the simplest way to create a point defect is to eliminate or modify one rod or hole from a perfect photonic crystal structure. For numerical modeling, one needs to set up a supercell as for the waveguide discussed earlier. This time, the supercell contains many unit cells along both directions in the two-dimensional plane, with the point defect placed at the center. The supercell may be terminated with the periodic boundary condition, which is mandated in the PW method. In this case, one actually models a system of periodically arranged defects.

Therefore, to simulate a single defect surrounded by an infinitely large PC structure, the supercell must be made large enough to prevent any significant coupling between neighboring defects artificially introduced by the periodic boundary condition. In a real space technique such as the FDTD method, one may also use an absorbing boundary condition. In this case, the size of the supercell represents the actual size of the PC structure. One can therefore investigate, for example, the effect of the size of the surrounding PC structure on the Q factor and field profile. Sakoda and Shiroma proposed a unique methodology based on the FDTD scheme for the calculation of defect modes [95]. In this method, a point dipole source is placed inside the photonic defect, and the electromagnetic energy emitted by the dipole is monitored by calculating the Poynting vector. Because the frequency dependence of the emitted energy should peak at the resonant frequency of the defect mode, one can determine the defect mode frequency by calculating the frequency spectrum of the dipole emission inside the defect.

Villeneuve et al. used the PW method to calculate the frequencies and field profiles of the allowed modes in a point defect in a two-dimensional PC structure [96]. The modeled structure is a square array of dielectric rods of radius $0.2a$, and the point defect is created by modifying the radius of one rod. This structure exhibits a wide PBG for TM polarization between normalized frequencies 0.29 and 0.42. Introduction of a point defect subsequently

creates allowed modes within the PBG. The defect mode frequency and field profile depend on the size of the defect rod, as shown in Fig. 40. The mode frequencies were calculated by the PW method using a 7×7 supercell. As the defect rod radius is decreased from the regular value, a new mode begins to appear from the band below the PBG. This is because the band below the PBG has the field concentrated in the dielectric region (dielectric band) and is thus perturbed most significantly by the reduction in size of the defect rod. This is exactly the behavior observed in PC waveguides, too. As the defect rod radius is further decreased, the mode gets pushed further into the PBG. The field profile indicates that the electric field is polarized along the rod axis (TM mode) and rapidly decays away from the defect. The field profile does not exhibit any node in the azimuthal direction and is thus named a monopole mode.

When the rod size is increased, instead, a new mode begins to appear from the band above the PBG. This band has its field mainly in the air region (air band) and is therefore perturbed significantly when the air region is taken away by the enlarged rod. As the defect rod size is increased further, the defect mode frequency decreases, and finally the modes disappear into the dielectric band. This mode is doubly degenerate, and the electric field profile is concentrated along one of the two symmetry directions of the crystal. Because they have two nodes along the azimuthal direction, they are labeled as the dipole mode. As the defect radius is further increased, more modes appear within the PBG, and their field profiles tend to become more and more complicated. One of the key parameters for a cavity is its quality factor, Q. By definition, it represents the rate of energy dissipation away from the cavity [97].

$$Q = \frac{\omega_o E}{P} \tag{56}$$

where ω_o is the resonant frequency, E is the energy stored in the cavity, and P is the rate of energy dissipation away from the cavity. Therefore, Q may be calculated by monitoring the dissipation of energy as a function of time after initially exciting a resonant mode. The FDTD method is then a natural choice for this type of calculations.

Villeneuve et al. investigated the Q factors of a defect formed by eliminating one rod in the square array of rods. The Q factor will obviously be affected by the size of the surrounding PC structure. Figure 41 shows the Q factors calculated for various sizes of surrounding PC matrix. The strong light confinement by the surrounding PBG medium is evident from the fact that the Q factor reaches as high as 10000 for a defect surrounded by the PC structure extending only four unit cells in each direction. The only loss mechanism is the tunneling through the PBG and into the outside medium, and therefore Q does not saturate. This result is, however, somewhat unrealistic, as the structure is assumed to be

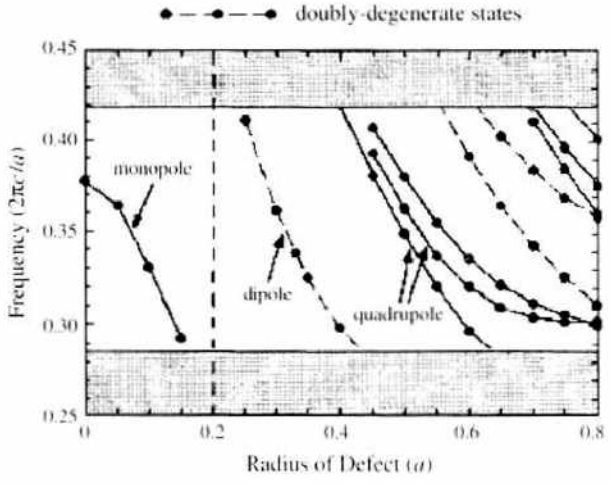


Figure 40. Frequency of the defect modes in a square array of dielectric rods ($r = 0.2a$). The shaded area indicates the allowed bands. Reprinted with permission from [96], P. R. Villeneuve et al., *Phys. Rev. B* 54, 7837 (1996). © 1996, American Physical Society.

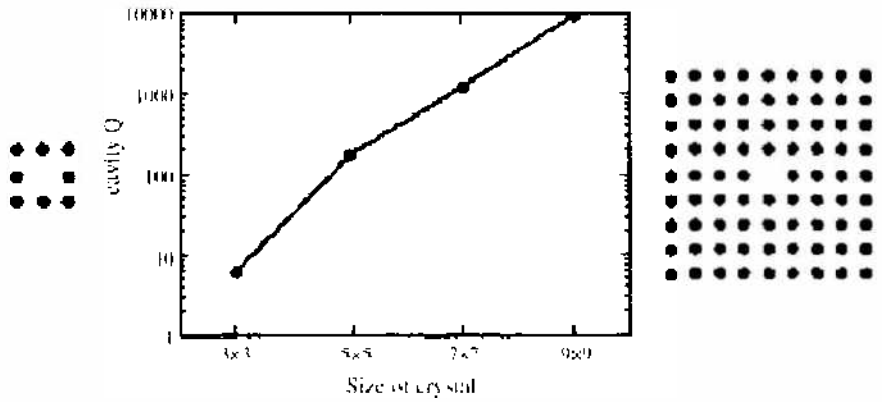


Figure 41. Quality factors as a function of the size of the surrounding photonic crystal matrix. Reprinted with permission from [96], P. R. Villeneuve et al., *Phys. Rev. B* 54, 7837 (1996). © 1996, American Physical Society.

infinitely long and the light is considered to be perfectly confined in the two-dimensional plane perpendicular to the rods. In practice, one has a two-dimensional periodic structure with a finite height and thus must be concerned about the light confinement and propagation in the direction perpendicular to the plane of two-dimensional periodicity.

A defect formed in a triangular array of holes in a dielectric slab has been studied by Painter et al., using the FDTD method [98]. The model structure consisted of a defect surrounded by the hexagonal shape of the triangular PC structure extending three unit cells along all six symmetry directions. The dielectric slab has a refractive index of 3.4 and the hole a radius of $0.3a$. The defect hole has the same radius but a different refractive index, which varied between 1.4 and 3.4. The computational cell was terminated by absorbing boundary conditions in all directions. As in the case of linear waveguides, the analysis of a defect mode in a two-dimensional slab PC must include the guiding condition by the slab. The perfect crystal was found to have a PBG for the TE-like (even) modes in which no guided modes exist. Introduction of a defect then produces allowed modes that are confined by the PBG within the slab and by the index difference along the direction perpendicular to the slab. As the refractive index of the defect hole is increased, the defect mode frequency was found to decrease. This is consistent with the previous observation, where the modification of the air region primarily affects the air band and pushes the mode down into the PBG. The defect mode turned out to be doubly degenerate, and the field profiles exhibited symmetry belonging to the point group C_{6v}, which is the symmetry of the surrounding lattice.

Figure 42 shows the field patterns for the two degenerate defect modes. Inside the cavity, the modes are concentrated along the x- and y-directions and are thus named x- and y-dipole modes, respectively. In both cases, the field was strongly confined within the cavity. In Fig. 42,

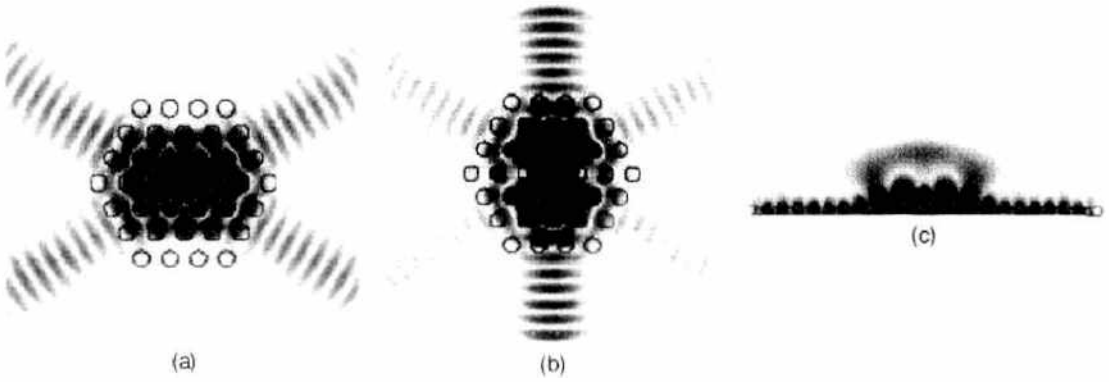


Figure 42. In-plane electric field profiles of (a) x-dipole and (b) y-dipole modes. (c) Electric field profile of the y-dipole in a vertical cross-section. The field amplitudes were enhanced to better show the leakage away from the cavity. Reprinted with permission from [98], O. Painter et al., *J. Opt. Soc. Am. B* 16, 275 (1999). © 1999, Optical Society of America.

the field amplitude was enhanced to show the coupling outside the cavity. Along the vertical direction, both modes exhibited an antinode in the center of the slab, which is important for efficient coupling with the light emission from the semiconductor. The in-plane coupling away from the cavity was found to be highly directional, as clearly shown in Fig. 42. This directional out-coupling must be considered when designing a coupled system of a waveguide and a cavity.

The radiation pattern along the vertical direction is also shown in Fig. 42. The mode volume estimated according to Eq. (55) was $\sim 2(\lambda_c/2n)^3$, which was much smaller than the one-dimensional PC cavity. The Q factor for this cavity was also calculated using the FDTD method where the power absorbed by the absorbing boundary layers was monitored and compared with the energy stored in the cavity. This method allows the separation of vertical and in-plane Q factors, $Q_\perp$ and $Q_{\parallel}$, whose harmonic mean gives the total Q factor, $Q_{\text{Total}}^{-1} = Q_\perp^{-1} + Q_{\parallel}^{-1}$. The Q factor has a maximum when the cavity mode exists in the middle of PBG and decreases rapidly as it approaches the continuum of slab PC bands. This means that the defect possessing a refractive index lower than that of the slab has much greater Q factor than the simple missing hole, which has the same index as the slab. The total Q factor was found to be limited by the in-plane Q factor, which is probably a result of the rather small size of the PC structure used in the simulation. When the number of layers of the surrounding PC structure was increased to seven, the in-plane Q factor exceeds the vertical Q, resulting a total Q factor of ~20000. The vertical Q factor was very sensitive to the relative position of the air band with respect to the light line. Using a larger air-hole size for the surrounding PC matrix tends to shift the air band closer to the light line, subsequently increasing the coupling of light into the vertical direction. This air-hole size can be used as an independent parameter to optimize the radiation pattern in the out-of-plane direction.

As discussed earlier, the Purcell factor has quantum mechanical origin and must be calculated by CQED. However, it has been shown that classical electrodynamics gives the same results as the quantum mechanical approach [87, 99]. This allows the use of the numerical modeling tools based on classical theory to directly compute the Purcell factor. In the classical description, spontaneous emission rate is given by the rate of radiation by an oscillating dipole, which can be calculated by the surface integral of Poynting vector. This can be done in a straightforward manner by using, for example, the FDTD method [100, 101].

In the scheme proposed by Hwang et al. [101], a point dipole source is placed inside a cavity, and the radiation power is computed by evaluating the Poynting vector over a surface enclosing the source. The infinite extent of the surrounding medium was simulated by the PML boundary condition, and the frequency spectrum is obtained from a single computation by imposing a Gaussian pulse excitation for the dipole source. The resulting time series of the fields were then Fourier transformed to obtain frequency spectrum. When the computational cell size is large, and thus the number of points at which the field values need be stored is large, the discrete Fourier transform [102] becomes more efficient than the popular fast Fourier transform (e.g., Ref. [103]). In the discrete Fourier transform, the summation of time series is carried out along with the progression of FDTD time evolution, and thus the storage of the entire time series is not necessary, which significantly relieves the memory requirement. It is also a significant advantage of the discrete Fourier transform that the frequency resolution can be chosen independent of the length of the times series.

The validity of this scheme was tested by calculating the spontaneous emission rate in a parallel mirror cavity, which showed excellent agreement with the analytical solution. It was then applied to the radiation rate calculations for point dipole sources in various PC structures. Figure 43a shows the PC-based nanocavity for which the emission from a point dipole source was investigated. As mentioned before, this structure exhibits a band gap for TE-like (even) modes if there were no defect, and the introduction of the missing hole produces cavity modes within the band gap. The radiation rate calculated for a point dipole polarized along the y-direction is shown in Fig. 43b. The radiation rate is strongly enhanced near the band edges, which could be attributed to the increased photon density of states. However, the photons emitted in this region will mostly be guided within the slab and therefore will not

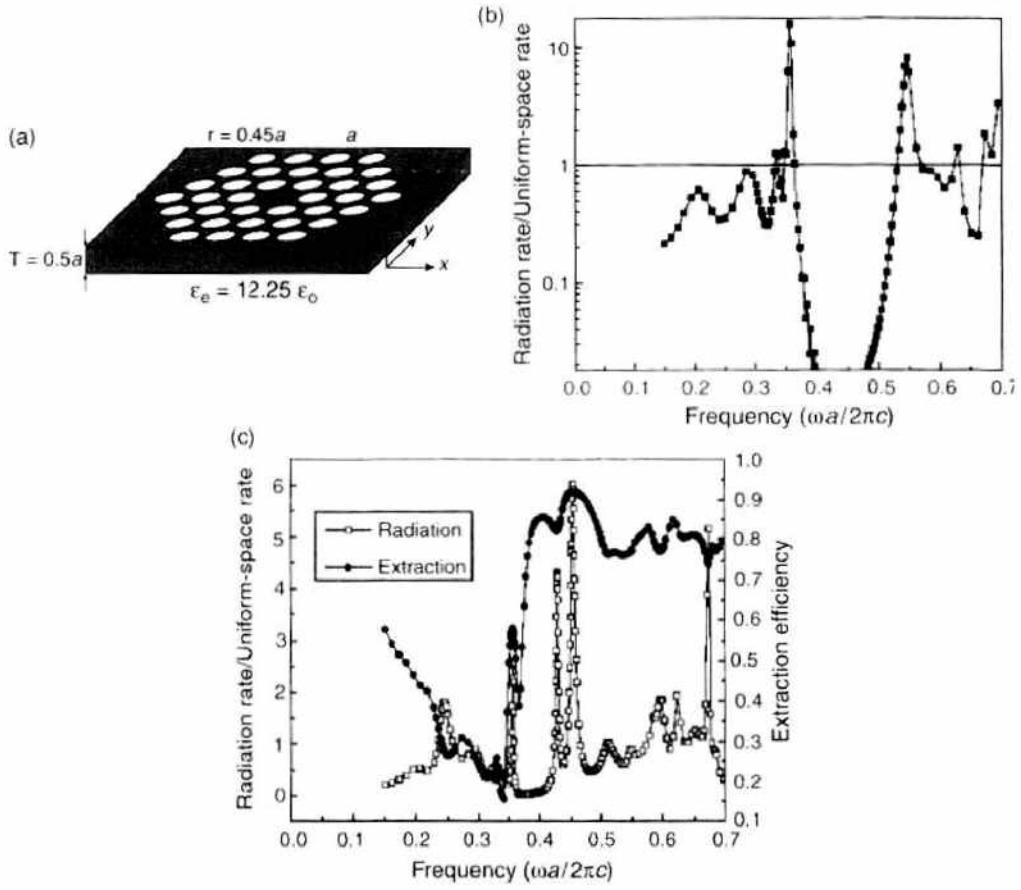


Figure 43. (a) Schematic diagram of the PC-based nanocavity structure investigated for spontaneous emission enhancement effect. (b) Radiation rate of a point dipole polarized along the y-direction located at the center of the defect. (c) Radiation rate of the same point dipole as in (b) but shifted along the x-direction by $0.45a$ from the center of the defect. Reprinted with permission from [101], J.-K. Hwang et al., *Phys. Rev. B* 60, 4688 (1999). © 1999, American Physical Society.

contribute to the emission in the vertical direction. However, the radiation rate was greatly suppressed within the band gap, and the emission caused by the localized mode was not observed despite the presence of the defect. This is because of the poor coupling between the defect mode and the point dipole source. The defect mode has very small electric field amplitude near the center of the cavity, which is where the point dipole was located in this simulation.

To confirm this argument, the radiation rate was calculated after shifting the point dipole to $0.45a$ along the x-direction. As shown in Fig. 43c, several lines associated with the defect modes were clearly visible. Both the radiation rates and extraction efficiencies were high at the defect mode frequencies. The extraction rate was defined as the ratio of light flux to the top and bottom surfaces of the slab, to the total emitted flux. This was calculated by taking the ratio of the z-component (the vertical component) of the Poynting vector integrated over a closed surface containing the dipole to the same integral performed for the normal component of the Poynting vector. Compared to the uniform dielectric slab, the light emission along the vertical direction at the defect mode frequencies was enhanced by as much as a factor of nearly 50, which was attributed to the combination of spontaneous emission enhancement and increased light coupling into the out-of-plane direction.

While the defect in a two-dimensional slab PC exhibits a large Purcell factor thanks to the small mode volume and high Q, for applications to thresholdless lasers, one needs to consider the spontaneous emission coupling factor, C, which is defined as the ratio of power emitted to the cavity mode to the total power emitted. We note that the same factor is denoted as β in some literature, but here we follow the notation by Baba et al. [9, 104]. The variable C is obviously dependent on the fractional solid angle subtended by the cavity mode β, which ranges between 10^{-10} and 10^{-5} in conventional laser cavities. Because of the small values

of fractional solid angle, the number of cavity mode photons increase substantially only when the pumping reaches the threshold for stimulated emission. If a substantial amount of spontaneously emitted photons are coupled into the cavity mode, one would expect a subsequent lowering of the threshold.

Yokoyama et al. solved the rate equation for a four-level laser as a function of the fractional solid angle subtended by the cavity mode, β [105]. Figure 44 shows the population inversion and output power as a function of input power for various values of β. It is clearly shown that the threshold pumping level to achieve population inversion decreases as β is increased. More remarkably, the threshold behavior in the output power becomes less pronounced with increasing β and completely disappears when β reaches unity. It should be noted that even for the case in which $\beta = 1$, the laser oscillation begins only after the pumping level reaches a certain threshold, as shown in the population inversion curve. Below this threshold, the emission consists primarily of spontaneously emitted photons. It is therefore observed that the thresholdless laser does in fact have a threshold for laser oscillation, but it does not appear in the output power curve.

The analyses of the spontaneous emission coupling factor have so far been performed mainly for cavities with simple geometry, but for a nanocavity defined in a PC structure, one needs to incorporate more rigorous numerical computation methods. Vuckovic et al. has proposed using the FDTD method for the calculation of the spontaneous emission coupling factor of PC-based nanocavities [106]. In this method, the resonant modes for the given cavity structure are first obtained by solving the Maxwell's equation with the FDTD method, as described previously. Then, the radiation from the semiconductor material inside the nanocavity is simulated by placing randomly polarized and randomly distributed point dipole sources in the cavity. The dipole sources are assumed to have a single frequency and a single lifetime, but random phase. The total radiation power is then calculated from the surface integral of the Poynting vector, and the total radiated energy W_{∞} is given by the time integral of the radiation power. The total energy radiated into the fundamental mode is given by

$$W_{0\infty} = \int_0^{\infty} dt \frac{1}{2} \mathrm{Re}\left[\int_V \mathbf{E}_0^*(\mathbf{r}, t) \cdot \frac{\partial \mathbf{P}(\mathbf{r}, t)}{\partial t} d^3 r\right] \tag{57}$$

where $\mathbf{P}(\mathbf{r},t)$ is the polarization resulting from the point dipole sources and $\mathbf{E}_0(\mathbf{r}, t)$ is the field profile of the fundamental ($0th$) mode whose time dependence is given by the wave equation. The spontaneous-emission coupling factor is then obtained by the ratio $W_{0\infty}/W_{\Sigma\infty}$. If inhomogeneous broadening is to be included, the radiated energy must be calculated for various frequencies and multiplied by the line shape function to obtain the total radiated energy over the broadened frequency range.

This approach has been applied to the structure shown in Fig. 42. The active region was assumed to be a quantum well structure located in the middle of the slab, and thus strong

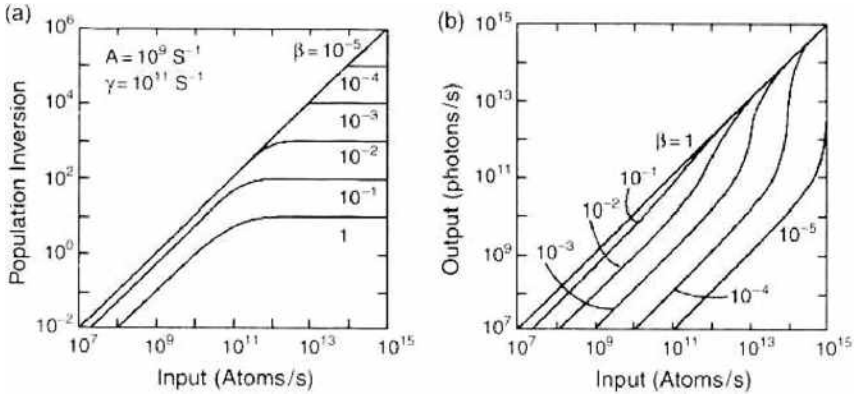


Figure 44. Calculated (a) population inversion and (b) light output for a four-level laser with various values of fractional solid angle subtended by the cavity mode, β. Reprinted with permission from [105]. Yokoyama et al., *Opt. Quantum Electron.* 24, S245, (1992). © 1992, Plenum.

coupling is expected for TE-like polarization. The dipoles were therefore assumed to be polarized in the plane of the quantum well, and their lifetime was set to be 0.1 ps. Figure 45 shows the spontaneous emission coupling factor calculated by assuming only homogeneous broadening. The coupling factor has a maximum when the oscillation frequency of the dipole sources is equal to the cavity resonance frequency and decreases rapidly as it gets detuned from the resonance. The effect of inhomogeneous broadening has also been investigated by assuming a Lorentzian line shape function. When the full width at half maximum (FWHM) was 25 nm, the coupling factor was 46%. As the broadening became more significant, the coupling factor decreased and became 40% for the FWHM value of 200 nm.

The cavity modes can be tailored in a similar fashion to the waveguides discussed in the previous section. As discussed earlier, the simplest form of defect cavity in a PC structure is created by removing a hole, as shown in Fig. 46a. This type of defect forms a symmetric cavity supporting a doubly degenerate mode, which are the x- and y-dipole modes shown in Fig. 42. The modes are located deep inside the band gap and thus exhibit small mode volume and high Q factor. Painter et al. suggested lifting the degeneracy and modifying the mode structure by breaking the cavity symmetry [107]. Figure 46b shows a cavity modified by vertically shifting the four holes located above and below the cavity. This cavity is no longer symmetric, and thus we expect the two originally degenerate cavity modes to split. The radiation rate calculation by the FDTD method showed the splitting, which was also confirmed by the photoluminescence spectroscopy.

Figure 47a shows the calculated and measured spectra for the modified cavity shown in Fig. 46b, where the shift $\Delta y = 0.05a$. It is clearly shown that the originally degenerate modes split into two modes at $\omega a/2\pi c = 0.385$ and 0.374. The field profile indicated that the mode with strong field amplitude along the vertical direction, along which the cavity dimension was decreased, possessed a higher frequency than the mode with fields concentrated along the lateral direction. The modification of the cavity geometry was also found to increase the out-coupling in the out-of-plane direction, decreasing the Q factors for both split modes. The additional modes, labeled as SA modes, were the result of the nonuniformity of the holes introduced in the fabrication process.

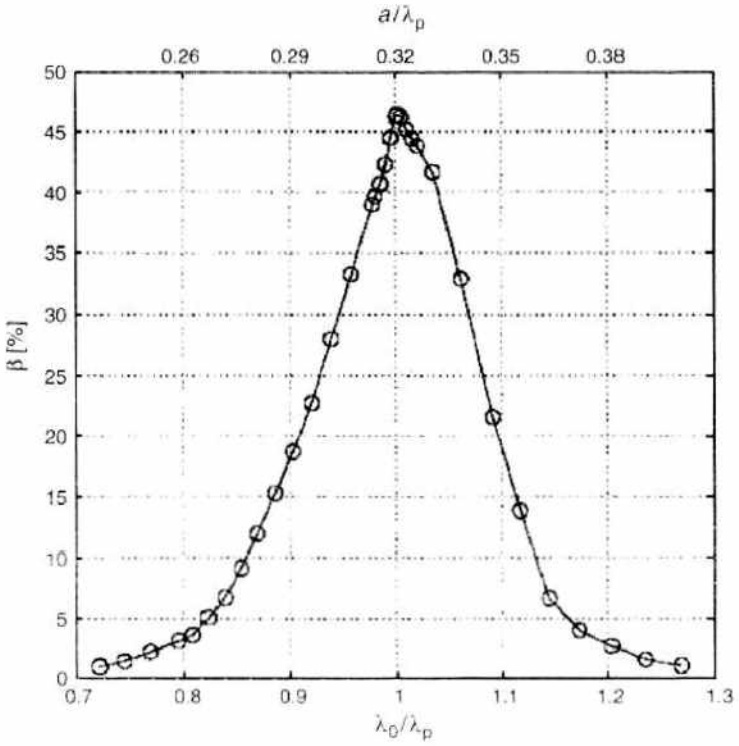


Figure 45. The spontaneous emission coupling factor for various dipole emission wavelength λ_0 for a photonic crystal-based cavity with resonant wavelength λ_p. Reprinted with permission from [106], J. Vuckovic et al., *IEEE J. Quantum Electron.* 35, 1168 (1999). © 1999, IEEE.

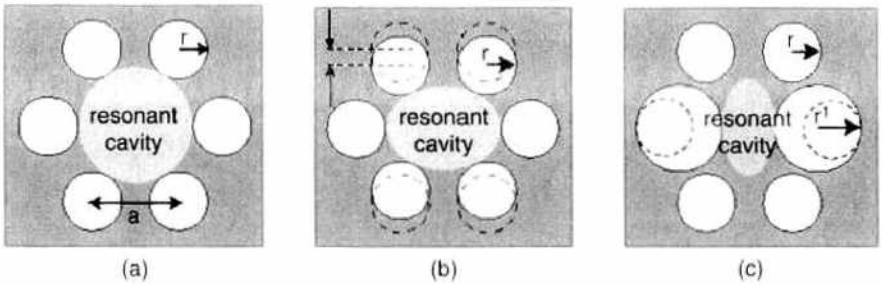


Figure 46. Various geometries for a photonic crystal cavity. (a) Symmetric cavity, (b) cavity with contracted vertical dimension, and (c) cavity with contracted lateral dimension. Reprinted with permission from [107]. O. Painter et al., *J. Opt. A Pure Appl. Opt.* 3, S161 (2001). © 2001, IOP Publishing Ltd.

Another type of modification is shown in Fig. 46c. This time, the lateral dimension of the cavity is decreased by enlarging the two nearest-neighbor holes located along the horizontal (Γ–M) direction and also moving them inward. The resultant emission spectra for this type of cavity are shown in Fig. 47b. The enlarged holes perturb the dielectric band and create new modes (acceptor-like modes) within the band gap. These modes are labeled DA modes in Fig. 47b. The originally degenerate modes are once again split, but this time the splitting is so large that the one of the split mode gets pushed out of the band gap. The remaining mode, labeled DD mode in Fig. 47b, is the one with its field profile concentrated along the vertical direction. The light confinement was found to be improved in this cavity, and both the in-plane and out-of-plane Q factors were greater than those for the symmetric cavity. This cavity also exhibited a much improved spontaneous emission coupling factor compared to the symmetric cavity. According to the FDTD simulations by Vuckovic et al., the coupling factor can be as high as 87%, which is almost twice as high as that for a symmetric cavity [106].

A different type of asymmetric cavity suggested by Vuckovic et al. involves elongation of holes on the horizontal axis, as shown in Fig. 48a [108]. FDTD simulations were carried out for a structure with $n_{\text{slab}} = 3.4$, $n_{\text{defect}} = 2.4$, hole radius $r = 0.3a$, and slab thickness $d = 0.6a$. With five layers of holes surrounding the defect, the perfectly symmetric cavity was found to exhibit $Q_{\parallel} = 2070$, $Q_{\perp} = 1290$, and mode frequency $\omega a/2\pi c = 0.316$. As the elongation is imposed, the cavity dimension is increased along the vertical (y) direction. This modification consequently affects the frequency and Q factor, particularly for the x-dipole mode, whose fields are concentrated along the y-direction. As the elongation was increased, the mode frequency was found to decrease. This behavior could be explained qualitatively by noting that the x-dipole mode has originated from the air band pushed down in frequency into the band gap as a result the perturbation caused by placing a high-index material where

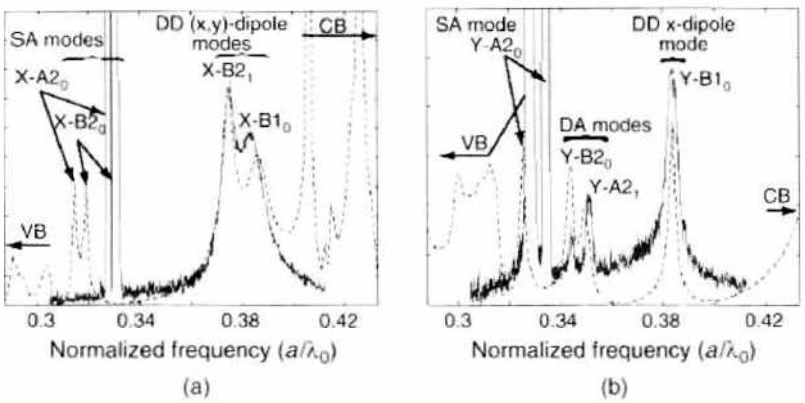


Figure 47. (a) Finite-difference time-domain and photoluminescence spectra for the modified cavity structure shown in Fig. 46(b). (b) Finite-difference time-domain and photoluminescence spectra for the modified cavity structure shown in Fig. 46(c). Reprinted with permission from [107]. O. Painter et al., *J. Opt. A Pure Appl. Opt.* 3, S161 (2001). © 2001, IOP Publishing Ltd.

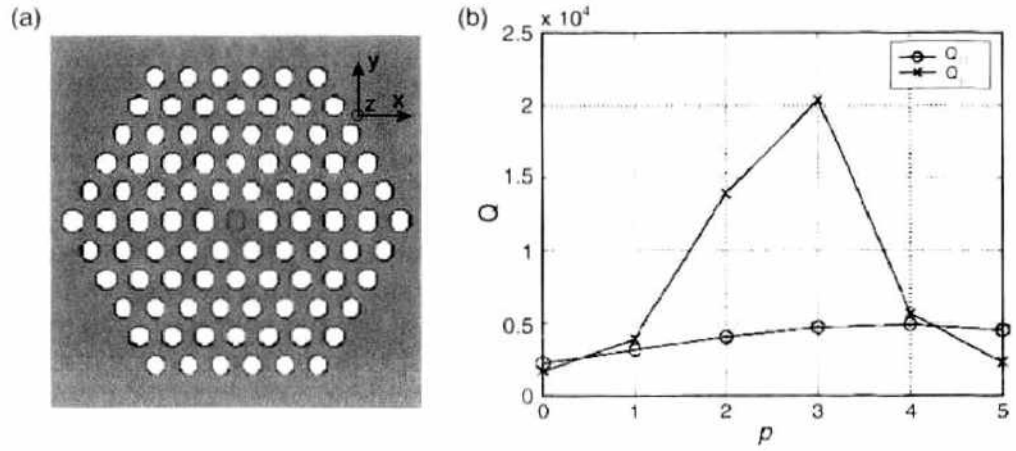


Figure 48. (a) Schematic diagram of a defect cavity modified by elongating a row of holes on the horizontal axis. (b) In-plane and out-of-plane Q factors calculated for various values of elongation parameter, p. $p = 1$ corresponds to $a/15$. Reprinted with permission from [108], J. Vuckovic et al., *Phys. Rev. E* 65, 016608 (2001). © 2001, American Physical Society.

a perfect PC structure would have air. The elongation adds even more high-index material inside the cavity at positions where air would be in a perfectly symmetric cavity. This would then perturb the mode even more, pushing the mode frequency farther down.

More remarkable is the changes in Q factor shown in Fig. 48b. As the structure is elongated, the in-plane Q factor showed only a modest increase, but the out-of-plane Q factor increased dramatically, reaching 20 000 when the elongation was $3a/15$. This was attributed to the fact that the elongation resulted in the suppression of wave vector components of the cavity mode that can couple into the radiation modes. The total Q factor is obviously limited by the in-plane Q factor, which is one order of magnitude smaller. However, the in-plane Q factor can be increased simply by increasing the number of PC layers surrounding the defect. Increasing the number of PC layers to seven yielded an in-plane Q factor exceeding the out-of-plane Q factor, raising the total Q factor above 10 000.

As discovered in Ref. [98], defect with a lowered refractive index is desirable for a high Q cavity. However, in practice such a structure is very difficult to fabricate. To circumvent this difficulty, using a smaller hole to mimic a defect with a lower refractive index was proposed [108, 109]. When this defect is coupled with the elongation discussed above, the Q factor is estimated to reach 4400 and the mode volume $0.43(\lambda/n)^3$. Although the theoretical Q factor was not as high as the lower index defect, this structure has actually been fabricated and shown to exhibit a Q factor of 2800, much larger than what has been measured for other types of PC-based nanocavities.

Although the cavities formed in a triangular lattice continue to be investigated extensively, the much less studied square lattice was recently found to support whispering gallery modes [110]. Ryu et al. performed FDTD simulations on the square lattice of air holes in a dielectric slab with $n = 3.4$. The cavity was formed by eliminating one hole and cavity was found to support two resonant modes, one nondegenerate mode, and one doubly degenerate mode. The doubly degenerate mode is the dipole mode similar to that observed in the cavity formed in a triangular PC structure.

The in-plane and vertical field distributions for the dipole mode are shown in Fig. 49d and 49f. Interestingly, the resonant frequencies of this mode calculated for various air hole radii were found to be located outside the complete in-plane band gap. The frequencies did fall within the pseudo band gap for Γ–X direction, along which the mode predominantly resonates. Unlike the triangular PC lattice, both x- and y-dipoles resonate along the Γ–X direction in the square lattice, which makes it possible to achieve light confinement only with the pseudogap along the Γ–X direction. The in-plane Q factor could become as high as ~1500 for an air-hole radius of $0.35a$.

The nondegenerate mode exhibited completely different in-plane field patterns, as shown in Fig. 49a and 49b. Both the electric and magnetic fields were found to have a node at the

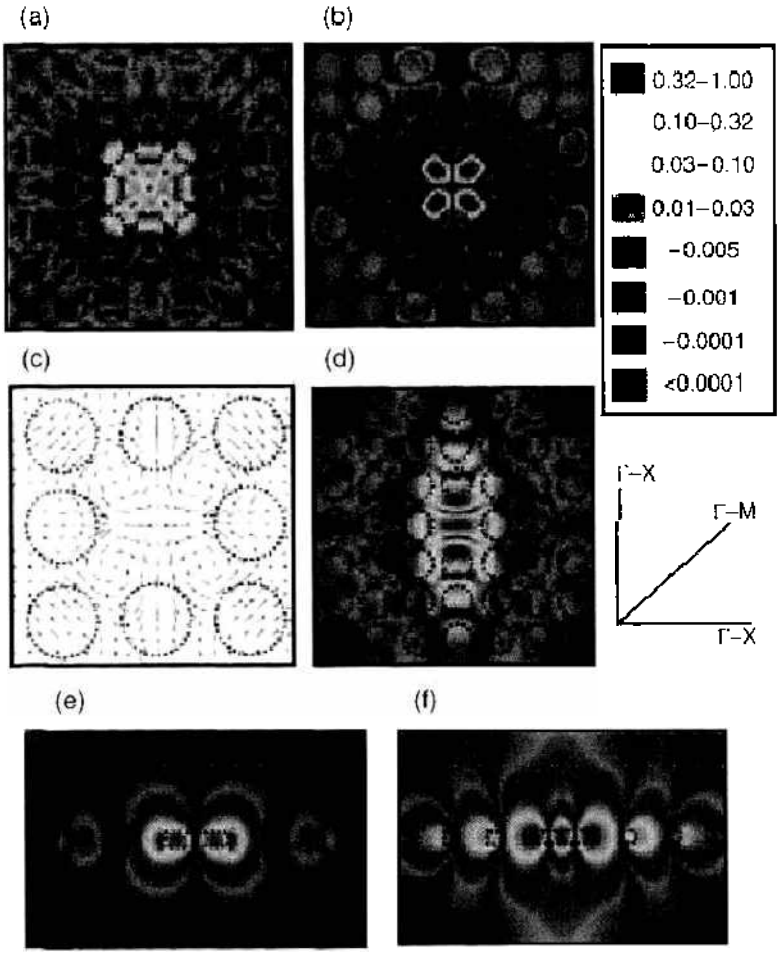


Figure 49. Resonant mode profiles of a free-standing square lattice single-cell cavity calculated by the three-dimensional finite-difference time-domain method. (a) Electric-field intensity profile of the lowest-order whispering gallery (LWG) mode calculated at the center of the slab. (b) Magnetic-field intensity profile of the LWG mode. (c) Electric-field directions of the LWG mode. (d) Electric-field intensity profile of the *y*-dipole mode. Side-view electric-field intensity profile of (e) the LWG mode and (f) the dipole mode are calculated along the Γ–X direction through the defect center. The color map of relative intensity scales is shown for reference. Reprinted with permission from [110], H.-Y. Ryu et al., *Appl. Phys. Lett.* 80, 3883 (2002). © 2002, American Institute of Physics.

center of the cavity and four maxima near the cavity boundary. For the electric field, the maxima occur near the four nearest neighbor holes along the Γ–X directions. The magnetic field, however, showed four maxima along the Γ–M directions. Furthermore, the electric field vector at the maxima pointed inward and outward in an alternating fashion, as shown in Fig. 49c. These are the characteristics of the whispering gallery mode with an azimuthal mode number 2. Consistent with the fact that the whispering gallery modes in microdisk cavities are strongly confined within the slab and radiate primarily in the radial direction, this mode was found to be well-confined in the vertical direction (see Fig. 49e and 49f), yielding much higher Q factors than the dipole modes. The vertical Q factor calculated from the surface integral of Poynting vector in the vertical direction was found to be as high as 40 000 for an 11×11 unit cell system with a hole radius of $0.38a$. The surface for integration was placed at $0.5a$ above the surface of the slab PC structure to avoid complications resulting from the evanescent waves. The high vertical Q factor is very important because that is what essentially limits the total Q in a slab PC cavity, and the in-plane Q can always be increased by increasing the size of the PC structure surrounding the cavity. The whispering gallery mode also exhibited a very small mode volume of $0.04\ \mu m^3$, which corresponds to $0.1(\lambda/2)^3$. The combination of small mode volume and high Q factor led to a very high Purcell factor, exceeding 1000.

In efforts to further improve the cavity characteristics, Ryu et al. investigated the effects of slab thickness and hole size [111]. As the slab thickness was increased, the mode became more confined within the slab, increasing the vertical Q factor. The in-plane Q factor also increases initially as a result of the fact that the in-plane PBG widens as the slab thickness is increased. However, the increased slab thickness also shifts the mode frequency lower. As a result, the cavity mode gets closer in frequency to the guided modes allowed to propagate within the slab, and this reduces the in-plane Q factor. Thus, there exists an optimum thickness at which the in-plane Q factor, and consequently the total Q factor, reach maximum. For a 13×13 structure with a hole radius of $0.38a$, the total Q factor has a maximum of $\sim 18\,000$ when the slab thickness is $0.6a$. One may also attempt to further increase the Q factor by modifying the hole sizes in a similar manner to what has been done for the triangular lattice.

Ryu et al. modeled the modified cavity structures in which the size of the nearest and second-nearest neighboring holes were different from the regular hole size [111]. In contrast to the triangular lattice, modification of neighboring hole sizes only decreased the Q factor, and the unmodified structure in which the hole sizes are the same for all holes exhibited the highest Q factor. The effective mode volume was also calculated using Eq. (55). With increasing slab thickness, the mode volume naturally increases. When expressed in units of cubic lattice constant a^3, the effective mode volume increased linearly with increasing slab thickness. However, because the change in slab thickness shifts the resonant frequency of the cavity mode, when expressed in units of cubic half-wavelength, $(\lambda/2)^3$, the mode volume was found to exhibit a minimum at a slab thickness of $0.5a$. The minimum mode volume was $0.093(\lambda/2)^3$. From the Q factor and effective mode volume, the Purcell factor was estimated and was found to reach a maximum of ~ 8000 for a slab thickness of $0.6a$ at which the Q factor was the largest.

6. CONCLUDING REMARKS

An overview was given on the basic theory and numerical techniques of photonic crystal modeling and on how they are applied to the forefront of PC research. The basic theoretical framework for PCs has been established on the marriage of condensed matter physics and optics. The analogy between PC and real crystal is quite remarkable and is clearly reflected in the theoretical development of PC. It has also become clear, however, that there are fundamental differences between the photonic and electronic systems: for example the vector nature of an electromagnetic field versus the scalar nature of an electronic wave. Thus, the theoretical development of PC often involved adaptation and extension of the condensed matter theory to properly describe the photonic system. The theoretical research has clearly been the driving force of initial PC research and has put the emerging new field on a firm scientific foundation. For example, the predictions of complete PBG by the early theoretical studies attracted much attention and spawned a wide array of experimental works to realize it. Light localization by the PBG has also received extensive theoretical investigations, leading to the conception of waveguide with a sharp bend and extremely high Q nanocavity. Even now, when the fabrication of the complex nanostructures and experimental demonstration of the theoretically predicted novel phenomena are increasingly emphasized, the modeling of PC remains a critical component of research as design and validation tool.

Furthermore, theoretical modeling studies continue to pioneer new areas and keep discovering new optical phenomena possible in PCs. For example, it was recently discovered that the interaction between shock wave and electromagnetic wave in a PC can lead to frequency up-conversion, slowing of light, and pulse narrowing in an optically linear medium [112]. Also fast emerging is the study of nonlinearity in photonic crystals. The novel optical properties and the design flexibility of photonic crystal structures provide unique opportunities to tailor the input and output characteristics of a nonlinear nanophotonic device. For example, Soljacic et al. recently reported that optical bistability observed a coupled system of nanocavity and waveguide based on two-dimensional PC [113]. These systems could potentially lead to all-optical photonic circuits for communication, interconnection, and information processing.

It should be pointed out that the theoretical development of photonic crystals has largely been based on the classical electrodynamics. Although some quantum effects such as the spontaneous emission enhancement in nanocavities have been extensively studied, it was found that classical theory adequately describes them, as discussed in the text. However, as the theoretical research on PCs continues to pioneer new fronts, it will be necessary to develop fully quantum mechanical treatments. Another important issue to be more completely addressed by theoretical studies is that of integration. One of the greatest advantages of PC is found in its potential for large-scale integration, which will involve numerous mutually interacting devices in a compact and complex geometry. Therefore, controlling the light propagation and coupling between various components will be critical to the realization of large-scale integrated photonic systems, and extensive theoretical studies will be needed to address this issue.

REFERENCES

1. E. Yablonovitch, *Phys. Rev. Lett.* 58, 2059 (1987).
2. S. John, *Phys. Rev. Lett.* 58, 2486 (1987).
3. A. Mekis, J. C. Chen, I. Kurland, S. Fan, P. R. Villeneuve, and J. D. Joannopoulos, *Phys. Rev. Lett.* 77, 3787 (1996).
4. M. Loncar, J. Vuckovic, and A. Scherer, *J. Opt. Soc. Am. B* 18, 1362 (2001).
5. S. G. Johnson, P. R. Villeneuve, S. Fan, and J. D. Joannopoulos, *Phys. Rev. B* 62, 8212 (2000).
6. S. Fan, P. R. Villeneuve, J. D. Joannopoulos, and H. A. Haus, *Phys. Rev. Lett.* 80, 960 (1998).
7. A. Mekis and J. D. Joannopoulos, *J. Lightwave Technol.* 19, 861 (2001).
8. S. Haroche and D. Kleppner, *Phys. Today* 42, 24 (1989).
9. T. Baba, T. Hamano, F. Koyama, and K. Iga, IEEE *J. Quantum Electron.* 27, 1347 (1991).
10. T. Yoshie, J. Vuckovic, A. Scherer, H. Chen, and D. Deppe, *Appl. Phys. Lett.* 79, 4289 (2001).
11. J. Vuckovic, M. Loncar, H. Mabuchi, and A. Scherer, *Phys. Rev. E* 65, 016608 (2001).
12. S.-Y. Lin, V. M. Hietala, L. Wang, and E. D. Jones, *Opt. Lett.* 21, 1771 (1996).
13. H. Kosaka, T. Kawashima, A. Tomina, M. Notomi, T. Tamamura, T. Sato, and S. Kawakami, *Phys. Rev. B* 58, 10096 (1998).
14. H. Kosaka, T. Kawashima, A. Tomita, M. Notomi, T. Tamamura, T. Sato, and S. Kawakami, *Appl. Phys. Lett.* 74, 1370 (1999).
15. M. Notomi, *Phys. Rev. B* 62, 10696 (2000).
16. R. E. Slusher and C. Weisbuch, *Solid State Commun.* 92, 149 (1994).
17. M. Tinkham, "Group Theory for Quantum Mechanics" McGraw-Hill, New York, 1964.
18. S. G. Johnson and J. D. Joannopoulos, *Optics Express* 8, no. 3, 173–190 (2001). http://www.opticsexpress.org/abstract.cfm?URI-OPEX-8-3-173.
19. K. Sakoda, *Phys. Rev. B* 52, 7982 (1995).
20. K. M. Ho, C. T. Chan, and C. M. Soukoulis, *Phys. Rev. Lett.* 65, 3152 (1990).
21. P. R. Villeneuve and M. Piche, *Prog. Quantum Electron.* 18, 153 (1994).
22. R. D. Meade, A. M. Rappe, K. D. Brommer, J. D. Joannopoulos, and O. L. Alerhand, *Phys. Rev. B* 48, 8434 (1993).
23. K. S. Kunz and R. J. Luebbers, "The Finite Difference Time Domain for Electromagnetics." CRC Press, Boca Raton, FL, 1993.
24. A. Taflove and S. C. Hagness, "Computational Electrodynamics: The Finite-Difference Time-Domain Method," 2nd ed., Artech House, Norwood, 2000.
25. K. S. Yee, *IEEE Trans. Antennas Propag.* AP-14, 302 (1966).
26. Finite-Difference Time-Domain Literature Database, www.fdtd.org.
27. G. Mur, *IEEE Trans. Electromagn. Compat.* EMC-23, 377 (1981).
28. B. Engquist and A. Majda, *Math. Comp.* 31, 629 (1977).
29. J. P. Berenger, *J. Comp. Phys.* 114, 185 (1994).
30. C. T. Chan, Q. L. Yu, and K. M. Ho, *Phys. Rev. B* 51, 16635 (1995).
31. K. Sakoda, N. Kawai, T. Ito, A. Chutinan, S. Noda, T. Mitsuyu, and K. Hirao, *Phys. Rev. B* 64, 045116 (2001).
32. J. B. Pendry and A. MacKinnon, *Phys. Rev. Lett.* 69, 2772 (1992).
33. J. B. Pendry, *J. Mod. Opt.* 41, 209 (1994).
34. X. Wang, X.-G. Zhang, Q. Yu, and B. N. Harmon, *Phys. Rev. B* 47, 4161 (1993).
35. N. Stefanou, V. Karathanos, and A. Modinos, *J. Phys. Condes. Matter* 4, 7389 (1992).
36. K. Ohtaka and Y. Tanabe, *J. Phys. Soc. Japan* 7, 2265 (1996).
37. A. Moroz, *Phys. Rev. B* 51, 2068 (1995).
38. K. Busch, *C. R. Physique* 3, 53 (2002).
39. N. Stefanou and A. Modinos, *J. Phys. Condens. Matter* 3, 8135 (1991).
40. M. Kerker, "Scattering of Light" Academic, New York, 1969.
41. N. Stefanou, V. Yannopapas, and A. Modinos, *Comp. Phys. Comm.* 113, 49 (1998).
42. P. Yeh and A. Yariv, *J. Opt. Soc. Am.* 67, 423 (1977).

43. J. D. Joannopoulos, R. D. Meade, and J. N. Winn, "Photonic Crystals: Molding the Flow of Light." Princeton University Press, Princetor, NJ, 1995.
44. C. M. Anderson and K. P. Giapis, *Phys. Rev. Lett.* 77, 2949 (1996).
45. P. R. Villeneuve and M. Piche, *Phys. Rev. B* 46, 4969 (1992).
46. X. H. Wang, B. Y. Gu, Z. Y. Li, and G. Z. Yang, *Phys. Rev. B* 60, 11417 (1999).
47. R. Wang, X.-H. Wang, B.-Y. Gu, and G.-Z. Yang, *J. Appl. Phys.* 90, 4307 (2001).
48. U. Gruning, V. Lehmann, S. Ottow, and K. Busch, *Appl. Phys. Lett.* 68, 747 (1996).
49. H. Masuda, M. Ohya, K. Nishio, H. Asoh, M. Nakao, M. Nohtomi, A. Yokoo, and T. Tamamura, *Jpn. J. Appl. Phys.* 39, L1039 (2000).
50. S. G. Johnson, S. Fan, P. R. Villeneuve, J. D. Joannopoulos, and L. A. Kolodziejski, *Phys. Rev. B* 60, 5751 (1999).
51. H. Benisty, D. Labilloy, C. Weisbuch, C. J. M. Smith, T. F. Krauss, D. Cassagne, A. Beraud, and C. Jouanin, *Appl. Phys. Lett.* 76, 532 (2000).
52. W. Bogaerts, P. Bienstman, D. Taillaert, R. Baets, and D. De Zutter, *IEEE Photonics Technol. Lett.* 13, 565 (2001).
53. E. Yablonovitch and T. J. Gmitter, *Phys. Rev. Lett.* 63, 1950 (1989).
54. E. Yablonovitch, T. J. Gmitter, and K. M. Leung, *Phys. Rev. Lett.* 67, 2295 (1991).
55. D. J. Norris and Y. A. Vlasov, *Adv. Mater.* 13, 371 (2001).
56. Y. Xia, B. Gates, Y. Yin, and Y. Lu, *Adv. Mater.* 12, 693 (2000).
57. K. M. Leung and Y. K. Liu, *Phys. Rev. B* 41, 10188 (1990).
58. K. M. Ho, C. T. Chan, and C. M. Soukoulis, *Phys. Rev. B* 46, 10650 (1992).
59. H. S. Sozuer, J. W. Haus, and R. Inguva, *Phys. Rev. B* 45, 13962 (1992).
60. K. Busch and S. John, *Phys. Rev. E* 58, 3896 (1998).
61. K. M. Ho, C. T. Chan, C. M. Soukoulis, R. Biswas, and M. Siglas, *Solid State Comm.* 89, 413 (1994).
62. S. Y. Lin, J. G. Fleming, D. L. Hetherington, B. K. Smith, R. Biswas, K. M. Ho, M. M. Sigalas, W. Zubrzycki, S. R. Kurtz, and Jim Bur, *Nature* 394, 251 (1998).
63. C. T. Chan, S. Datta, K. M. Ho, and C. M. Soukoulis, *Phys. Rev. B* 50, 1988 (1994).
64. S. G. Johnson and J. D. Joannopoulos, *Appl. Phys. Lett.* 77, 3490 (2000).
65. I. El-Kady, M. M. Siglas, R. Biswas, K. M. Ho, and C. M. Soukoulis, *Phys. Rev. B* 62, 15299 (2000).
66. V. Kuzmiak and A. A. Maradudin, *Phys. Rev. B* 55, 7427 (1997).
67. A. Moroz, *Phys. Rev. Lett.* 83, 5274 (1999).
68. W. Y. Zhang, X. Y. Lei, Z. L. Wang, D. G. Zhang, W. Y. Tam, C. T. Chan, and P. Sheng, *Phys. Rev. Lett.* 84, 2853 (2000).
69. Z. Wang, C. T. Chan, W. Zhang, N. Ming, and P. Sheng, *Phys. Rev. B* 64 113108 (2001).
70. S. Fan, P. R. Villeneuve, and J. D. Joannopoulos, *Phys. Rev. B* 54, 11245 (1996).
71. J. G. Fleming, S. Y. Lin, I. El-Kady, R. Biswas, and K. M. Ho, *Nature* 417, 52 (2002).
72. W. Park and C. J. Summers, *Opt. Lett.* 27, 1387 (2002).
73. L. Zhao and A. C. Cangellaris, *IEEE Trans. Microwave Theory Tech.* 44, 2555 (1996).
74. K. Busch and S. John, *Phys. Rev. Lett.* 83, 967 (1999).
75. K. Yoshino, Y. Shinoda, Y. Kawagishi, K. Nakayama, and M. Ozaki, *Appl. Phys. Lett.* 75, 932 (1999).
76. S. W. Leonard, J. P. Mondia, H. M. van Driel, O. Toader, S. John, K. Busch, A. Birmer, U. Gosele, and V. Lehmann, *Phys. Rev. B* 61, R2389 (2000).
77. W. Park and C. J. Summers, "The Fourth International Workshop on Photonic and Electromagnetic Crystal Structures." IPAM, Los Angeles, CA, Oct. 28–31, 2002.
78. D. Scrymgeour, N. Malkova, S. Kim, and V. Gopalan, *Appl. Phys. Lett.* 82, 3176 (2003).
79. S. Xiong and H. Fukshima, *J. Appl. Phys.* 94, 1286 (2003).
80. C. J. Summers and W. Park, *J. Nonlinear Opt. Phys. Mater.* 12, 582 (2003).
81. M. Notomi, *Phys. Rev. B* 62, 10696 (2000).
82. C. Luo, S. G. Johnson, J. D. Joannopoulos, and J. B. Pendry, *Phys. Rev. B* 65, 201104(R) (2002).
83. A. Mekis, S. Fan, and J. D. Joannopoulos, *Phys. Rev. B* 58, 4809 (1998).
84. S. G. Johnson, P. R. Villeneuve, S. Fan, and J. D. Joannopoulos, *Phys. Rev. B* 62, 8212 (2000).
85. M. Loncar, T. Doll, J. Vuckovic, and A. Scherer, *J. Lightwave Technol.* 18, 1402 (2000).
86. M. Loncar, J. Vuckovic, and A. Scherer, *J. Opt. Soc. Am. B* 18, 1362 (2001).
87. E. A. Hinds, in "Cavity Quantum Electrodynamics" (P. R. Berman, Ed.), Academic Press, New York, 1994.
88. S. Haroche and J. M. Raimond, in "Advances in Atomic and Molecular Physics" (D. Bates and B. Bedersor, Eds.), Vol. 20. Academic Press, Orlando, FL, 1985.
89. E. M. Purcell, *Phys. Rev.* 69, 681 (1946).
90. Y. Yamamoto and R. E. Slusher, *Phys. Today* 66 (1993).
91. J.-M. Gerard and B. Gayral, in "Confined Photon Systems: Fundamentals and Applications" (H. Benisty, J.-M. Gerard, R. Houdre, J. Rarity, and C. Weisbuch, Eds.). Springer, Berlin, 1998.
92. P. R. Villeneuve, S. Fan, J. D. Joannopoulos, K.-Y. Lim, G. S. Petrich, L. A. Kolodziejski, and R. Reif, *Appl. Phys. Lett.* 67, 167 (1995).
93. P. R. Villeneuve, D. S. Abrams, S. Fan, and J. D. Joannopoulos, *Opt. Lett.* 21, 2017 (1996).
94. J. S. Foresi, P. R. Villereuve, J. Ferrera, E. R. Thoen, G. Steinmeyer, S. Fan, J. D. Joannopoulos, L. C. Kimmerling, H. I. Smith, and E. P. Ippen, *Nature* 390, 143 (1997).
95. K. Sakoda and H. Shiroma, *Phys. Rev. B* 56, 4830 (1997).
96. P. R. Villeneuve, S. Fan, and J. D. Joannopoulos, *Phys. Rev. B* 54, 7837 (1996).

97. A. Yariv, "Optical Electronics," Saunders, Philadelphia, 1991.
98. O. Painter, J. Vuckovic, and A. Scherer, *J. Opt. Soc. Am. B* 16, 275 (1999).
99. J. P. Dowling and C. M. Bowden, *Phys. Rev. A* 46, 612 (1992).
100. Y. Xu, J. S. Vuckovic, R. K. Lee, O. J. Painter, A. Scherer, and A. Yariv, *J. Opt. Soc. Am. B* 16, 465 (1999).
101. J.-K. Hwang, H.-Y. Ryu, and Y.-H. Lee, *Phys. Rev. B* 60, 4688 (1999).
102. C. M. Furse and O. P. Gandhi, *IEEE Microwave and Guided Wave Lett.* 5, 326 (1995).
103. W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, "Numerical Recipes in Fortran: The Art of Scientific Computing" 2nd ed. Cambridge University Press, New York, 1992.
104. T. Baba, T. Hamano, F. Koyama, and K. Iga, *IEEE J. Quantum Electron.* 28, 1310 (1992).
105. H. Yokoyama, K. Nishi, T. Anan, Y. Nambu, S. D. Brorson, E. P. Ippen, and M. Suzuki, *Opt. Quantum Electron.* 23, S245 (1992).
106. J. Vuckovic, O. Painter, Y. Xu, A. Yariv, and A. Scherer, *IEEE J. Quantum Electron.* 35, 1168 (1999).
107. O. Painter, K. Srinivasan, J. D. O'Brien, A. Scherer, and P. D. Dapkus, *J. Opt. A: Pure Appl. Opt.* 3, S161 (2001).
108. J. Vuckovic, M. Loncar, H. Mabuchi, and A. Scherer, *Phys. Rev. E* 65, 016608 (2001).
109. T. Yoshie, J. Vuckovic, A. Scherer, H. Chen, and D. Deppe, *Appl. Phys. Lett.* 79, 4289 (2001).
110. H.-Y. Ryu, S.-H. Kim, H.-G. Park, J.-K. Hwang, Y.-H. Lee, and J.-S. Kim, *Appl. Phys. Lett.* 80, 3883 (2002).
111. H.-Y. Ryu, J.-K. Hwang, and Y.-H. Lee, *IEEE J. Quantum Electron.* 39, 314 (2003).
112. E. J. Reed, M. Soljacic, and J. D. Joannopoulos, *Phys. Rev. Lett.* 90, 203904 (2003).
113. M. Soljacic, M. Ibanescu, S. G. Johnson, Y. Fink, and J. D. Joannopoulos, *Phys. Rev. E* 66, 055601(R) (2002).

CHAPTER 6

Decoherence, Quantum Information, and Quantum-State Measurement in Quantum Optics

Luiz Davidovich

Instituto de Física, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil

CONTENTS

ISBN: 1-58883-049-7

Handbook of Theoretical and Computational Nanotechnology
Edited by Michael Rieth and Wolfram Schommers
Volume 7: Pages (329–394)

1. INTRODUCTION

The area of quantum information owes its fast development in the past years to the ability, developed in several labs, of controlling and measuring simple microscopic systems, to the discovery of fast quantum algorithms, and to the recognition that Moore's law will soon lead to the single-atom limit of elementary computing gates. This could possibly lead to new architectures of hardware and new methods of computation, using the properties of the quantum world [1–4]. The development of quantum computers faces, however, a major obstacle, due to the fact that quantum software deals with entangled states of many qubits (ideally of the order of 1000): the fast loss of coherence between any two different macroscopic states, which transforms the entangled states into mixtures and the quantum computer into a classical one. This problem, of great practical importance, is related to a very fundamental question in physics.

Indeed, one of the most subtle problems in physics is the relation between the macroscopic world, described by classical physics, and the microscopic world, ruled by the laws of quantum physics [5–7]. Among the several questions involved in the quantum-classical transition, one stands out in a striking way. As pointed out by Einstein in a letter to Max Born in 1954 [8],

it concerns the inexistence at the classical level of the majority of states allowed by quantum mechanics, namely coherent superpositions of classically distinct states. Indeed, whereas in the quantum world one frequently comes across coherent superpositions of states (like in Young's two-slit interference experiment, in which each photon is considered to be in a coherent superposition of two wave packets, centered around the classical paths that stem out of each slit), one does not see macroscopic objects in coherent superpositions of two distinguishable classical states, localized for instance around two distinct positions in space. There is an important difference between a state of this kind and one that would involve just a classical alternative: the existence of quantum coherence between the two localized states would allow in principle the realization of an interference experiment, complementary to the simple observation of the position of the object. We know all this already from Young's experiment: the observation of the photon path (that is, a measurement that is able to distinguish through which slit the photon has passed) unavoidably destroys the interference fringes.

If one assumes that the usual rules of quantum dynamics are valid up to the macroscopic level, then the existence of quantum interference at the microscopic level necessarily implies that the same phenomenon should occur between distinguishable macroscopic states. This was emphasized by Schrödinger in his famous "cat paradox" [9]. An important role is played by this fact also in quantum measurement theory, as pointed out by Von Neumann [10, 11]. Indeed, let us assume for instance that a microscopic two-level system (states $|+\rangle$ and $|-\rangle$) interacts with a macroscopic measuring apparatus in such a way that the pointer of the apparatus points to different (and classically distinguishable!) positions for each of the two states, that is, the interaction transforms the joint atom-apparatus initial state into

$$|+\rangle|\uparrow\rangle \rightarrow |+\rangle'|\nearrow\rangle$$
$$|-\rangle|\uparrow\rangle \rightarrow |-\rangle'|\nwarrow\rangle$$

where one has allowed for a change in the state of the two-level system due to its interaction with the measurement apparatus.

The linearity of quantum mechanics implies that, if the quantum system is prepared in a coherent superposition of the two states, say $|\psi\rangle = (|+\rangle + |-\rangle)/\sqrt{2}$, the final state of the complete system should be a coherent superposition of two product states, each of which corresponds to a different position of the pointer:

$$\left(\frac{1}{\sqrt{2}}\right)(|+\rangle + |-\rangle)|\uparrow\rangle \rightarrow \left(\frac{1}{\sqrt{2}}\right)(|+\rangle'|\nearrow\rangle + |-\rangle'|\nwarrow\rangle) = \left(\frac{1}{\sqrt{2}}\right)(|\nearrow\rangle' + |\nwarrow\rangle') \qquad (1)$$

where in the last step it was assumed that the two-level system is incorporated into the measurement apparatus after their interaction (for instance, an atom that gets stuck to the detector). One gets, therefore, as a result of the interaction between the microscopic and the macroscopic system, a coherent superposition of two classically distinct states of the macroscopic apparatus. This is actually the situation in Schrödinger's cat paradox: the cat can be viewed as a measuring apparatus of the state of a decaying atom, the state of life or death of the cat being equivalent to the two positions of the pointer. This would imply that one should be able in principle to get interference between the two states of the pointer: it is precisely the lack of evidence of such phenomena in the macroscopic world that motivated Einstein's concern.

Faced with this problem, Von Neumann introduced through his collapse postulate [10] two distinct types of evolution in quantum mechanics: the deterministic and unitary evolution associated with the Schrödinger equation, which describes the establishment of a correlation between states of the microscopic system being measured and distinguishable classical states (for instance, distinct positions of a pointer) of the macroscopic measurement apparatus; and the probabilistic and irreversible process associated with measurement, which transforms coherent superpositions of distinguishable classical states into a statistical mixture. This separation of the whole process into two steps has been the object of much debate

[11–15]; indeed, it would not only imply an intrinsic limitation of quantum mechanics to deal with classical objects, but it would also pose the problem of drawing the line between the microscopic and the macroscopic world.

Several possibilities have been explored as solutions to this paradox, including the proposal that a small nonlinear term in the Schrödinger equation, although unnoticeable for microscopic phenomena, could eliminate the coherence between macroscopic states, thus transforming the quantum superpositions into statistical mixtures [12, 13]. The nonobservability of the coherence between the two positions of the pointer has been attributed both to the lack of nonlocal observables with matrix elements between the two corresponding states [16] as well as to the fast decoherence due to dissipation [17–26]. This last approach has been emphasized in recent years: decoherence follows from the irreversible coupling of the observed system to a reservoir. In this process, the quantum superposition is turned into a statistical mixture, for which all the information on the system can be described in classical terms, so our usual perception of the world is recovered. Furthermore, for macroscopic superpositions, quantum coherence decays much faster than the macroscopic observables of the system, its decay time being given by the dissipation time divided by a dimensionless number measuring the "separation" between the two parts. The statement that these two parts are macroscopically separated implies that this separation is an extremely large number. Such is the case for biological systems like "cats" made of a huge number of molecules. In the simple case mentioned by Einstein [8] of a particle split into two spatially separated wave packets by a distance d, the dimensionless measure of the separation is $(d/\lambda_{dB})^2$, where λ_{dB} is the particle de Broglie wavelength [6]. For a particle with mass equal to 1 g at a temperature of 300 K, and $d = 1$ cm, this number is about 10^{40}, and the decoherence is for all purposes instantaneous. This would provide an answer to Einstein's concern: decoherence of macroscopic states would be too fast to be observed.

Careful consideration of the interaction between the macroscopic system and the reservoir, and especially of its local character, leads to the understanding of a puzzling aspect of decoherence: why, among the huge variety of states in Hilbert space, are some states preferred in the classical limit? Indeed, localized states of macroscopic objects tend to be much more stable than their coherent superpositions. One knows now that the form of the interaction with the environment plays an essential role in inducing the selection of a set of states that are more resistant to decoherence [7]. These states have been called "pointer states" by Zurek [7]. One may say that the environment induces a superselection rule ("environment induced selection" or "ein-selection," according to Zurek), which inhibits coherent superpositions of pointer states.

In this article, it will be shown that the study of the interaction between atoms and electromagnetic fields in cavities can help us understand some aspects of this problem. In fact, many recent contributions in the field of quantum optics have led not only to the investigation of the subtle frontier between the quantum and the classical world, but also of hitherto unsuspected quantum mechanical processes like teleportation [27–37]. Research on quantum optics is therefore intimately entangled with fundamental problems of quantum mechanics.

It is also of interest to the area of quantum information, for several reasons. First, the individual control of atoms and photons attained in the area known as "cavity quantum electrodynamics" (cavity QED) has led to the demonstration of elementary quantum gates, which are the building blocks of quantum computation. Second, as mentioned before, the phenomenon of decoherence, which has been the subject of detailed studies in cavity QED, plays a central role in the quest for a quantum computer, as quantum software involves entangled states of many individual subsystems (many entangled "qubits"). These are true quantum superpositions of macroscopic states and therefore highly sensitive to decoherence: if special care is not taken, and the relevant states are not protected from decoherence, the quantum computer becomes a classical one within a very short time!

The whole area of cavity quantum electrodynamics is a very recent one. It concerns the interactions between atoms and discrete modes of the electromagnetic field in a cavity, under conditions such that losses due to dissipation and atomic spontaneous emission are very

small. Usually, one deals with atomic beams crossing cavities with a high quality factor Q (defined as the product of the angular frequency of the mode and its lifetime, $Q = \omega\tau$). Experiments have been realized both in the optical [38] and in the microwave domain [39–44]. In the strong-coupling regime, attained in the experiments realized in the past 10 years, the characteristic coupling frequency is much larger than the atomic and field decay rates and also the inverse of the interaction time between the atom and the cavity field.

Several factors contributed to the development of this area, among them the production of very good cavities and the development of techniques for controlling atomic beams. In the microwave region, the production of superconducting Niobium cavities, with extremely high quality factors, up to the order 10^{10}, allows one to keep a photon in the cavity for a time of the order up to 1 s. New techniques of atomic excitation [45, 46] (alkaline atoms, like rubidium and cesium, are frequently used for this purpose) to highly excited levels (principal quantum numbers of the order 50)—the so-called Rydberg atoms—with maximum angular momentum ($\ell = n - 1$) (so that the electrons describe circular orbits, a configuration known as "planetary" or "circular" atom) have led to the production of atomic beams that interact strongly even with very weak fields in the microwave domain, of the order one photon, due to the large magnitude of the relevant electric dipoles. Besides, the lifetime of these states is large—of the order 30 ms for $n = 50$—which may be understood semiclassically, from the correspondence principle (which should be valid for $n \sim 50$): the electron is always very far away from the nucleus, and therefore its acceleration is small, implying weak radiation and a long lifetime. One should also mention the new techniques of atomic velocity control, which allow the production of approximately monokinetic atomic beams, leading to a precise control of the interaction time between atom and field.

Although in the optical domain open geometries have been adopted for the cavities, made with two highly polished mirrors facing each other, both open [39, 40, 43] and closed [41, 42] geometries have been adopted in microwave experiments. The closed geometry, typically a cylinder with holes on its bases that allow the injection of atoms, has the advantage of increasing the quality factor of the cavity, as the leakage of photons is thus reduced. It does not allow, however, the injection of atoms in planetary states, as these are very unstable. This instability is due to the high degeneracy of these states: for $\ell = n - 1$, there are $2\ell + 1 = 2n - 1$ states with the same energy, corresponding to different directions in space of the electron orbit. Rydberg states with small values of the orbital angular momentum have substantially smaller lifetimes, of the order 200 to 400 μs in recent experiments. On the other hand, the open geometry, though leading to lower quality factors, allows the removal of this degeneracy, by submitting the atom to an electric field in the space between the two mirrors. In this way, planetary atomic states, with larger lifetimes, can be used.

In the following two sections, some of the basic properties of the quantized electromagnetic field and its interaction with two-level atoms are reviewed. Sections 4 and 5 apply this formalism respectively to the description of optical lattices and the trapping of a single atom by a single photon. Section 6 shows that two-level atoms interacting with photons in cavities may be used to demonstrate basic building blocks of quantum computation, as well as subtle quantum effects, like teleportation. Section 7 analyzes recent experiments that led to the monitoring in real-time of the decoherence process and of the dynamics of the quantum-classical transition. The quantum nondemolition measurement of a field in a cavity is discussed in Section 8, and the full characterization of the state of the system, through the measurement of the corresponding Wigner function, is discussed in Section 9. Section 10 reviews a technique, recently demonstrated experimentally, for directly determining the Wigner function of an electromagnetic field in a cavity. The measurement of the corresponding characteristic function and of the Husimi function are also discussed in that section, and the effect of dissipation is treated in Section 11. Section 12 shows how entanglement between two cavity modes can be characterizec through the Wigner function. Several methods to control decoherence are reviewed in Section 13. The use of the Wigner function to describe the quantum-classical transition for classically chaotic systems is discussed in Section 14. The conclusions are summarized in Section 15.

2. THE QUANTIZED ELECTROMAGNETIC FIELD

The free-field Hamiltonian for a mode of the electromagnetic field is given by the harmonic oscillator expression [16]

$$\hat{H} = \hbar\omega\left(\hat{N} + \frac{1}{2}\right) \tag{2}$$

where $\hat{N} = \hat{a}^\dagger\hat{a}$ is the *number operator*, and $\hat{a}$ and $\hat{a}^\dagger$ satisfy the commutation relation

$$[\hat{a}, \hat{a}^\dagger] = 1 \tag{3}$$

The eigenstates of $\hat{H}$ are denoted by $|n\rangle$, and satisfy the equation

$$\hat{N}|n\rangle = n|n\rangle, \quad n = 0, 1, 2, \ldots \tag{4}$$

and the eigenenergies are given by $E_n = (n + 1/2)\hbar\omega$. It is easy to show that

$$\hat{a}^\dagger|n\rangle = \sqrt{n+1}|n+1\rangle, \qquad \hat{a}|n\rangle = \sqrt{n}|n-1\rangle \tag{5}$$

The eigenvalue n of the number operator $\hat{N}$ is interpreted as the number of photons in the field, while, in view of Eq. (5), $\hat{a}$ and $\hat{a}^\dagger$ are the photon *annihilation* and *creation* operators.

The states $|n\rangle$ are the so-called *Fock states*, and have a well-defined number of photons. The state corresponding to $n = 0$ is the *vacuum state*. It is easy to show from the above relations that

$$|n\rangle = \frac{\hat{a}^{\dagger n}}{\sqrt{n!}}|0\rangle$$

The electric field is expressed in terms of the annihilation and creation operators by

$$\vec{E}(r) = E_\omega[\hat{a}u(\vec{r})\vec{\epsilon} + \hat{a}^\dagger u^*(\vec{r})\vec{\epsilon}^*] \tag{6}$$

where $u(\vec{r})$ is a function that describes the spatial dependence of the field mode, $\vec{\epsilon}$ is the polarization vector, and $E_\omega = \sqrt{\hbar\omega/V}$ is the *field per photon*. Here $V = \int |u(\vec{r})|^2 d^3r$ is the effective volume of the mode, defined so that the expectation value of the electromagnetic energy in the vacuum state, $(1/4\pi)\int\langle 0|[\vec{E}(\vec{r})]^2|0\rangle d^3r$ (in CGS units), is equal to the zero-point energy $\hbar\omega/2$. Typically, the mode function $u(\vec{r})$ is a standing wave along the axis of the cavity, and a Gaussian $\exp(-r^2/w^2)$ along the transverse direction, where w is the mode waist, of the order the micrometer in the optical case and the centimeter for microwave cavities. The volume of the mode is then given by $V = (\pi/4)Lw^2$, where L is the cavity length (distance between mirrors). In the microwave region, one has typically $V \approx 0.7$ cm^3, so that $E_\omega \approx 1.5$ mV/m. In the optical region, one may have $V \approx 10^3$ μm^3, so that $E_\omega \approx$ 150 V/cm. It is interesting to note that the much smaller transition dipole in the optical case (of the order 1 atomic unit versus 1000 atomic units for the Rydberg states) is compensated by a much larger electric field per photon, which is due to both the larger frequency and the smaller mode volume in the optical domain.

One should note that $\langle n|\hat{E}|n\rangle = 0$, that is, the average electric field is zero in a Fock state.

A special role will be played in the following by the *phase displacement operator*:

$$\hat{U}(\theta) = \exp(-i\theta\hat{N}) \tag{7}$$

It follows from the commutation relations that

$$\hat{U}^\dagger(\theta)\hat{a}\hat{U}(\theta) = \hat{a}\exp(-i\theta) \tag{8}$$

For $\theta = \omega t$, the phase displacement operator coincides, up to a factor $\exp(-i\omega t/2)$ coming from the zero-point energy, with the evolution operator corresponding to the Hamiltonian (2), and (8) yields the time evolution of the Heisenberg operator associated with $\hat{a}$.

2.1. Quadratures of the Electromagnetic Field

The quadratures of the electromagnetic field correspond to the position and momentum of a harmonic oscillator:

$$\hat{q} = \frac{1}{\sqrt{2}}(\hat{a} + \hat{a}^\dagger), \qquad \hat{p} = \frac{i}{\sqrt{2}}(\hat{a}^\dagger - \hat{a}), \qquad [\hat{q}, \hat{p}] = i \tag{9}$$

This commutation relation implies the Heisenberg inequality $\Delta q \Delta p \geq 1/2$.

From Eqs. (8) and (9), we see that, for $\theta = \pi$,

$$\hat{U}^\dagger(\pi)\hat{q}\hat{U}(\pi) = -\hat{q}, \qquad \hat{U}^\dagger(\pi)\hat{p}\hat{U}(\pi) = -\hat{p}$$

so that $\hat{U}(\pi)$ is the *parity operator*.

Setting $u(\vec{r}) = |u(\vec{r})|\exp[-i\phi(\vec{r})]$ in Eq. (6), we have, in the Heisenberg picture, for the electric field operator in terms of these quadratures (for a real polarization vector):

$$\vec{E}(r, t) = E_\omega |u(\vec{r})|\sqrt{2}[\hat{q}\cos(\omega t + \phi) - \hat{p}\sin(\omega t + \phi)]\vec{\epsilon}$$

This expression is analogous to the one that yields the position of a harmonic oscillator at time t in terms of its initial position and momentum:

$$x(t) = x(t_0)\cos\omega(t - t_0) + \frac{p(t_0)}{m\omega}\sin\omega(t - t_0) \tag{10}$$

The quadrature eigenstates (which correspond to states with well-defined position and momentum for the harmonic oscillator) will be denoted by

$$\hat{q}|q\rangle = q|q\rangle, \qquad \hat{p}|p\rangle = p|p\rangle$$

Exactly as for the position and momentum eigenstates, these states provide two non-normalizable bases. The corresponding quadrature wave functions are given by

$$\psi(q) = \langle q|\psi\rangle, \qquad \tilde{\psi}(p) = \langle p|\psi\rangle$$

where $|\psi\rangle$ is an arbitrary state of the field.

Using the phase-displacement operator given by Eq. (7), it is possible to define *generalized quadratures*:

$$\hat{q}_\theta = \hat{U}^\dagger(\theta)\hat{q}\hat{U}(\theta) = \left(\frac{1}{\sqrt{2}}\right)(\hat{a}e^{-i\theta} + \hat{a}^\dagger e^{i\theta}) = \hat{q}\cos\theta + \hat{p}\sin\theta \tag{11}$$

$$\hat{p}_\theta = \hat{U}^\dagger(\theta)\hat{p}\hat{U}(\theta) = -\hat{q}\sin\theta + \hat{p}\cos\theta \tag{12}$$

where of course $\hat{p}_\theta = \hat{q}_{\theta+\pi/2}$.

It is clear from these expressions that $\hat{U}(\theta)$ is the rotation operator in phase space. For a harmonic oscillator, and with $\theta = \omega t$, Eqs. (11) and (12) correspond respectively to the position and the momentum of the oscillator at time t, expressed in terms of the position $\hat{q}$ and momentum $\hat{p}$ at time $t = 0$.

2.2. Coherent States

We have seen that the average value of the electric field operator vanishes in a Fock state. Therefore, we cannot associate Fock states to classical fields with amplitude different from zero and well-defined phase. Let us look now for "quasi-classical" field states [47]. We require that the average value of the electromagnetic field in these states, which we denote by $|\alpha\rangle$, coincides with the classical expression for an electromagnetic field with complex amplitude α:

$$\langle\alpha|\hat{E}(\vec{r})|\alpha\rangle = \sqrt{\hbar\omega/V}\,u(\vec{r})\vec{\epsilon}\alpha + c.c. \tag{13}$$

From (6) and (13), it follows that

$$\langle\alpha|\hat{a}|\alpha\rangle = \alpha \tag{14}$$

We also require that the expectation value of the electromagnetic energy in the state $|\alpha\rangle$ coincides with the classical expression for this energy, expressed in terms of the complex amplitude α, at least in the limit when $|\alpha| \gg 1$. It is easy to show that the classical energy, when the electric field is expressed in terms of α like in (13), is given by $E_{\rm cl} = \hbar\omega|\alpha|^2$. Comparing this expression with (2), we must have therefore

$$\langle\alpha|\hat{a}^\dagger\hat{a}|\alpha\rangle = |\alpha|^2 \tag{15}$$

From (14) and (15), it follows that

$$\langle\alpha|(\hat{a}^\dagger - \alpha^*)(\hat{a} - \alpha)|\alpha\rangle = 0 \tag{16}$$

and therefore the state $|\alpha\rangle$ must be an eigenstate of the annihilation operator $\hat{a}$ with eigenvalue α:

$$\hat{a}|\alpha\rangle = \alpha|\alpha\rangle \tag{17}$$

These are the *coherent states* [48], which play an important role in quantum optics, and also in the understanding of the classical limit of quantum mechanics.

It is clear from the above discussion that the average number of photons in a coherent state $|\alpha\rangle$ is given by

$$\langle n\rangle = \langle\alpha|\hat{a}^\dagger\hat{a}|\alpha\rangle = |\alpha|^2$$

It also follows from the definition (17) and the commutation relations that, for a coherent state,

$$\Delta q = \Delta p = \frac{1}{\sqrt{2}}$$

or, more generally, $\Delta q_\theta = 1/\sqrt{2}$. Therefore, coherent states are minimum uncertainty states. This property can be pictorially depicted by drawing a circle in phase space, with a radius equal to the uncertainty in Δq_θ, as shown in Fig. 1

In terms of Fock states, coherent states can be expressed in the following way:

$$|\alpha\rangle = e^{-|\alpha|^2/2}\sum_{n=0}^{\infty}\frac{\alpha^n}{\sqrt{n!}}|n\rangle = e^{-|\alpha|^2/2}\sum_{n=0}^{\infty}\frac{(\alpha\hat{a}^\dagger)^n}{n!}|0\rangle \tag{18}$$

corresponding to the photon number distribution

$$p(n) = |\langle n|\alpha\rangle|^2 = \exp(-|\alpha|^2)\frac{|\alpha|^{2n}}{n!} = \exp(-\langle n\rangle)\frac{\langle n\rangle^n}{n!} \tag{19}$$

This is a Poisson distribution, with the variance $(\Delta n)^2 = \langle n\rangle$.

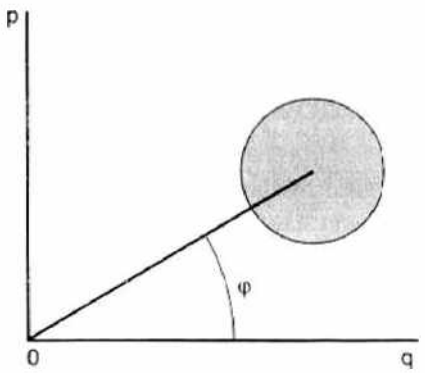


Figure 1. Pictorial representation of a coherent state in phase space.

The *displacement operator* is defined by

$$\hat{D}(\alpha, \alpha^*) = \exp(\alpha \hat{a}^\dagger - \alpha^* \hat{a}) \tag{20}$$

Note that the right-hand side of (18) implies that

$$|\alpha\rangle = e^{-|\alpha|^2/2} e^{\alpha \hat{a}^\dagger} |0\rangle = e^{-|\alpha|^2/2} e^{\alpha \hat{a}^\dagger} e^{-\alpha \hat{a}} |0\rangle = \exp(\alpha \hat{a}^\dagger - \alpha^* \hat{a})|0\rangle = \hat{D}(\alpha, \alpha^*)|0\rangle \tag{21}$$

where in the last step the Baker–Hausdorff transformation has been used to entangle the annihilation and creation operators in the exponent. Therefore, a coherent state $|\alpha\rangle$ can be obtained by applying the displacement operator $\hat{D}(\alpha, \alpha^*)$ to the vacuum state. The displacement operator is closely connected to the evolution operator corresponding to the interaction of the electromagnetic field with a classical current. Indeed, this interaction is described by the Hamiltonian

$$H_{\text{int}} = \int \vec{J} \cdot \vec{A} d^3 r$$

which can be written in the form $H_{\text{int}} = i(\alpha \hat{a}^\dagger - \alpha^* \hat{a})$. The evolution operator corresponding to this interaction coincides, up to a phase, with $D(\alpha, \alpha^*)$. Therefore, Eq. (21) implies that classical currents generate coherent states from the vacuum. In an analogous way, it is easy to show that for a harmonic oscillator, a coherent state can be physically realized by applying a classical force to the oscillator, initially in the ground state.

From the expansion of the coherent states in terms of the Fock states, one easily derives the following scalar-product and completeness relations:

$$|\langle \alpha | \alpha' \rangle|^2 = e^{-|\alpha - \alpha'|^2} \tag{22}$$

$$\frac{1}{\pi} \int d^2\alpha |\alpha\rangle\langle\alpha| = \mathbb{1} \tag{23}$$

where $d^2\alpha \equiv d(\Re e\alpha) d(\Im m\alpha)$.

Equation (22) implies that two different coherent states are not necessarily orthogonal. This is not surprising, as, according to Eq. (17), coherent states are eigenstates of a non-Hermitian operator. On the other hand, if $|\alpha - \alpha'| \gg 1$, then the coherent states $|\alpha\rangle$ and $|\alpha'\rangle$ will be approximately orthogonal. In view of this non-orthogonality, Eq. (23) allows one to express any coherent state in terms of the others: for this reason, coherent states are said to form an over-complete set.

In terms of the quadrature eigenstates $|q\rangle$, one may write:

$$\langle q | \alpha_0 \rangle = \left(\frac{1}{\pi}\right)^{1/4} e^{-iq_0 p_0/2} e^{ip_0 q} e^{-(q - q_0)^2/2}$$

with $\alpha_0 = (q_0 + ip_0)/\sqrt{2}$.

Therefore, the probability density of finding a quadrature $\hat{q}$ of the field with a value q, for a coherent state $|\alpha_0\rangle$, is given by a Gaussian (for the vacuum, $q_0 = p_0 = 0$):

$$P(q) = \left(\frac{1}{\pi}\right)^{1/2} \exp[-(q - q_0)^2]$$

This expression coincides with the probability density of measuring the value q for the position of a harmonic oscillator, when its ground state is displaced by q_0, as expected.

2.3. Measurement of Quadratures

Several methods have been proposed to measure quadratures of the electromagnetic field (for a review, see for instance Ref. [49]). The general idea consists in mixing the signal to be detected with an intense coherent signal, called *local oscillator*, before detection [50–52].

The discussion here is restricted to the method of *balanced homodyne detection*, sketched in Fig. 2. The field to be measured (complex amplitude E_a) is sent on a beam splitter, together with a coherent field (complex amplitude E_b) with the same frequency. One measures then the difference of intensity of the two beams emerging from the beam splitter (complex amplitudes E_c and E_d). The detection is said to be balanced when the mirror transmits 50% of the incident light.

Let r and t be the reflection and transmission coefficients of the mirror, respectively. Let us set:

$$E_c = rE_a - tE_b \tag{24}$$

$$E_d = tE_a + rE_b \tag{25}$$

or yet, in matrix form,

$$\begin{pmatrix} E_c \\ E_d \end{pmatrix} = \begin{pmatrix} r & -t \\ t & r \end{pmatrix} \begin{pmatrix} E_a \\ E_b \end{pmatrix} \tag{26}$$

Energy conservation (assuming that losses are negligible) implies that

$$|E_c|^2 + |E_d|^2 = |E_a|^2 + |E_b|^2 \tag{27}$$

From (26) and (27), one gets:

$$|r|^2 + |t|^2 = 1 \tag{28}$$

$$r^*t - rt^* = 0 \tag{29}$$

If one takes r real and equal to $\sqrt{\eta}$ (one should note that phases in r and t can be removed by redefining the phases of the incoming and outgoing fields), it follows from Eq. (28) that $t = \pm(1-\eta)^{1/2}$. Choosing the positive sign, one gets then

$$\begin{pmatrix} E_c \\ E_d \end{pmatrix} = \begin{pmatrix} \sqrt{\eta} & -\sqrt{1-\eta} \\ \sqrt{1-\eta} & \sqrt{\eta} \end{pmatrix} \begin{pmatrix} E_a \\ E_b \end{pmatrix} \tag{30}$$

Normalizing the intensity to the photon number, and introducing annihilation operators corresponding to the several fields involved through

$$E_a \to \hat{a}, \qquad E_b \to \hat{b}$$

$$E_c \to \hat{c}, \qquad E_d \to \hat{d}$$

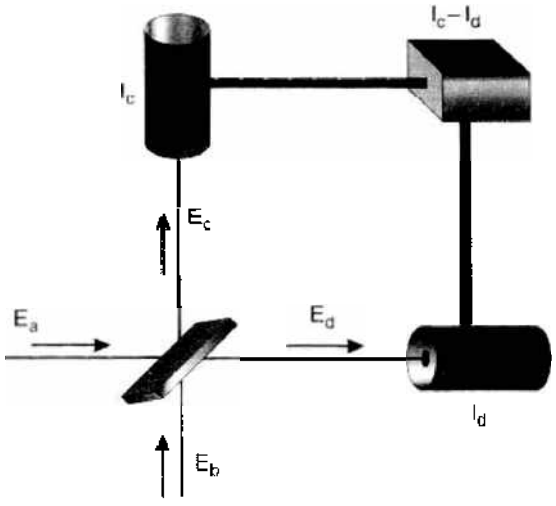


Figure 2. Method of balanced homodyne detection.

one gets, from (30),

$$\hat{c} = \sqrt{\eta}\hat{a} - \sqrt{1-\eta}\hat{b}, \qquad \hat{d} = \sqrt{1-\eta}\hat{a} + \sqrt{\eta}\hat{b} \tag{31}$$

For balanced detection, $\eta = 1/2$, so that

$$\hat{c} = \frac{1}{\sqrt{2}}(\hat{a} - \hat{b}), \qquad \hat{d} = \frac{1}{\sqrt{2}}(\hat{a} + \hat{b}) \tag{32}$$

These expressions relate the field operators after the beam splitter with those before this device. They should be understood therefore as Heisenberg operators, which evolve under the action of the beam splitter. Note that the conditions expressed by Eqs. (28) and (29) imply that the transformation between the field operators corresponding to Eq. (26) is unitary (this is the requirement for operators that corresponds to energy conservation for the classical fields).

The difference between the intensities of the fields E_d e E_c is given then by

$$I = \hat{d}^\dagger\hat{d} - \hat{c}^\dagger\hat{c} = \hat{a}^\dagger\hat{b} + \hat{b}^\dagger\hat{a} \tag{33}$$

Assuming that the field E_b may be described classically (this would be the case for a coherent state with large average photon number), one replaces $\hat{b}$ by $\beta = Be^{-i(\omega t - \theta)}$, so that (33) gets transformed into

$$I = B\left[\hat{a}e^{i(\omega t - \theta)} + \hat{a}^\dagger e^{-i(\omega t - \theta)}\right] \tag{34}$$

Because $\hat{a} = \hat{a}_0 e^{-i\omega t}$ (all fields are taken in the Heisenberg picture), one gets finally,

$$I = B(\hat{a}_0 e^{-i\theta} + \hat{a}_0^\dagger e^{i\theta}) \tag{35}$$

This equation shows that the difference of intensities, measured by the method of homodyne detection, is directly proportional to the quadrature $\hat{X}(\theta)$ of the field E_a, defined by

$$\hat{X}(\theta) = \frac{1}{\sqrt{2}}(\hat{a}_0 e^{-i\theta} + \hat{a}_0^\dagger e^{i\theta}) \tag{36}$$

Therefore, by detecting the difference of intensities, as the phase of the local oscillator E_b is changed, one may measure an arbitrary quadrature of the field E_a. For a stationary field, a sequence of measurements for the same phase yields the probability distribution for the corresponding quadrature, $P(X_\theta)$. In practice, one deals with a continuum of modes, and the above analysis applies to the situation when the frequency window of the detector is much smaller than the linewidth of the light that is being measured.

In order to keep the phase difference constant between the local oscillator and the field to be measured, in spite of the unavoidable phase diffusion of the local oscillator, one frequently uses part of the local oscillator beam to pump the medium that produces the field to be analyzed. In this way, the local oscillator is used both for analyzing and for providing a phase reference for the detected field.

3. THE ATOM-FIELD INTERACTION

3.1. The Interaction Hamiltonian

We consider now the interaction of atoms and fields. We will be considering situations in which the atom is resonant or quasi-resonant with one of the modes of the electromagnetic field in a cavity. Under these conditions, it is possible to consider just two of the atomic states and therefore reduce the atom to a two-level system [we will call e the upper level and g the lower level, and define their energy difference as $\hbar\omega_0$]. The basic Hamiltonian describing this system can then be expressed in terms of Pauli spin operators, if we let

$$|e\rangle \leftrightarrow \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad |g\rangle \leftrightarrow \begin{pmatrix} 0 \\ 1 \end{pmatrix} \tag{37}$$

Then,

$$\hat{H} = \hat{H}_A + \hat{H}_F + \hat{H}_{AF} \tag{38}$$

where

$$\hat{H}_A = \left(\frac{\hbar\omega_0}{2}\right)\hat{\sigma}_z \tag{39}$$

is the free-atom Hamiltonian, with $\hat{\sigma}_z$ a Pauli matrix:

$$\hat{\sigma}_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{40}$$

$$\hat{H}_F = \hbar\omega\left(\hat{a}^\dagger\hat{a} + \frac{1}{2}\right) \tag{41}$$

is the free-field Hamiltonian, and

$$\hat{H}_{AF} = \left(\frac{\hbar\Omega_0}{2}\right)(\hat{\sigma}_+\hat{a} + \hat{\sigma}_-\hat{a}^\dagger) \tag{42}$$

with

$$\hat{\sigma}_+ = \frac{\hat{\sigma}_x + i\hat{\sigma}_y}{2} = (\hat{\sigma}_-)^\dagger = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \tag{43}$$

The operator $\hat{\sigma}_+$ applied to $|g\rangle$ yields the state $|e\rangle$. The interaction in Eq. (42) is thus easily interpreted: the term $\hat{\sigma}_+\hat{a}$ describes the excitation of the atom from state $|g\rangle$ to state $|e\rangle$, through the absorption of a photon, whereas the term $\hat{\sigma}_-\hat{a}^\dagger$ describes the inverse process, with the atom emitting a photon while it gets deexcited from $|e\rangle$ to $|g\rangle$.

In Eq. (42), we have neglected terms of the form $\hat{\sigma}_+\hat{a}^\dagger$ and $\hat{\sigma}_-a$, which do not conserve energy in first order and which lead to small corrections in the results to be obtained, as long as $|\omega - \omega_0| \ll \omega_0$ and $\Omega_0\sqrt{\langle n\rangle} \ll \omega, \omega_0$, where $\langle n\rangle$ is the average number of photons in the field. We have also adopted the dipole approximation, neglecting the variation of the electromagnetic field within the atom.

The coupling constant $\hbar\Omega_0/2$ depends on the transition dipole $\vec{d}_{eg}$ between the two levels, on the polarization vector $\vec{\epsilon}$ and the frequency ω of the electromagnetic field, as well as the effective volume of the mode V. From Eq. (6), it follows that

$$\frac{\hbar\Omega_0}{2} = -\vec{d}_{eg}\cdot\vec{\epsilon}\sqrt{\frac{\hbar\omega}{V}}u(\vec{R}) \tag{44}$$

where the mode function $u(\vec{R})$ is evaluated on the center-of-mass position $\vec{R}$ of the atom interacting with the field (this is a consequence of the dipole approximation). One should note that, as only two atomic states are involved, one may always choose their phases so that Ω_0 is real and positive. If $u(\vec{R})$ is real, then this choice will not depend on $\vec{R}$. The frequency Ω_0 is the resonant *vacuum Rabi frequency*.

The above equations define the *Jaynes–Cummings model* [53], a very useful model in quantum optics, which has successfully described many experiments in cavity QED. One should note that this model neglects dissipative processes, which will be considered in a while. The most interesting regime, from the point of view of demonstrating quantum features of the system, corresponds to the so-called strong-coupling regime, when the frequency $\Omega_0 \gg \kappa, \gamma, 1/t_{\text{int}}$, where κ and γ are respectively the atomic and field damping rates, and t_{int} is the interaction time between the atom and the field in the cavity.

3.2. Semiclassical Approximation and Bloch Equations

We note that the above Hamiltonian can be written as

$$\hat{H} = \hbar\omega\left(\hat{a}^\dagger\hat{a} + \frac{1}{2} + \frac{1}{2}\hat{\sigma}_z\right) + \left(\frac{\hbar\delta}{2}\right)\hat{\sigma}_z + \left(\frac{\hbar\Omega_0}{2}\right)(\hat{\sigma}_+\hat{a} + \hat{\sigma}_-\hat{a}^\dagger) \tag{45}$$

where $\delta = \omega_0 - \omega$ is the detuning between the two-level atom and the cavity mode.

We transform now the above Hamiltonian to the interaction picture with respect to

$$\hat{H}_0 = \hbar\omega\left(\hat{a}^\dagger a + \frac{1}{2} + \frac{1}{2}\hat{\sigma}_z\right) \tag{46}$$

This transformation eliminates the free evolution of the field and at the same time rotates the components $\hat{\sigma}_x, \hat{\sigma}_y$ of the pseudo-spin associated with the two-level atom around the z axis with the field frequency ω. That is why it is called a "rotating-frame transformation." Indeed,

$$\exp\left(\frac{iH_0 t}{\hbar}\right)\hat{a}\exp\left(\frac{-iH_0 t}{\hbar}\right) = \hat{a}\exp(-i\omega t) \equiv \hat{a}_I \tag{47}$$

$$\exp\left(\frac{iH_0 t}{\hbar}\right)\hat{\sigma}\ \exp\left(\frac{-iH_0 t}{\hbar}\right) = \hat{\sigma}\ \exp(-i\omega t) = \hat{\sigma}_{-,I} \tag{48}$$

where the subscript I stands for interaction picture.

We get then, for the corresponding interaction-picture Hamiltonian,

$$\hat{H}_I = \left(\frac{\hbar\delta}{2}\right)\hat{\sigma}_z + \left(\frac{\hbar\Omega_0}{2}\right)(\hat{\sigma}_{+,I}\hat{a}_I + \hat{\sigma}_{-,I}\hat{a}_I^\dagger) \tag{49}$$

The semiclassical approximation corresponds to setting $\hat{a}_I \to \alpha$, where α stands for the complex slowly varying envelope of the field [that is, with the factor $\exp(-i\omega t)$ taken out], which now becomes an external classical field. Both its quantum character and the reaction of the atom on the field are neglected in this approximation. It is usually assumed that this should hold for sufficiently intense fields, described by coherent states. This is not necessarily so, however: this approximation can be shown to be valid even for coherent states with an average number of photons of the order one, as long as the dampening of the field in the cavity is very strong (low Q cavity), and the cavity is pumped by a classical source [54]. Of course, in this case both the pumping and the losses should be included in the model. This is the situation in recent experiments held at Ecole Normale Supérieure, in Paris [55], where indeed the number of photons in the cavity mode that interacts with the two-level atom is of the order one, and yet the field can be treated classically.

With this approximation, the semiclassical Hamiltonian becomes:

$$\hat{H}_I = \frac{\hbar\delta}{2}\hat{\sigma}_z + \frac{\hbar\Omega_0}{2}(\hat{\sigma}_{+,I}\alpha + \hat{\sigma}_{-,I}\alpha^*) = \frac{\hbar}{2}\vec{\hat{\sigma}}_I \cdot \vec{\Omega} \tag{50}$$

where

$$\vec{\Omega} \equiv \Omega\hat{n} = (V_1, V_2, \delta) \tag{51}$$

the components V_1 and V_2 are defined by

$$\Omega_0\alpha \equiv V \equiv V_1 - iV_2 \tag{52}$$

and $\hat{n}$ is a unit vector.

We note that $\hat{H}_I$ corresponds to the precession of a pseudo-spin around a pseudo-magnetic field $\vec{\Omega}$. The corresponding evolution operator is given by

$$U_{\vec{\Omega}} = \exp\left(-i\vec{\hat{\sigma}}_I \cdot \frac{\hat{n}\Omega t}{2}\right) = \begin{bmatrix} \cos\left(\frac{\Omega t}{2}\right) - in_z\sin\left(\frac{\Omega t}{2}\right) & (-in_x - n_y)\sin\left(\frac{\Omega t}{2}\right) \\ (-in_x + n_y)\sin\left(\frac{\Omega t}{2}\right) & \cos\left(\frac{\Omega t}{2}\right) + in_z\sin\left(\frac{\Omega t}{2}\right) \end{bmatrix} \tag{53}$$

where we have replaced the Pauli operators in the interaction picture by the same matrices as before (this is allowed because their commutation relations are the same as before—of course the corresponding basis will differ from the one defined before by the product with oscillating exponentials, which stem from the transformation of the previous basis to the interaction picture).

Setting

$$r_1 \equiv \langle \hat{\sigma}_{x,I} \rangle, \qquad r_2 \equiv \langle \hat{\sigma}_{y,I} \rangle, \qquad r_3 \equiv \langle \hat{\sigma}_{z,I} \rangle \tag{54}$$

we get, from Eq. (50), via the Heisenberg equations of motion for the atomic operators, the expected result:

$$\frac{d\vec{r}}{dt} = \vec{\Omega} \times \vec{r} \tag{55}$$

This is the Bloch equation (except for the absence of phenomenological decay terms): it represents the atomic state by a pseudo-spin $\vec{r}$ (Bloch vector), which precesses around the pseudo-magnetic field $\vec{\Omega}$, as shown in Fig. 3. The precession frequency, given by

$$\Omega = \sqrt{|V|^2 + \delta^2} \tag{56}$$

is the *Rabi frequency*.

One should note that, in terms of the atomic density matrix, in the same interaction picture considered before,

$$\hat{\rho}_A^I = \begin{pmatrix} \rho_{ee}^I & \rho_{eg}^I \\ \rho_{ge}^I & \rho_{gg}^I \end{pmatrix}$$

one may write:

$$r_1 = \mathrm{Tr}[\hat{\sigma}_x \hat{\rho}_A^I(t)] = \rho_{eg}^I + \rho_{ge}^I = 2\Re e(\rho_{ge}^I) \tag{57}$$

$$r_2 = \mathrm{Tr}[\hat{\sigma}_y \hat{\rho}_A^I(t)] = i(\rho_{eg}^I - \rho_{ge}^I) = 2\Im m(\rho_{ge}^I) \tag{58}$$

$$r_3 = \mathrm{Tr}[\hat{\sigma}_z \hat{\rho}_A^I(t)] = \rho_{ee}^I - \rho_{gg}^I \tag{59}$$

and therefore the components of the Bloch vector are easily expressed in terms of the matrix elements of the atomic density operator in the rotating frame.

The third component of the Bloch vector represents the atomic population, whereas the equatorial projection is associated with the atomic polarization ($\vec{P}$ in Fig. 3). This picture of the evolution of a two-level atom is due to Feynman, Vernon, and Hellwarth [56].

For a pure state, $|\psi\rangle = c_e|e\rangle + c_g|g\rangle$, one has

$$\hat{\rho}_A = \begin{pmatrix} |c_e|^2 & c_e^* c_g \\ c_e c_g^* & |c_g|^2 \end{pmatrix}, \qquad r_1^2 + r_2^2 + r_3^2 = 1$$

Therefore, in this case, the tip of the Bloch vector is situated on a sphere of unit radius. In general, $r_1^2 + r_2^2 + r_3^2 \leq 1$, the equality being valid if and only if the atomic state is pure.

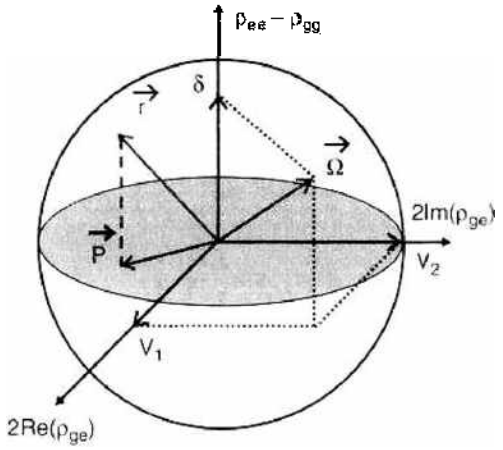


Figure 3. The Bloch vector $\vec{r}$ precesses around the pseudo-magnetic field $\vec{\Omega}$.

Two limiting cases of this expression correspond to the resonant and to the dispersive interaction. When the interaction is resonant, $\delta = 0$, $\vec{\Omega} = (V_1, V_2, 0)$, and the Bloch vector precesses around a vector in the equatorial plane. One gets in this case maximum population transfer. The precession frequency is then $|V|$.

In particular, it follows from Eq. (53) that a sequence of $\pi/2$ rotations for $\vec{\Omega} = (\pi/2)(-\sin\phi, \cos\phi)$ would yield, if one starts with the atom in state $|e\rangle$:

$$|e\rangle \rightarrow \frac{1}{\sqrt{2}}(|e\rangle + e^{i\phi}|g\rangle) \rightarrow e^{i\phi}|g\rangle \rightarrow \frac{e^{i\phi}}{\sqrt{2}}(-e^{-i\phi}|e\rangle + |g\rangle) \rightarrow -|e\rangle \tag{60}$$

so that after a 2π rotation, one recovers the initial state, except for a minus sign, characteristic of the behavior of a spin 1/2 system.

In the dispersive limit, $|\delta| \gg |V|$, and $\vec{\Omega} \rightarrow (0, 0, \delta)$. The Bloch vector precesses then around a vector parallel to the axis 3 (population axis), with a frequency equal to δ, as shown in Fig. 4. Therefore, the atomic population does not change.

3.3. Quantum Theory: The Dressed Atom

We go back now to the Hamiltonian given by Eq. (38) and consider the effects resulting from the quantization of the electromagnetic field.

The eigenstates of the Hamiltonian in Eq. (38) define the *dressed atom* [57–59]. Whereas H_A has two energy levels, H_F has an infinite number of discrete levels, given by $\hbar\omega(n + 1/2)$, $n = 0, 1, 2, \ldots$. The interaction H_{AF} couples these levels, leading to a discrete structure of levels of the composed system, which one could call the "atom-field molecule." We study first the structure of the uncoupled system, and then we analyze the energy levels of the coupled system.

3.3.1. Uncoupled States

The state corresponding to the atom in state e and n photons in the field will be denoted by $|e, n\rangle$; analogously for $|g, n\rangle$. Let again $\delta = \omega_0 - \omega$ be the detuning between the atom and the field. If $|\delta| \ll \omega_0$, the energies of the states $|e, n\rangle$ and $|g, n + 1\rangle$ will be very close to each other. If $\delta > 0$, the energy of the state $|g, n + 1\rangle$ will be smaller than the energy of the state $|e, n\rangle$ (Fig. 5). In fact, we can write

$$\begin{aligned} E_{e,n} &= \hbar\omega(n + 1) + \frac{\hbar\delta}{2} \\ E_{g,n+1} &= \hbar\omega(n + 1) - \frac{\hbar\delta}{2} \end{aligned} \tag{61}$$

We have therefore a sequence of quasi-degenerate subspaces $\varepsilon(n) \equiv \{|g, n + 1\rangle; |e, n\rangle\}$. Note that $E_{e,0} = \hbar(\omega + \omega_0)/2$, and $E_{g,0} = -\hbar\delta/2 = \hbar(\omega - \omega_0)/2$, consistently with the fact that the zero-point energy is $\hbar\omega/2$ and the energies of the two atomic states are $\pm\hbar\omega_0/2$.

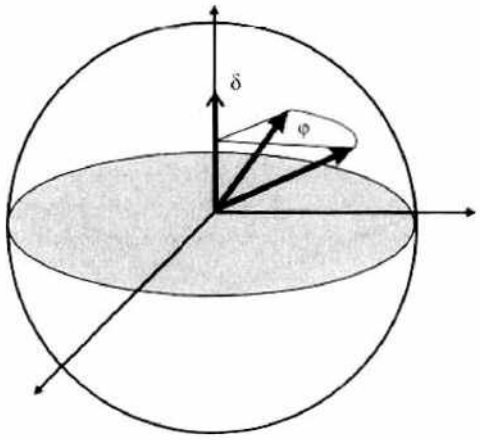


Figure 4. Dispersive limit: the Bloch vector precesses around the vertical axis.

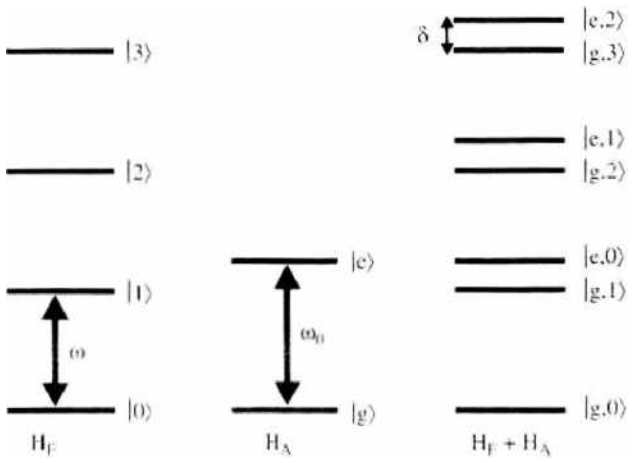


Figure 5. Energy level diagram for the uncoupled atom-field system. Here $\delta = \omega_0 - \omega$ is taken to be positive.

Figure 6 shows the energy diagram for the uncoupled states as a function of the detuning $\delta = \omega_0 - \omega$. The two straight lines correspond to the energy levels given by Eq. (61). There is a crossing for $\delta = 0$, when the two states $|e, n\rangle$ and $|g, n+1\rangle$ become degenerate.

3.3.2. Coupled States

The Hamiltonian in Eq. (38), with H_{AI} given by Eq. (42), couples only states within the same subspace (this is a consequence of the rotating-wave approximation; the counter-rotating terms, neglected in H_{AF}, connect states belonging to different subspaces, which leads to small corrections to the results considered here, due to the large energy differences involved). Therefore, in order to calculate the eigenvalues of the complete Hamiltonian, one has to diagonalize a 2×2 matrix, given in the subspace $\mathcal{E}(n)$ by:

$$[H]_n = \hbar \begin{pmatrix} \omega(n+1) - \dfrac{\delta}{2} & \dfrac{\Omega_0}{2}\sqrt{n+1} \\ \dfrac{\Omega_0}{2}\sqrt{n+1} & \omega(n+1) + \dfrac{\delta}{2} \end{pmatrix} \tag{62}$$

The eigenvalues of this matrix define the energy levels of the dressed atom:

$$\begin{aligned} E_{\pm,n} &= (n+1)\hbar\omega \pm \left(\frac{\hbar\Omega}{2}\right) \\ E_{g,0} &= -\frac{\hbar\delta}{2} \end{aligned} \tag{63}$$

where

$$\Omega = [\Omega_0^2(n+1) + \delta^2]^{1/2} \tag{64}$$

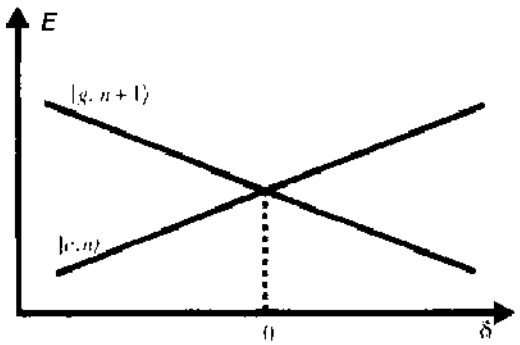


Figure 6. Energies as a function of the detuning δ for the uncoupled states $|e, n\rangle$ and $|g, n+1\rangle$.

is the quantum *Rabi frequency* of the system, whereas Ω_0 as defined before is the resonant vacuum Rabi frequency, which coincides with Ω when $n = 0$ and $\delta = 0$. The quantum Rabi frequency given by Eq. (64) coincides precisely with the classical expression in Eq. (56) if one identifies $\Omega_0\sqrt{n+1}$ with $|V|$. One should note, however, that contrary to the classical expression, the quantum Rabi frequency remains different from zero even when the number of photons in the mode is equal to zero. The remaining contribution is associated with spontaneous emission into the mode, which couples the state $|e, 0\rangle$ to the state $|g, 1\rangle$.

Equation (63) shows that the two states are separated by the coupling, the energy difference between them going from $\hbar\delta$ to $\hbar\Omega$. This effect is displayed in Fig. 7.

One should note that the coupling Ω_0 depends on the position of the atom, as the field amplitude depends on the mode profile. For a Gaussian profile, the interaction is stronger at the mode center and vanishes as the atom approaches the edges of the cavity. Figure 8 displays the lower energy levels of the dressed atom as functions of the radial distance with respect to the center of the mode.

The corresponding eigenstates are given by:

$$|+, n\rangle = \sin\theta|g, n+1\rangle + \cos\theta|e, n\rangle \tag{65}$$

$$|-, n\rangle = \cos\theta|g, n+1\rangle - \sin\theta|e, n\rangle \tag{66}$$

with

$$\cot 2\theta = \frac{\delta}{\Omega_0\sqrt{n+1}}, \qquad 0 \le 2\theta < \pi \tag{67}$$

3.3.3. Resonant Interaction

At resonance, $\delta = 0$, and therefore $\theta = \pi/4$, so that

$$|+, n\rangle = \frac{1}{\sqrt{2}}(|g, n+1\rangle + |e, n\rangle) \tag{68}$$

$$|-, n\rangle = \frac{1}{\sqrt{2}}(|g, n+1\rangle - |e, n\rangle) \tag{69}$$

In this case, each subspace $\mathcal{E}(n)$ becomes two-fold degenerate, and the dressed states are expressed in terms of the sum and the difference of the corresponding uncoupled states, with equal weights. The corresponding energies are, for $n \neq 0$,

$$E_{\pm,n} = (n+1)\hbar\omega \pm (\hbar\Omega_0\sqrt{n+1}/2) \tag{70}$$

3.3.4. Dispersive Interaction

For large detuning ($|\delta| \gg \Omega_0\sqrt{n+1}$), one gets $\theta \to \pi/2$ if $\delta < 0$, and $\theta \to 0$ if $\delta > 0$, and therefore

$$|+, n\rangle \to |g, n+1\rangle, \qquad |-, n\rangle \to |e, n\rangle, \qquad (\delta < 0) \tag{71}$$

$$|+-n\rangle \to |e, n\rangle, \qquad |-, n\rangle \to |g, n+1\rangle, \qquad (\delta > 0) \tag{72}$$

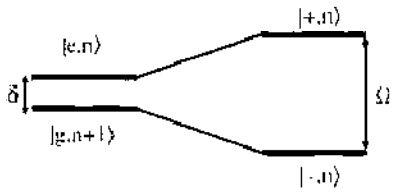


Figure 7. Energy displacement of levels, produced by the coupling.

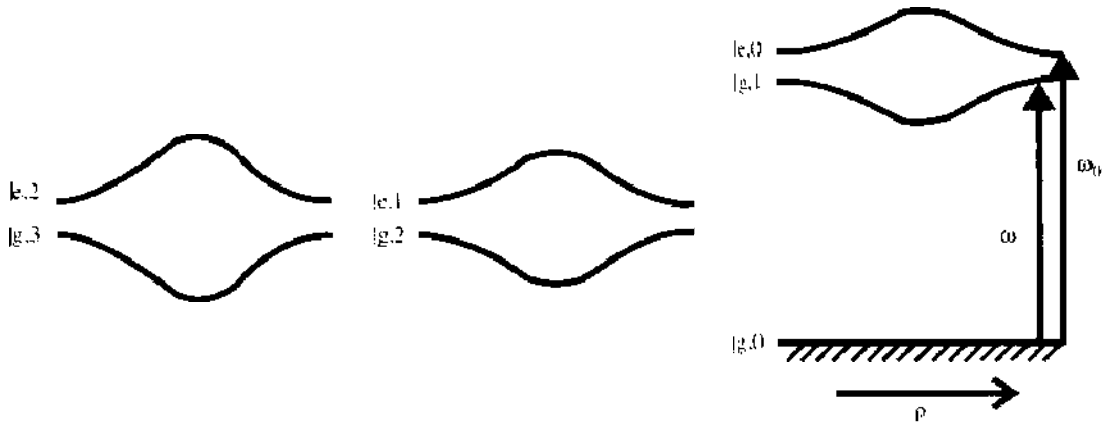


Figure 8. Lower energy levels of the dressed atom as functions of the radial distance with respect to the center of the mode.

These equations show that, for a dispersive interaction, the coupled states approach the uncoupled states, with an energy shift obtained from Eq. (63):

$$E_{+,n} \approx (n+1)\hbar\omega \pm \frac{\hbar|\delta|}{2} \pm \hbar\Omega_0^2(n+1)/4|\delta| \tag{73}$$

In any case, we have in this limit:

$$\Delta E_{e,n} \approx \hbar\left(\frac{\Omega_0^2}{4\delta}\right)(n+1) \tag{74}$$

$$\Delta E_{g,n} \approx -\hbar\left(\frac{\Omega_0^2}{4\delta}\right)n \tag{75}$$

The two energy levels of each subspace get displaced in opposite directions. These displacements coincide precisely with those that would be obtained using second-order perturbation theory, and constitute the *AC Stark effect*.

These considerations are summarized in Fig. 9, which displays the energies given by Eqs. (63) and (64) as functions of the detuning δ. The interaction transforms the crossing exhibited in Fig. 6 into an anti-crossing.

3.4. Dynamics of the Interaction

Once the Hamiltonian is diagonalized, one can easily describe the dynamical behavior of the system. From Eq. (65), one has:

$$|e,n\rangle = \cos\theta|+,n\rangle - \sin\theta|-,n\rangle \tag{76}$$

$$|g,n+1\rangle = \sin\theta|+,n\rangle + \cos\theta|-,n\rangle \tag{77}$$

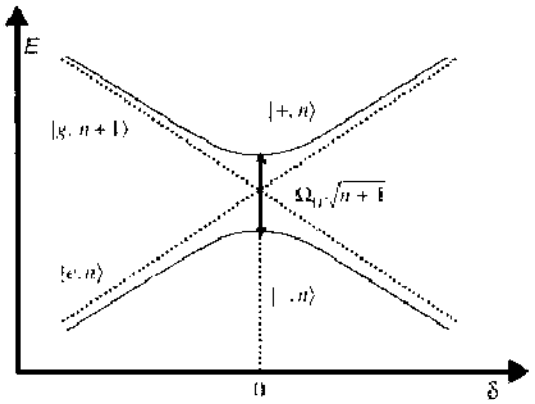


Figure 9. Energy levels as functions of the detuning $\delta = \omega_0 - \omega$, for the dressed states.

and therefore, if the initial state of the system is $|\psi(0)\rangle = |e, n\rangle$, we have at time t:

$$|\psi(t)\rangle = \cos\theta e^{-i(n+1/2)\omega t}|+, n\rangle - \sin\theta e^{-i(n-1/2)\omega t}|-, n\rangle \tag{78}$$

or yet, reexpressing in terms of the uncoupled states:

$$|\psi(t)\rangle = e^{-i\omega(n-1)t}\left\{-i\sin(2\theta)\sin\left(\frac{\Omega t}{2}\right)|g, n+1\rangle + \left[\cos\left(\frac{\Omega t}{2}\right) - i\cos(2\theta)\sin\left(\frac{\Omega t}{2}\right)\right]|e, n\rangle\right\} \tag{79}$$

which is an atom-field entangled state. The probabilities of finding the system in the states $|e, n\rangle$ e $|g, n+1\rangle$ are thus given by

$$P_{e,n} = \cos^2\left(\frac{\Omega t}{2}\right) + \cos^2(2\theta)\sin^2\left(\frac{\Omega t}{2}\right) \tag{80}$$

$$P_{g,n+1} = \sin^2(2\theta)\sin^2\left(\frac{\Omega t}{2}\right) \tag{81}$$

oscillating therefore with the Rabi frequency Ω. At resonance, when $\theta = \pi/4$, one gets $P_{e,n} = \cos^2(\Omega t/2)$, $P_{g,n+1} = \sin^2(\Omega t/2)$, so the oscillation has maximum amplitude.

These considerations extend to the quantum case the description of the atomic evolution in terms of the Bloch vector, previously discussed within the semiclassical approximation. One should note that, if one starts from the state $|e, n\rangle$, the quantum system evolves even when the number of photons in the mode is equal to zero, contrary to what would happen if the field is not quantized. This extra quantum feature is again due to spontaneous emission by the atomic excited state into the cavity mode.

On the other hand, when $n \gg 1$, the above expressions for the probabilities coincide with those obtained from the semiclassical theory, if one identifies the amplitude α of the classical field with $\sqrt{n}$. One should note, however, that Eq. (79) displays an entanglement between atom and field states, which could never be obtained from a semiclassical theory.

One should also note that, in the dispersive limit, neither the number of photons nor the populations of states e and g change, exactly as in the semiclassical treatment. In this case, the dynamics of the atom is well represented by the precession of the Bloch vector around the vertical axis (population axis): if there are n photons in the field, the angle of precession is given by $(\Delta E_{e,n} - \Delta E_{g,n})t/\hbar$, with the Stark energy displacements given by Eqs. (74) and (75).

3.5. Single-Atom Refraction Index

According to Eq. (74), if the two-level atom is in state $|e\rangle$ and the field in the cavity is in a coherent state with amplitude α (and frequency ω), one should have, if the atom interacts dispersively with the field:

$$|e\rangle|\alpha e^{-i\omega t}\rangle \rightarrow |e\rangle \exp\left(-\frac{|\alpha|^2}{2}\right)\sum_{n=0}^{\infty}\frac{\alpha^n}{\sqrt{n!}}e^{-in\omega t}e^{i(n+1)\Omega_0^2 t/4\delta}|n\rangle = e^{i\Delta\omega t}|e\rangle|\alpha e^{-i(\omega-\Delta\omega)t}\rangle \tag{82}$$

where $\Delta\omega = \Omega_0^2/4\delta$. The frequency of the field is shifted, becoming smaller or larger depending on whether the detuning δ is positive or negative.

In this situation, the atom plays the same role as a slab of transparent material introduced into the cavity. If the refraction index of the material is different from one, the speed of light inside the cavity should change, and therefore the frequency of the standing wave should also change, as the wavelength is fixed by the boundary conditions. The frequency of the field will increase or decrease, depending on whether δ is negative or positive. This results in a change of the phase of the field, equal to $\Delta\omega t_{\text{int}}$, where t_{int} is the interaction time

between the material (atom or slab) and the cavity mode. From Eq. (82), one can see that the quantity

$$\phi = \frac{\Omega_0^2 t_{\text{int}}}{4\delta} \tag{83}$$

is the phase shift per photon.

In a recent experiment [60], a single Rydberg atom, with a speed of about 150 m/s, crossing a cavity mode of 51.1 GHz with a 6-mm waist, produced a phase shift of about $\pi/2$ for each level e and g, thanks to the huge transition dipole involved. The corresponding Rabi frequency was $\Omega/2\pi = 49$ kHz, and the detuning was $\delta/2\pi = 105$ kHz.

If the field changes its frequency, its energy also changes. This means that the slab (or the atom) introduced into the cavity realizes work, and therefore suffers a force, which can be attractive or repulsive, depending on the sign of the detuning.

We discuss in the following four nice applications of these concepts, in optical lattices, in the demonstration of single-atom effects in cavities in the optical region, in the realization of quantum gates and teleportation, and in the study of the dynamics of the decoherence process.

4. DISPERSIVE FORCES AND OPTICAL LATTICES

The dispersive force has been explored in experiments involving optical lattices: four laser beams produce a three-dimensional standing-wave pattern in space, similar to an egg carton. With the proper detuning, nonresonant atoms can get trapped in the valleys of the standing wave, thus mimicking the situation in crystals, with the atoms here replacing the electrons and the standing wave replacing the periodic potential in the crystal. This system has been used to demonstrate a quantum-mechanical phase transition, from an insulator to a superfluid phase (Mott transition) [61], following a proposal by Jaksch et al. [62]. In this experiment, ^{87}Rb atoms from a Bose–Einstein condensate are loaded into a three-dimensional optical lattice potential. The interaction between the atoms is repulsive, and the atom sample is sufficiently dilute so that the number of atoms per site (valleys of the standing-wave pattern) is of the order one to three. The condensate is a superfluid and is described by a wavefunction that exhibits long-range phase coherence. As the amplitude of the lasers is smoothly increased, one observes a phase transition from a situation in which a macroscopic phase is defined (this is the phase of the Bose–Einstein condensate) to a situation in which the atoms get localized at the valleys, the system thus behaving as an insulator. The system remains in the superfluid phase as long as the atom–atom interactions are small compared to the tunnel coupling through the crests of the lattice potential. In the opposite limit, when the repulsive atom–atom interactions are large compared to the tunnel coupling, the total energy is minimized when each lattice site is filled with the same number of atoms. The reduction of fluctuations in the atom number in each site leads to increased fluctuations in the phase. Thus, in the state with a fixed atom number per site, phase coherence is lost. In addition, a gap in the excitation spectrum appears.

In this way, one was able to demonstrate experimentally, in a situation in which one gets a much better control of the parameters than with experiments done with real crystals, the Bose–Hubbard Hamiltonian [63]:

$$\hat{H} = J \sum_{i \neq j} a_i^\dagger a_j + \sum_i \varepsilon_i n_i + \frac{1}{2} U \sum_i \hat{n}_i (\hat{n}_i - 1) \tag{84}$$

where the first term on the right-hand side stands for the tunneling of the bosonic atoms between different sites, the second term represents the energies of the different sites (due to the harmonic potential that confines the cold atoms), and the third term, with $U > 0$, corresponds to the repulsive short-range interaction between atoms in the same site. The phase transition originates here from the competition between the tunneling term, measured by the energy J, and the repulsion term, associated with the energy U, as opposed to normal temperature-dependent phase transitions, where the competition is between the inner energy and the entropy.

In the experiment described in Ref. [61], the lattice had about 150,000 sites (about 65 sites in each direction), and the state of the atomic system was probed by suddenly turning off the combined trapping potential. The atomic wavefunctions were then allowed to expand freely and interfere with each other. In the superfluid regime, where all atoms are delocalized over the entire lattice with equal relative phases between different lattice sites, a high-contrast three-dimensional interference pattern was obtained. As the lattice potential depth was increased, the resulting interference pattern changed markedly, with an incoherent background of atoms gaining more and more strength until no interference pattern was visible at all, thus demonstrating the continuous decrease of the superfluid fraction for increasing ratios U/J.

In experiments with optical lattices, not only the tunneling matrix elements can be controlled (by changing the intensity of the lasers that generate the optical lattice), but also the interaction between the atoms, through Feshbach resonances [64, 65]. Furthermore, this setup allows the realization of quantum gates with neutral atoms [66]. Optical lattices have also been considered for the simulation of other quantum Hamiltonians [67].

5. TRAPPING A SINGLE ATOM WITH A SINGLE PHOTON

The dependence on the atomic position of the energy levels of the dressed atom, displayed in Fig. 8, creates an effective potential well for the center-of-mass motion, which can be used to trap the atom in the cavity field [68–70]. Also, the variation of the energy of the dressed mode with the position of the atom can be used to control its position inside the cavity with great accuracy.

This was demonstrated in a beautiful experiment done by Jeff Kimble's group, in CALTECH [38]. The experimental scheme is sketched in Fig. 10. A high-Q optical cavity is pumped by a laser field P, tuned to the frequency ω_p displayed in Fig. 11. This frequency is detuned with respect to the relevant mode of the empty cavity, with frequency ω. The initial intensity inside the cavity is very small, corresponding to an average photon number equal to 0.05, so under these conditions the relevant cavity mode is excited by a small, off-resonance field.

A magneto-optical atom trap (MOT) is placed 3 mm above the cavity (about 10^4 cesium atoms, cooled to a temperature of about 20 μK), so that the released atoms fall under the action of the gravity, in a vacuum chamber at 10^{-8} Torr, crossing the cavity mode. The atoms are initially in the ground state $|g\rangle$, and each time the trapped atoms are released, at most one or two atoms cross the standing-wave mode of the cavity. The initial mean velocity of the atoms is about 4 cm/s, and after the 3-mm fall they reach the cavity with a velocity of about 24 cm/s. As the atom accelerates toward the center of the mode, under the action of the

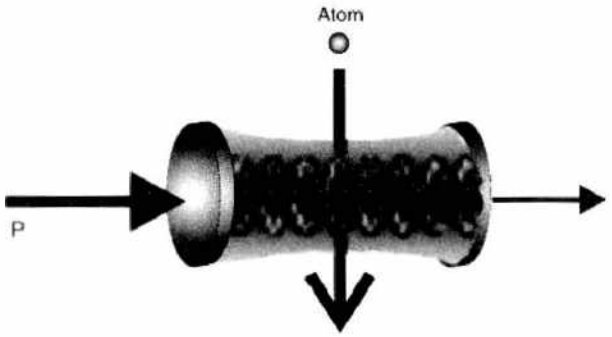


Figure 10. Sketch of the experiment that led to the trapping of a single atom by a field with one photon in the average. The cavity is pumped by a laser detuned from the empty-cavity mode. As a cold atom, released from an atom trap placed above the cavity axis, crosses the cavity mode, the pumping field gets resonant with the dressed-system mode. While the atom approaches the center of the mode, the intensity of the pumping field is suddenly switched up, at some predetermined threshold, so that the atom gets trapped in the cavity. The average number of intracavity photons, under these circumstances, is of the order one. The transmitted field depends on the position of the atom in the cavity. Measurement of this field allows the reconstruction of the orbit followed by the atom trapped by the cavity field.

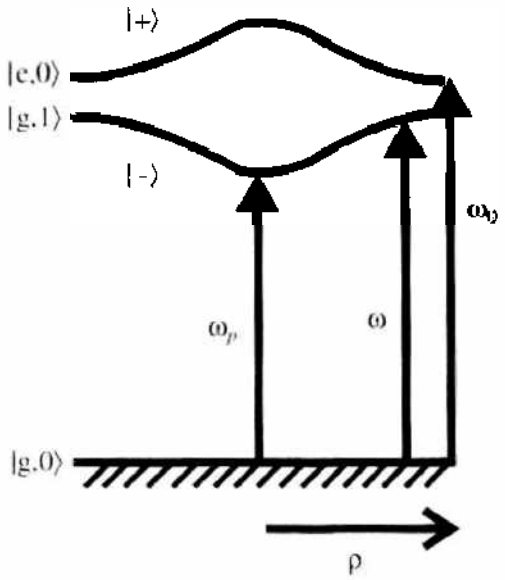


Figure 11. Dressed levels relevant for Kimble's experiment as functions of the transverse distance ρ. The atom gets trapped in the potential well corresponding to the unperturbed level $|g, 1\rangle$.

potential well depicted in Fig. 11, the frequency ω_p approaches resonance with the dressed mode sketched in Fig. 11, implying that the intensity of the transmitted field increases. This intensity is a function of the position of the atom in the cavity, and it is possible to find an inversion algorithm that gives this position as a function of the intensity [38]. On the other hand, by suddenly increasing the pumping intensity after some predetermined threshold, in such a way that the depth of the well gets higher than the kinetic energy of the atom, it is possible to trap the atom in the cavity.

In the experiment described in Ref. [38], the pumping power was switched up to a level of 0.3 intracavity photons, thus creating a deep confining potential around the atom, which trapped it. The depth of the potential was about 2.3 mK, greater than the initial kinetic energy of the atoms, of about 0.46 mK. When the atom is in the center of the mode, the average number of photons in the cavity is of the order one. This experiment demonstrates therefore the trapping of a single atom by a quantum field with an average number of photons of the order of one.

6. ATOMS AND PHOTONS AS QUBITS

Qubits are the natural generalization of bits in the quantum world: they are described by states of two-level systems, of the form $a|0\rangle + b|1\rangle$. In cavity quantum electrodynamics, one may associate qubits to two-level atoms, the corresponding states being $|\psi_{\text{atom}}\rangle = c_e|e\rangle + c_g|g\rangle$, and to field modes in high-Q cavities, under the restriction that there is at most one photon in the mode. The state of the field is then expressed as $c_0|0\rangle + c_1|1\rangle$.

6.1. Measuring the Atomic and Field Qubits

Measurement of the atomic qubit is achieved by sending the atom through an ionization chamber, which consists of two conducting plates with a potential difference between them, in such a way that a static electric field is produced with an amplitude that increases linearly along the direction of atomic motion. This field ionizes the atom, at a time that depends on the atomic state: the higher excited state is ionized under a weaker field than the less excited one. The ionized electron gives rise to a current, and the position of the current pulse on the screen of an oscilloscope will determine if the atom was measured in the upper or the lower state. Atomic coherence can also be measured by first letting the atom interact with a classical field (produced, for instance, in a low-Q cavity by a classical current—say a microwave generator, as described in Section 3.2), so that it undergoes a rotation in

state space, and then measuring its state. The resulting population will be a function of the rotation angle and the state coherences.

On the other hand, the field in the cavity is measured by sending an atom in state $|g\rangle$ through the cavity, interacting resonantly with the field. If the interaction time is such that the atom undergoes a π rotation, the field state is mapped onto the atomic state, which can then be measured by the techniques described above. Indeed, if the atom enters the cavity in the state $|g\rangle$, and the field is in the state $c_0|0\rangle + c_1|1\rangle$, the π interaction yields:

$$|g\rangle \otimes [c_0|0\rangle + c_1|1\rangle] \rightarrow [c_0|g\rangle - c_1|e\rangle] \otimes |0\rangle \tag{85}$$

so that the field probability amplitudes can be determined by measuring the atomic state.

For an open cavity geometry (two mirrors facing each other), the interaction time between the atoms and the cavity mode can be calibrated by applying a potential difference between the mirrors of the high-Q cavity. The atomic levels are Stark shifted by the resulting electric field, so that they can be taken into or out of resonance with the cavity mode, as they cross the cavity, by changing the potential difference applied to the mirrors.

6.2. Production of Entangled States

Note that an interaction time corresponding to a $\pi/2$ rotation would lead to the entangled state

$$|e\rangle \otimes |0\rangle \rightarrow \frac{1}{\sqrt{2}}(|e\rangle \otimes |0\rangle + |g\rangle \otimes |1\rangle) \tag{86}$$

Multiparticle entanglement can be achieved in the following way [71]. After the above state is prepared, one sends a second atom through the cavity, prepared in the state $(|g_2\rangle + |i_2\rangle)/\sqrt{2}$, where $|i_2\rangle$ is nonresonant with the cavity mode, and $|g_2\rangle$ undergoes a 2π rotation if there is one photon in the cavity. One gets then, denoting by the index 1 the first atom:

$$\begin{aligned}&\frac{1}{\sqrt{2}}(|g_2\rangle + |i_2\rangle) \otimes \frac{1}{\sqrt{2}}(|e_1\rangle \otimes |0\rangle + |g_1\rangle \otimes |1\rangle)\\ &\quad \rightarrow \frac{1}{2}[(|g_2\rangle + |i_2\rangle) \otimes |e_1\rangle \otimes |0\rangle + (-|g_2\rangle + |i_2\rangle) \otimes |g_1\rangle \otimes |1\rangle]\end{aligned} \tag{87}$$

This is already an entangled state of three subsystems: atom 1, atom 2, and the cavity mode. Indeed, if one defines $|\pm_2\rangle = (\pm|g_2\rangle + |i_2\rangle)/\sqrt{2}$, the above state can be written in the following way:

$$|\Psi_3\rangle = \frac{1}{\sqrt{2}}(|+_2\rangle \otimes |e_1\rangle \otimes |0\rangle + |-_2\rangle \otimes |g_1\rangle \otimes |1\rangle) \tag{88}$$

Sending a third atom in state $(|g_3\rangle + |i_3\rangle)/\sqrt{2}$ would produce an entangled state of four subsystems, and so on. One should note that, after entangling N atoms, using the above procedure, one may send a final atom in state $|g_{N+1}\rangle$, with a π interaction with the field mode if one photon is present. This atom absorbs the photon and gets excited to state $|e_{N+1}\rangle$, so that the cavity state is factored out and one is left with an entangled state of $N+1$ atoms:

$$|\psi_{N+1}\rangle = \frac{1}{2}(|e_1\rangle \otimes |+_2\rangle \cdots \otimes |+_N\rangle \otimes |g_{N+1}\rangle - |g_1\rangle \otimes |-_2\rangle \cdots \otimes |-_N\rangle \otimes |e_{N+1}\rangle) \tag{89}$$

This is a generalized state of the GHZ type, originally defined by Greenberger, Horne, and Zeilinger [72–74]. A GHZ state involving three photons was produced by Bowmeester et al. [75]. An entangled state involving four atoms in a trap was demonstrated by the NIST group [76].

These interactions actually provide the basic building blocks for doing in principle any quantum computation, as it will be shown now.

6.3. Quantum Computation with Atoms and Photons

The main ideas of quantum computation were set off by Paul Benioff [77], Richard Feynman [78], and David Deutsch [79] in the 1980s (for reviews, see Refs. [1–4]). In 1995, David DiVincenzo et al. [80, 81] established that unitary transformations on a single qubit, plus a special kind of two-qubit operation, the controlled-not gate, form a universal set for quantum computation: this implies that any computation can in principle be done if one knows how to implement these operations.

6.3.1. The Controlled-not Gate

The controlled-not (CNOT) gate is already defined in classical computation by the following "truth table," where the first and second digits correspond to the values of the first and second bits, respectively: $00 \rightarrow 00, 01 \rightarrow 01, 10 \rightarrow 11, 11 \rightarrow 10$. That is, the value of the second bit is changed (implying a "not" operation) if and only if the value of the first bit is equal to 1 (therefore, the "not" is controlled by the value of the first bit). The first bit, which is not changed by the operation, is the "control" bit, whereas the controlled bit is the "target" bit. This gate can be represented by a circuit diagram, as shown in Fig. 12. It is easy to check that this gate is reversible.

One should note that, even though this gate is already defined in classical computation, its behavior acquires an interesting twist if one allows the inputs to be states in Hilbert space. Indeed, if one takes the control bit in the state $c_0|0\rangle + c_1|1\rangle$ and the target bit in state $|0\rangle$, one gets, following the above truth table, the transformation

$$(c_0|0\rangle + c_1|1\rangle) \otimes |0\rangle \rightarrow c_0|00\rangle + c_1|11\rangle \tag{90}$$

which is an entangled state. This shows that entangled states appear naturally in quantum computing.

Equation (90) is basically the same as Eq. (1) specialized to the case in which the state of the system does not change upon measurement. This shows that the CNOT operation plays a basic role in quantum measurement.

The quantum CNOT can be expressed in terms of a Pauli σ_x matrix, as shown in Fig. 12, as

$$\sigma_x|0\rangle \rightarrow |1\rangle \quad \sigma_x|1\rangle \rightarrow |0\rangle \tag{91}$$

Or yet, in the two-qubit product basis, by the matrix

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \tag{92}$$

6.3.2. Building Phase Gates

A related gate, which can also be taken as a building block of quantum computation, together with the single-qubit unitary transformations, is the *Z gate*, which is represented by a circuit analogous to the one corresponding to the CNOT, except that σ_x is replaced by σ_z. It does not have any classical equivalence, as it involves a quantum phase, and it is defined through the truth table $|00\rangle \rightarrow |00\rangle, |01\rangle \rightarrow |01\rangle, |10\rangle \rightarrow |10\rangle, |11\rangle \rightarrow -|11\rangle$. That is, the phase of

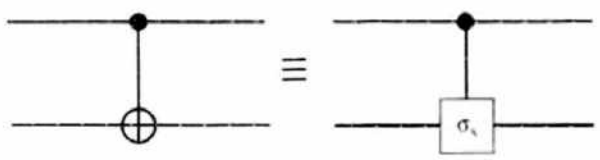


Figure 12. Diagramatic representation of a controlled-not gate. The upper line represents the control bit, which controls the state of the target bit, represented by the lower line: the state of the control bit does not change, whereas the target bit changes if and only if the control bit has the value one. If one associates to each bit a two-line column matrix, the negation operation can be represented by the Pauli matrix σ_x.

the two-qubit state is changed by π if and only if both qubits are in the state $|1\rangle$. In the product basis, it corresponds to the matrix

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \tag{93}$$

It is also a reversible transformation, as can be readily seen.

The CNOT and the Z gates are simply related via a *Hadamard* transformation:

$$H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} = \frac{1}{\sqrt{2}} (\sigma_x + \sigma_z) \tag{94}$$

Because $H^2 = 1$, this is also a reversible transformation. One should note that $-iH$ is a π rotation around an axes in the $x - z$ plane forming an angle of $\pi/4$ with both axes. This can be seen from Eq. (53), with $\Omega t = \pi$, or from the representation of H in terms of σ_x and σ_z.

As can be seen from the algebra of Pauli matrices, $H\sigma_z H = \sigma_x$, and therefore the CNOT gate can be obtained from the Z gate by sandwiching the target bit between two Hadamard operations, as shown in Fig. 13.

We have already seen how to implement unitary transformations on a single atomic qubit, by submitting it to a classical electromagnetic field. Let us show how two-qubit gates can be implemented.

The Z gate may be implemented in the following way. Let us consider three atomic levels, specified by the letters e, g, and i, as shown in Fig. 14. We assume that the atom has a resonant interaction with the cavity mode when the atom is in state $|e\rangle$, so that if there is a photon in the cavity, the atom undergoes a 2π transition, going from state $|e\rangle$ to state $|i\rangle$ and then back to state $|e\rangle$. This will multiply the state by a minus sign. On the other hand, if there is no photon in the cavity or if the atom is in state $|g\rangle$, nothing happens to the state. Therefore, if the atom is in level $|e\rangle$ and there is one photon in the cavity, the atom-field state is multiplied by a minus sign, whereas nothing happens in the other cases. This is precisely what one expects from a Z gate.

More generally, a phase gate can be built, corresponding to the following matrix:

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & e^{i\phi} \end{pmatrix} \tag{95}$$

This gate is implemented by detuning, through a Stark shift, the atomic transition with respect to the field mode, and letting again the Bloch vector describe a full circle on the Bloch sphere, if there is one photon in the field and the atom is in state $|e\rangle$. As the detuning increases, one may show that the phase ϕ goes from π, corresponding to a Z gate, to a value close to 0, as the detuning gets much higher than the resonant Rabi frequency. This was demonstrated experimentally in Ref. [82].

By adding now to this setup two low-Q cavities, which rotate the atomic state by $\pi/2$, in the way described before, one gets, as the atom crosses the three-cavity system, that if there

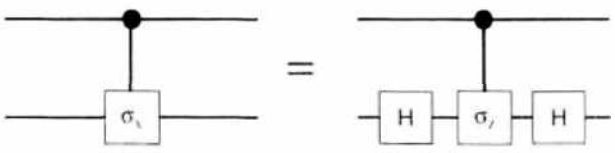


Figure 13. Relation between a controlled-not and a Z gate. The controlled-not gate is obtained from the Z gate by sandwiching the Z operation between two Hadamard gates.

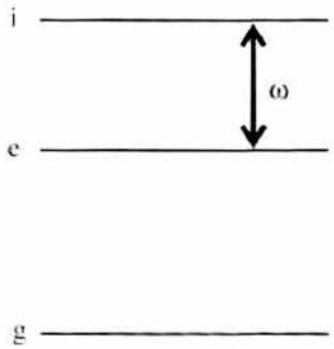


Figure 14. Level scheme for the implementation of a Z gate. The transition $|e\rangle \rightarrow |i\rangle$ is resonant with a cavity mode. If there is one photon in the mode, the atom initially in state $|e\rangle$ undergoes a 2π rotation in state space, going from e to i and then back to e, the state acquiring in the process a minus sign.

is no photon in the cavity, the atomic state is just flipped by the interaction with the two low-Q cavities. On the other hand, if the atom is in state $|e\rangle$ and there is one photon in the cavity, the atom-field state undergoes the following transformations, as the atom crosses the system:

$$|e\rangle \otimes |1\rangle \rightarrow \frac{1}{\sqrt{2}}(|e\rangle + |g\rangle) \otimes |1\rangle \rightarrow \frac{1}{\sqrt{2}}(-|e\rangle + |g\rangle) \otimes |1\rangle \rightarrow -|e\rangle \otimes |1\rangle \tag{96}$$

In the same way, one shows that the state $|g\rangle \otimes |1\rangle$ does not change. One gets therefore the following table:

$$|e\rangle \otimes |0\rangle \rightarrow |g\rangle \otimes |0\rangle, |g\rangle \otimes |0\rangle \rightarrow -|e\rangle \otimes |0\rangle, |e\rangle \otimes |1\rangle \rightarrow -|e\rangle \otimes |1\rangle, |g\rangle \otimes |1\rangle \rightarrow |g\rangle \otimes |1\rangle \tag{97}$$

If one calls the state with zero photons in the mode as bit "one" and the state with one photon as bit "zero," we get a gate that looks similar to the CNOT gate, in the sense of flipping the atomic bit if the field bit is 1, and doing nothing, except for a sign change, if the field bit is zero. Of course, this is not strictly a CNOT gate, as there are phase changes, in addition to the flipping. Indeed, one has here $\pi/2$ rotations, as opposed to the π rotation associated with iH. This is not a problem, as the resulting two-qubit transformations, together with the single-atom transformations, also constitute a universal set for quantum computation.

An immediate application of these gates is the teleportation of an atomic state, which will be discussed in the following section.

6.4. Teleporting an Atomic State

The interactions and transformations discussed above could be used to demonstrate the phenomenon of teleportation. The possibility of teleporting a quantum state using an entangled state as communication channel was proposed in Ref. [27]. This proposal is summarized in the following.

Suppose Alice wants to transmit to Bob a quantum state of a system in her possession. For instance, she could have a spin in the state

$$|\phi\rangle = c_+|\uparrow\rangle_1 + c_-|\downarrow\rangle_1 \tag{98}$$

If she would have only classical means for that purpose (for instance, a telephone line), she would face serious problems. First, if she only has a single spin with her, and she ignores its state, then it would be impossible for her to measure it, as this would require an infinite number of systems prepared in the same way. So she would not be able to know the state to be transmitted! Even if she had produced the state somehow (for instance, through a Stern–Gerlach apparatus), so that she would have full knowledge about it, she would still run into problems. Indeed, she would need to transmit the information on the complex

coefficients c_1 and c_2, and this would require, in most cases, an infinite numbers of bits, as these coefficients would involve, in general, an arbitrary number of digits.

Alice can do much better by sharing with Bob a two-spin entangled state, of the form (singlet state):

$$|\psi_{23}^{(-)}\rangle = \frac{1}{\sqrt{2}}(|\uparrow\rangle_2|\downarrow\rangle_3 - |\downarrow\rangle_2|\uparrow\rangle_3) \tag{99}$$

The combined system (Alice's spin plus entangled pair) is described by the state:

$$\begin{aligned}|\psi_{123}\rangle &= (c_+|\uparrow\rangle_1 + c_-|\downarrow\rangle_1)\otimes\frac{1}{\sqrt{2}}(|\uparrow\rangle_2|\downarrow\rangle_3 - |\downarrow\rangle_2|\uparrow\rangle_3)\\ &= \frac{c_+}{\sqrt{2}}(|\uparrow\rangle_1|\uparrow\rangle_2|\downarrow\rangle_3 - |\uparrow\rangle_1|\downarrow\rangle_2|\uparrow\rangle_3) + \frac{c_-}{\sqrt{2}}(|\downarrow\rangle_1|\uparrow\rangle_2|\downarrow\rangle_3 - |\downarrow\rangle_1|\downarrow\rangle_2|\uparrow\rangle_3)\end{aligned}$$

The products of the states of spins 1 and 2 that show up in the above equation are now reexpressed in terms of the Bell states, defined as:

$$\begin{aligned}|\Psi_{12}^{(-)}\rangle &= \frac{1}{\sqrt{2}}(|\uparrow\rangle_1|\downarrow\rangle_2 - |\downarrow\rangle_1|\uparrow\rangle_2)\\ |\Psi_{12}^{(+)}\rangle &= \frac{1}{\sqrt{2}}(|\uparrow\rangle_1|\downarrow\rangle_2 + |\downarrow\rangle_1|\uparrow\rangle_2)\\ |\Phi_{12}^{(-)}\rangle &= \frac{1}{\sqrt{2}}(|\uparrow\rangle_1|\uparrow\rangle_2 - |\downarrow\rangle_1|\downarrow\rangle_2)\\ |\Phi_{12}^{(+)}\rangle &= \frac{1}{\sqrt{2}}(|\uparrow\rangle_1|\uparrow\rangle_2 + |\downarrow\rangle_1|\downarrow\rangle_2)\end{aligned}$$

One gets then:

$$\begin{aligned}|\psi_{123}\rangle = \frac{1}{2}\big[&|\Psi_{12}^{(-)}\rangle(-c_+|\uparrow\rangle_3 - c_-|\downarrow\rangle_3) + |\Psi_{12}^{(+)}\rangle(-c_+|\uparrow\rangle_3 + c_-|\downarrow\rangle_3)\\ &+ |\Phi_{12}^{(-)}\rangle(c_+|\downarrow\rangle_3 + c_-|\uparrow\rangle_3) + |\Phi_{12}^{(+)}\rangle(c_+|\downarrow\rangle_3 - c_-|\uparrow\rangle_3)\big]\end{aligned}$$

One should note that the Bell state $|\Psi_{12}^{(-)}\rangle$ is correlated with the state of spin 3 that coincides with the original state of spin 1 (except for an irrelevant overall minus sign), whereas the other Bell states are correlated with states of spin 3 that are obtained from the original state by π rotations around the z, y, or x axis. Alice now performs a Bell-state measurement on the two spins in her possession (spins 1 and 2), thus projecting the above state onto one of the four Bell components. She then informs Bob about the result of the measurement (note that this implies two bits of information, as there are four Bell states). Depending on the two bits of information he gets, he applies to particle 3 either no rotation at all (if system 1–2 is found in state $|\Psi_{12}^{(+)}\rangle$), or a π rotation around the z, y, or x axis, thus replicating the original state.

In this way, Alice succeeds in transmitting to Bob the full information about the quantum state, using only two bits of information!

In Ref. [27], it was not made clear how the measurement of Bell states could be made, in concrete terms.

A logical circuit that implements this measurement is shown in Fig. 15. This circuit builds up the four Bell states from product-state inputs and inversely transforms Bell states into product states, according to the table:

In	Out
$\|\uparrow\uparrow\rangle$	$(\|\uparrow\uparrow\rangle + \|\downarrow\downarrow\rangle)/\sqrt{2} \equiv \|\Phi^+\rangle$
$\|\uparrow\downarrow\rangle$	$(\|\uparrow\downarrow\rangle + \|\downarrow\uparrow\rangle)/\sqrt{2} \equiv \|\Psi^+\rangle$
$\|\downarrow\uparrow\rangle$	$(\|\uparrow\uparrow\rangle - \|\downarrow\downarrow\rangle)/\sqrt{2} \equiv \|\Phi^-\rangle$
$\|\downarrow\downarrow\rangle$	$(\|\uparrow\downarrow\rangle - \|\downarrow\uparrow\rangle)/\sqrt{2} \equiv \|\Psi^-\rangle$

(100)

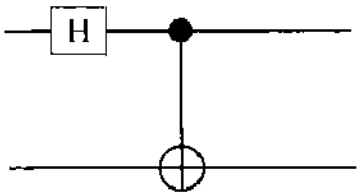


Figure 15. Circuit leading to the production and the detection of Bell states. When the input on the left-hand side is one of the product states $|\uparrow\rangle|\uparrow\rangle$, $|\uparrow\rangle|\downarrow\rangle$, $|\downarrow\rangle|\uparrow\rangle$, and $|\downarrow\rangle|\downarrow\rangle$, the output on the right-hand side is one of the Bell states. On the other hand, when Bell states are entered on the right-hand side, the output on the left-hand side is a product state. Measurement of the two output spins in this case allows one to know which of the Bell states was fed into the right-hand side of the circuit.

Upon measuring the spins of the product states, one can uniquely determine the Bell state at the input.

The first experimental proposal for implementation of this teleportation scheme was published in Ref. [28]. We present here a modified version of that proposal.

The basic setup is illustrated in Fig. 16. A high-Q cavity (**C**), used to keep the field qubit, is placed between two low-Q cavities ($\mathbf{R}_1$ and $\mathbf{R}_2$), used to transform the atomic qubits.

The high-Q cavity does not have any photons, initially. We start by sending through this cavity a two-level atom in the excited state $|e_3\rangle$, so that it undergoes a $\pi/2$ interaction in **C** (the two low-Q cavities do not play any role in this part of the process—one may assume for instance that the atom is Stark-shifted as it crosses these cavities, so as to be highly detuned with respect to the corresponding cavity modes):

$$|e_3\rangle|0\rangle \rightarrow \frac{1}{\sqrt{2}}(|e_3\rangle|0\rangle + |g_3\rangle|1\rangle) \tag{101}$$

This atom is then sent to Bob. The quantum channel is here formed by the entangled pair photon + atom 3.

Once this state is prepared, the teleportation process is carried out in the following way. One sends an atom prepared in the state $|\psi\rangle_a = c_e|e_1\rangle + c_g|g_1\rangle$ (the state to be teleported) through the system of cavities, so that the atom undergoes a 2π resonant interaction in **C** if it is in state $|e\rangle$ and this cavity contains one photon, thus adding a minus sign to the state. The atom also undergoes $\pi/2$ resonant interactions with the classical fields in cavities $\mathbf{R}_1$ and $\mathbf{R}_2$. This setup was discussed in Section 6.3, and according to Eq. (97), we get for the state that includes atom 1, atom 3, and the cavity field:

$$\begin{aligned}(c_g|g_1\rangle + c_e|e_1\rangle) \otimes \frac{1}{\sqrt{2}}(|e_3, 0\rangle + |g_3, 1\rangle) \\ \rightarrow \frac{c_g}{\sqrt{2}}(-|e_1, e_3, 0\rangle + |g_1, g_3, 1\rangle) + \frac{c_e}{\sqrt{2}}(|g_1, e_3, 0\rangle - |e_1, g_3, 1\rangle)\end{aligned} \tag{102}$$

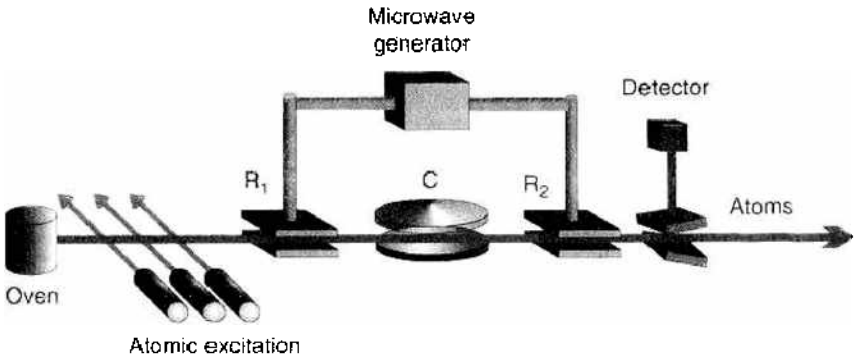


Figure 16. Teleportation experiment. The atoms are sent through the three cavities $\mathbf{R}_1$, **C**, and $\mathbf{R}_2$, and may be measured by a set of ionizing plates, with a static electric field that has an increasing amplitude along the direction of the atomic beam.

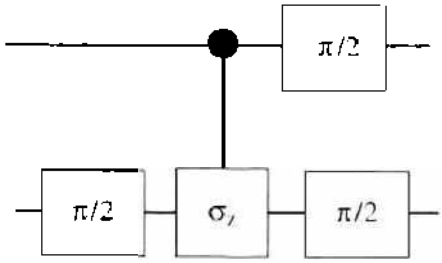


Figure 17. Circuit corresponding to the teleportation experiment with atoms and photons. The $\pi/2$ rotations of the atomic state, plus the Z gate, replace here the Hadamard gate and the CNOT in Fig. 15.

After atom 1 exits the cavity, another atom, which we call atom 2, prepared in the state $|g_2\rangle$, is sent through the same cavity (for this atom, the interaction with the field in the first low-Q cavity $\mathbf{R}_1$ is turned off), and it has now a resonant interaction with the field, so that if there is one photon in the cavity, this photon is absorbed by the atom, which ends up in state $-|e\rangle$. This second atom undergoes then a $\pi/2$ rotation, by interacting with the field in the cavity $\mathbf{R}_2$.

The state of the system becomes now:

$$\frac{1}{2}[|g_1, g_2\rangle(-c_g|g_3\rangle + c_e|e_3\rangle) - |g_1, e_2\rangle(c_g|g_3\rangle + c_e|e_3\rangle) + |e_1, g_2\rangle(-c_g|e_3\rangle + c_e|g_3\rangle) + |e_1, e_2\rangle(c_g|e_3\rangle + c_e|g_3\rangle)] \tag{103}$$

This expression makes it clear that, if the populations of atoms 1 and 2 are measured, one projects the state of the third atom either on the same state atom 1 had initially (if atom 1 is detected in state $|g_1\rangle$ and atom 2 is detected in state $|e_2\rangle$), or on states that are easily transformed onto this state by simple π rotations around the axes x, y, and z, generated by the three Pauli matrices. The proper rotation depends on which of the four possibilities (g_1, g_2), (g_1, e_2), (e_1, g_2), or (e_1, e_2) is measured by Alice. By calling Bob, she can tell him which one she got (this amounts to two bits of classical information), and with this information Bob can apply the proper transformation to the atom in his possession, so as to recover the state of atom 1. In this way, one succeeds in teleporting the state of atom 1 to atom 3.

The circuit corresponding to the above proposal is shown in Fig. 17. It is not quite the one shown in Fig. 15. Indeed, one deals here with $\pi/2$ rotations in state space, rather than Hadamard gates.

One should notice two characteristic features of the teleportation process: first, the original state is destroyed, by letting atom 1 interact with the photon in cavity C and then detecting its state. This can be seen as an illustration of the "no-cloning" theorem [83, 84]: it is not possible to replicate the state of a single quantum system. The second feature is that Bob is able to reconstruct the state of the atom that was with Alice without either Alice or Bob knowing about it. Indeed, it is impossible to know the quantum state of a single quantum system (one needs an ensemble of identically prepared systems in order to measure a quantum state).

The teleportation of atomic states, as described above, has not been realized experimentally yet. Teleportation of spin states using nuclear magnetic resonance techniques was demonstrated by Nielsen et al. [32]. Demonstrations of teleportation with light beams have been done by several groups [29–31, 33–37].

7. COHERENT SUPERPOSITIONS OF MESOSCOPIC STATES IN CAVITY QUANTUM ELECTRODYNAMICS

7.1. Building the Coherent Superposition

We now show how, by carefully tailoring the interactions between two-level atoms and one mode of the electromagnetic field in a cavity, one can produce quantum superpositions of

distinguishable coherent states of the field, thus mimicking the superposition of two classically distinct states of a pointer.

The method for generating the quantum superposition of two coherent states, proposed in Ref. [85], and sketched in Fig. 18, involves a beam of circular Rydberg atoms [45] crossing a high-Q cavity **C** in which a coherent state is previously injected (this is accomplished by coupling the cavity to a classical source—a microwave generator—through a wave guide). The use of circular levels is due to their strong coupling to microwaves and their very long radiative decay times, which makes them ideally suited for preparing and detecting long-lived correlations between atom and field states [86]. On either side of the high-Q cavity there are two low-Q cavities ($\mathbf{R}_1$ and $\mathbf{R}_2$), which remain coupled to a microwave generator, as in the scheme proposed for the teleportation experiment. The fields in these two cavities can be considered as classical, for the reasons mentioned before.

This set of two low-Q cavities constitutes the usual experimental arrangement in the Ramsey method of interferometry [86, 87]. Two of the (highly excited) atomic levels, which we denote by $|e\rangle$ (the upper level) and $|g\rangle$ (the lower one), are resonant with the microwave fields in cavities $\mathbf{R}_1$ and $\mathbf{R}_2$. The intensity of the fields in $\mathbf{R}_1$ and $\mathbf{R}_2$ is chosen so that, for the selected atomic velocity, effectively a $\pi/2$ pulse is applied to the atom as it crosses each cavity. For a properly chosen phase of the microwave field, this pulse transforms the state $|e\rangle$ into the linear combination $(|e\rangle + |g\rangle)/\sqrt{2}$, and the state $|g\rangle$ into $(-|e\rangle + |g\rangle)/\sqrt{2}$.

Therefore, if each atom is prepared in the state $|e\rangle$ just prior to crossing the system, after leaving $\mathbf{R}_1$ the atom is in a superposition of two circular Rydberg states $|e\rangle$ and $|g\rangle$:

$$|\psi_{\text{atom}}\rangle = \frac{1}{\sqrt{2}}(|e\rangle + |g\rangle) \tag{104}$$

On the other hand, the superconducting cavity is assumed not to be in resonance with any of the transitions originating from those two atomic states. This means that the atom does not suffer a transition and does not emit or absorb photons from the field. This property is further enhanced by the fact that the cavity mode is such that the field slowly rises and decreases along the atomic trajectory, following a Gaussian mode profile, so that, for sufficiently slow atoms, the atom-field coupling changes adiabatically. However, the cavity is tuned in such a way that it is much closer to resonance with respect to one of those transitions, say the one connecting $|e\rangle$ to some intermediate state $|i\rangle$. The relevant level scheme is illustrated in Fig. 19. This implies that, if the atom crosses the cavity in state $|e\rangle$, dispersive effects can induce an appreciable phase shift on the field in the cavity. That is, the atom acts like a refraction index, changing the frequency of the field while the interaction is on—the corresponding energy change is just the AC-Stark shift, which for a Fock state of the electromagnetic field is proportional to the number of photons in the cavity. This frequency shift, multiplied by the interaction time between the atom and the mode, leads to a phase shift

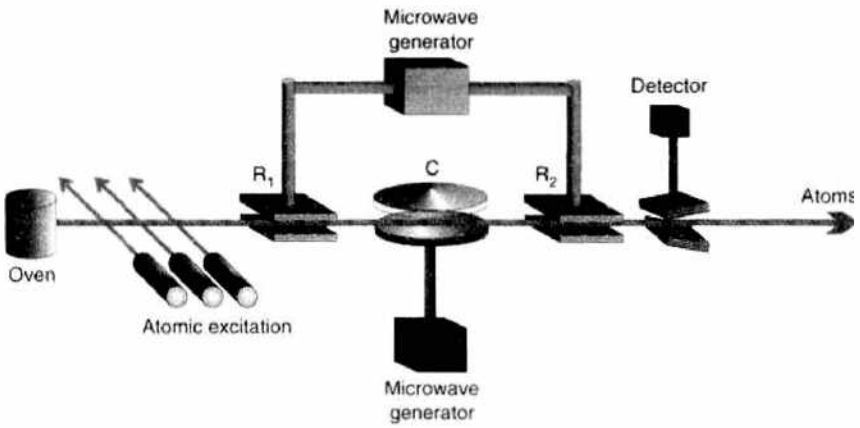


Figure 18. Experimental arrangement for producing and measuring a coherent superposition of two coherent states of the field in cavity **C**.

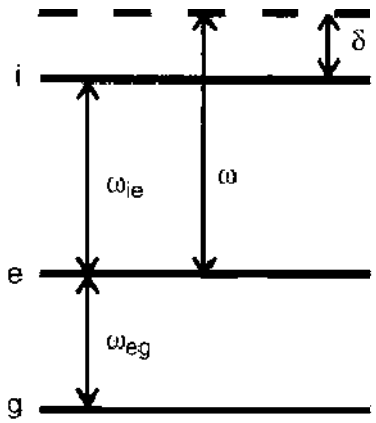


Figure 19. Atomic level scheme used for the production and measurement of a coherent superposition of two coherent states of the electromagnetic field in a cavity: The transition $i \leftrightarrow e$ is detuned by δ from the frequency ω of a mode of cavity **C**, whereas the transition $e \leftrightarrow g$ is resonant with the fields in $\mathbf{R}_1$ and $\mathbf{R}_2$. State $|g\rangle$ is not affected by the field in **C**.

of the field in the cavity, if the atom is in state $|e\rangle$. The phase shift is negligible, however, if the atom is in state $|g\rangle$.

Note that, as discussed in Section 3.5 and shown in Eq. (82), if there is a coherent state in the cavity, a phase shift of ϕ per photon if the atom is in state $|e\rangle$ implies that the phase of the coherent state is shifted by ϕ. After the atom has crossed the cavity, in a time short compared to the field relaxation time and also to the atomic radiative dampening time, the state of the combined atom-field system can be written as

$$|\psi_{\text{atom-field}}\rangle = \frac{1}{\sqrt{2}}(|e; -\alpha\rangle + |g; \alpha\rangle) \tag{105}$$

assuming that the phase shift is π per photon if the atom is in the excited state. The entanglement between the field and atomic states is analogous to the correlated two-particle states in the Einstein–Podolski–Rosen (EPR) paradox [88–91]. The two possible atomic states e and g are here correlated to the two field states $|-\alpha\rangle$ and $|\alpha\rangle$, respectively. This entangled state may also be interpreted as an example of the measurement process discussed in the introductory section of this article: the two coherent states are pointers of the measuring apparatus in this case, their phases depending on the state of the microscopic system being measured.

One should also note the analogy with the CNOT gates discussed in Section 6. Here the atom plays the role of control bit, whereas the cavity field is the target bit. If one associates the bit 0 to state $|\alpha\rangle$ and the bit 1 to state $|-\alpha\rangle$, then one sees that state $|e\rangle$ changes the target bit, whereas state $|g\rangle$ does not change it. The atomic state remains the same, as expected from a control bit.

After the atoms leave the superconducting cavity, one can detect them in the e or g states, by sending them through the ionization chamber described before. This measurement projects the field in the cavity either onto the state $|\alpha\rangle$ (if the atom is detected in state g), or onto the state $|-\alpha\rangle$ (if the atom is detected in state e). However, as in an EPR experiment [91], one may choose to make another kind of measurement, letting the atom cross, after it leaves the superconducting cavity, a second classical microwave field ($\mathbf{R}_2$ in Fig. 18), which amounts to applying to the atom another $\pi/2$ pulse. The state in Eq. (105) gets transformed then into

$$|\psi'_{\text{atom-field}}\rangle = \frac{1}{2}(|e; -\alpha\rangle - |e; \alpha\rangle + |g; \alpha\rangle + |g; -\alpha\rangle) \tag{106}$$

If one detects now the atom in the state $|g\rangle$ or $|e\rangle$, the field is projected onto the state

$$|\psi_{\text{cat}}\rangle = \frac{1}{N_+}(|\alpha\rangle + e^{i\phi_+}|-\alpha\rangle) \tag{107}$$

where $N_1 = \sqrt{2[1 + \cos\psi_1 \exp(-2|\alpha|^2)]}$ and $\psi_1 = 0$ or π, according to whether the detected state is g or e, respectively. One produces therefore a coherent superposition of two coherent states, with phases differing by π. For $|\alpha|^2 \gg 1$, this is a "Schrödinger cat-like" state: a coherent superposition of two classical distinguishable states.

Superpositions of coherent states were first prepared with trapped ions [92], using similar techniques. In this case, the superposition involved different states of the center-of-mass motion. The vibrational mode of the ion plays the role of the cavity mode, whereas the ion's electronic states stand for the two-level atom.

Superpositions of coherent states of the field were produced in the experiment reported in Ref. [55] and were detected by a procedure proposed in Refs. [93, 94]. This was the first time the continuous decoherence process was monitored in real-time. The method for following the decoherence is discussed in the following Section.

7.2. Measuring the Coherent Superposition

Once the quantum superposition is produced, how could one tell the difference between such a superposition and a statistical mixture of the two coherent states? This can be done by simply sending another atom, in the same initial state as the first one. It can be shown then [94] that, for the state (107), with $|\alpha| \gg 1$, there is a perfect correlation between the measurements of the first and the second atom: both are always detected in the same state. On the other hand, for the corresponding statistical mixture, the probability of detecting the second atom in state $|e\rangle$ is 50%, independently of which state was detected for the first atom. By delaying the sending of the second atom, one may thus explore the dynamical process by which the quantum superposition is transformed into a statistical mixture, due to the always present dissipation in a nonperfect cavity.

The time-dependent behavior of the conditional probability for measuring the second atom in the upper state, knowing that the first atom was also measured in the upper state, is displayed in Fig. 20. The sharp decay of this conditional probability from the perfectly coherent situation to the plateau associated with an incoherent superposition defines the *decoherence time*. This time can be shown to be equal to the dissipation time for the field in the cavity divided by twice the average number of photons in the field [94]. More generally, for a coherent superposition $|\psi\rangle \propto |\alpha_1\rangle + |\alpha_2\rangle$, the decoherence time is equal to the dissipation time divided by the "distance" in phase space $D = |\alpha_1 - \alpha_2|^2/2$. Thus, it becomes shorter as the field becomes more macroscopic. Note also that the plateau eventually disappears, and the probability for measuring the second atom in the state $|e\rangle$ goes to zero. This can be easily understood: the field in the cavity **C** leaks out, and therefore the sole effect on the atom initially prepared in the state $|e\rangle$ is the sum of two $\pi/2$ pulses in the cavities $\mathbf{R}_1$ and $\mathbf{R}_2$, that is a π pulse, which takes the atom into the state $|g\rangle$.

This analysis shows that the coherent states are "pointer states" in the current context: while they decay with the characteristic dampening time of the mode, t_{cav}, the coherence between these states decays with a much faster rate.

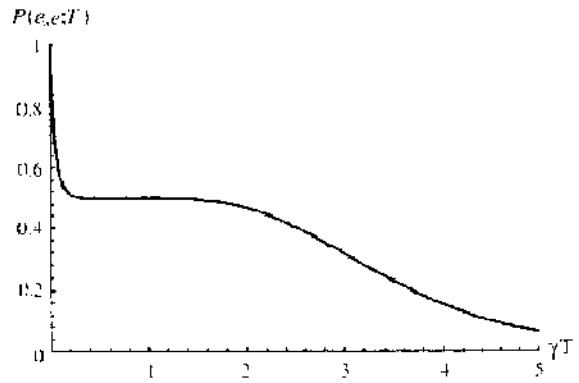


Figure 20. Conditional probability for finding the second atom in state $|e\rangle$ if the first atom was detected in state $|e\rangle$, as a function of time (measured in units of the field damping time).

An experimental realization of this proposal was made in 1996 by Haroche's group at Ecole Normale Supérieure, in Paris [55]. The dynamical measurement of the decoherence process, as proposed above, was in agreement with the theoretical predictions. In the actual experiment, instead of the three-level configuration discussed above, both the states $|e\rangle$ and $|g\rangle$ underwent a dispersive interaction with the cavity mode, resulting in a phase displacement of the field by some angle ϕ for state $|e\rangle$ and $-\phi$ for state $|g\rangle$. This scheme has the advantage that one is able to get the same separation between the two coherent states as in the previous configuration, but with a smaller interaction time, which helps to keep this interaction time smaller than the decoherence time.

In the experiment described in Ref. [55], the states e and g were circular Rydberg levels of rubidium atoms, with principal quantum numbers 51 and 50 (transition frequency equal to 51.099 GHz), with a long radiative lifetime (30 ms) and a strong coupling to radiation (resonant Rabi frequency $\Omega_0/2\pi = 48$ kHz). The Q of the superconducting Niobium cavity, cooled to 0.6 K, was 5.1×10^7, corresponding to a photon lifetime equal to 160 μs. At this temperature, the number of thermal radiation photons is negligible (of the order 0.05 photons). The cavity was tuned by adjusting the mirror separation, thus varying $\delta/2\pi$ between 70 and 800 kHz. The effective interaction time t_{int} between the atoms and the cavity mode was 19 ms, corresponding to atoms with a velocity of 400 m/s. The average number of photons in the cavity was changed from 3.3 to 5.1. Different detunings give rise to different phase shifts and therefore to different distances between the two coherent states in phase space. This allows testing the dependence on this distance of the decoherence time. For $\delta = 100$ kHz, ϕ is 0.69 radian, a large single-atom refraction index effect. The separation between the two successive atoms, the one used to build the coherent superposition and the probe atom, was varied from 30 to 250 μs.

7.3. Physical Interpretation of the Measurement of Decoherence

The measurement of decoherence discussed above has a simple physical interpretation, when the phase difference between the two coherent states is equal to π. In this case, one should notice that

$$|\alpha\rangle + |-\alpha\rangle \propto \sum_{k=0}^{\infty} \frac{\alpha^{2k}}{(2k)!}|2k\rangle \tag{108}$$

$$|\alpha\rangle - |-\alpha\rangle \propto \sum_{k=0}^{\infty} \frac{\alpha^{2k+1}}{(2k+1)!}|2k+1\rangle \tag{109}$$

so that the parity of the field is well-defined in these two cases. Therefore, if one starts with the atom in state $|e\rangle$ (Bloch vector up), it will undergo a $\pi/2$ rotation in the first low-Q cavity, around the pseudo-magnetic field (which in this resonant situation will be on the equatorial plane of the Bloch sphere), implying that the corresponding Bloch vector will be taken to the equatorial plane of the Bloch sphere. Upon crossing the high-Q cavity, the Bloch vector is rotated around the vertical axis, due to the dispersive interaction, by an odd multiple of π, if the parity of the field is odd, or an even multiple of 2π, if the parity is even. In the first case, after the atom crosses the second low-Q cavity, the corresponding Bloch vector will be pointing upward (so the atom will again be in the state $|e\rangle$). Otherwise, the Bloch vector will point downward. This sequence of transformations is displayed in Fig. 21.

This scheme amounts therefore to a measurement of the parity of the field in the cavity [95]: as one measures the state of the atom, after the second Ramsey region, one gains information about the field in the cavity. If the atom is measured in state $|e\rangle$, one knows that the parity of the field (initially not well-defined, as the field was in a coherent state) should be odd, otherwise the parity of the field should be even. This acquired information amounts to the projection of the original coherent state onto a "cat-like" state of the electromagnetic field. If nothing else happens to the cavity field, then sending the second atom in the same state as the first should confirm the first measurement: that is the reason why there is perfect correlation between the two measurements, if no dissipation is present.

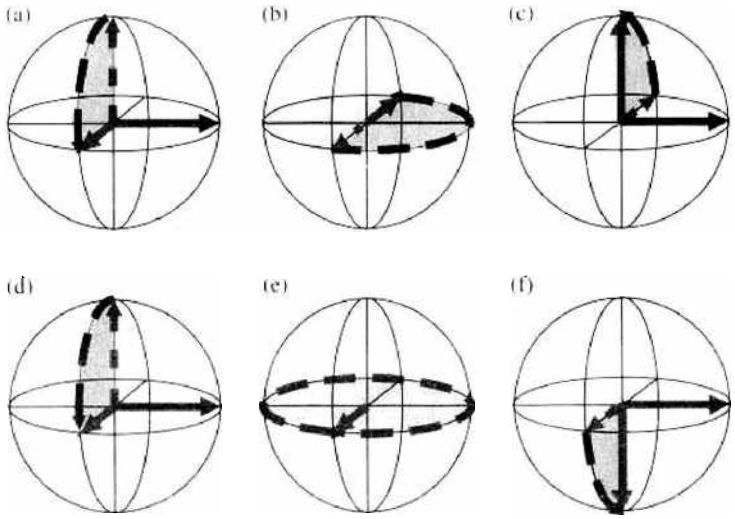


Figure 21. Evolution of the Bloch vector (gray arrow) of an atom that crosses the proposed setup, interacting with the eletromagnetic field (black arrow) in the two Ramsey zones, and having a dispersive interaction with the field in the superconducting cavity C, when the number of photons of the field in C has a well defined parity. The atom is initially in the state e. As the atom crosses the first Ramsey zone, its Bloch vector is rotated by $\pi/2$ around the vector representing the electromagnetic field along the real polarization axis, as shown in (a) and (d). As the atom crosses the cavity C, the Bloch vector rotates around the population axis. If the number of photons in the cavity is odd (b), the Bloch vector ends up pointing toward the opposite direction, and the rotation in the second zone leads the atom back to state $|e\rangle$ (c). On the other hand, if the number of photons in C is even, the Bloch vector turns by an integer multiple of 2π, so its direction does not change (e). The second Ramsey zone then brings the atom to $|g\rangle$ (f).

On the other hand, dissipation will imply that, if one starts with a state of well-defined parity, the holes in the photon-number distribution will quickly get filled, as the cavity may lose photons, the characteristic time for losing a single photon being the decay time for the field, t_{cav}, divided by the average number of photons in the state. After the holes get filled, again parity is not well-defined, and the atoms will exit the system either in the $|e\rangle$ or the $|g\rangle$ state.

The physical origin of the decoherence process is actually very simple: as the field in the cavity leaks into the external reservoir, the states of the field get correlated with states of the reservoir that become approximately orthogonal after the time it takes for one photon to leave the cavity, thus implying the disappearance of interference effects between the two internal states. This is more easily seen for the state $|\psi'\rangle = (1/\mathcal{N})[|2\alpha\rangle + |0\rangle]$, which is obtained by displacing the state (107) in the cavity by α. For the state $|2\alpha\rangle$, a photon leaves the cavity in a time of the order $(1/4|\alpha|^2\Gamma)$, whereas for the state $|0\rangle$, no photon leaves the cavity. Because the probability for finding the system in each of these states is $1/2$ for $|\alpha| \gg 1$, it follows that the effective lifetime of a photon is $(1/2|\alpha|^2\Gamma)$, which is precisely the decoherence time obtained for the state (107): this is a consequence of the fact that the distance D is the same for both states. After this time, the state $|2\alpha\rangle$ becomes correlated with a state of the reservoir containing approximately one photon, whereas the state $|0\rangle$ remains correlated with the vacuum. Decoherence of the system under observation is therefore closely connected with entanglement between this system and the reservoir.

A formal treatment of the effect of dissipation will be developed in Section 11.

7.4. Decoherence due to Classical Noise

The decoherence discussed in the previous sections involved entanglement between states of the system and the environment. Classical noise may also induce decoherence, which is however quite different conceptually. In fact, if one starts with a system in a pure state and one turns on classical noise on the system and then after sometime one turns it off, the system will remain pure, whereas with entanglement a mixed state unavoidably sets in. Classical noise may serve useful purposes, however, for demonstration purposes. Also, it is a matter of concern for the realization of quantum computers, and therefore it should be understood

and controlled. It was used by Wineland's group at NIST to investigate the decoherence of several states of the center-of-mass harmonic motion of a trapped ion under different kinds of noise [96, 97]. In these experiments, the action of both classical noise and quantum reservoirs was studied.

8. QUANTUM NONDEMOLITION MEASUREMENT OF THE FIELD POPULATIONS

In the previous section, decoherence was measured by sending a single atom through the cavity and measuring the parity of the field. This amounts to one bit of information on the field in the cavity: either the photon distribution is odd or even. It is translated into one bit of information on the exiting atom: either it is in state e or in state g. Getting more information on the field in the cavity implies sending other atoms through it. It will now be shown that the experimental scheme discussed above could be used to implement a quantum nondemolition (QND) measurement of the field populations.

The QND method consists in measuring a signal observable A_S of a quantum system S by detecting a change in an observable A_P of a probe P coupled to S during the measurement time, without perturbing the subsequent evolution of A_S [98–101]. QND measurements of the photon number in a cavity would allow the detection of very small classical forces acting on the cavity walls, which would be the case if these cavities were used as sensors of oscillations of gravitational-wave bar detectors.

QND methods are generally based on dispersive and nonlinear effects. One may, for instance, send a light beam through a dispersive medium with a nonlinear refraction index (Kerr cell). As long as the absorption is kept negligible, the intensity of the beam remains constant (it would correspond to the observable A_S mentioned above). Only its phase changes, but this change does not affect the intensity. On the other hand, the change in the refraction index, and therefore the intensity of the beam, can be determined by measuring the dephasing of another beam sent through the same medium [102]. The phase of this probe beam would correspond to the observable A_P. This possibility was demonstrated experimentally by Shelby et al. [103] and Grangier et al. [104]. This measurement can be used to reduce quantum fluctuations in the original beam through the technique of active stabilization [105, 106]: the intensity fluctuations of the beam are counteracted according to the result of the nondemolition measurement.

QND measurements can be used to reconstruct the photon-number distribution of a field in a cavity. Furthermore, through the successive increase of quantum-mechanical knowledge about the state of the field, acquired by a continuous QND measurement, one may produce a Fock state of the field in a cavity [87, 107]. This method is now discussed in detail.

The proposed experimental scheme is the same as the one used to produce and detect a coherent superposition of two coherent states, sketched in Fig. 18. It consists of a superconducting cavity with a high quality factor, containing the electromagnetic field to be measured, and placed between two other cavities $\mathbf{R}_1$ and $\mathbf{R}_2$.

A beam of excited atoms interact resonantly with the fields in $\mathbf{R}_1$ and $\mathbf{R}_2$, and dispersively with the field in the superconducting cavity. Before crossing the first cavity $\mathbf{R}_1$, the atoms are prepared in a circular Rydberg state, that is, a state with high principal quantum number and maximum angular momentum, so as to increase its decay time and increase the atom-field coupling. This state will be denoted by $|e\rangle$ in the following. The relevant atomic levels are shown in Fig. 19. Levels i and g are of the same parity, opposite to the parity of level e. The cavity mode, with angular frequency ω, is slightly detuned from the $e \rightarrow i$ transition, which corresponds to the angular frequency ω_{ie}.

The field in $\mathbf{R}_1$ induces resonant transitions between $|e\rangle$ and the lower-lying Rydberg state $|g\rangle$. One may assume that the interaction time and the intensity of the field in $\mathbf{R}_1$ (which is essentially classical, as is the field in $\mathbf{R}_2$) are such that the state of the atom, after it leaves $\mathbf{R}_1$, is $(|e\rangle + |g\rangle)/\sqrt{2}$ (one says then that the atom suffers a $\pi/2$ pulse).

The atom enters then the superconducting cavity, where due to the dispersive coupling between the field and the transition $e \rightarrow i$, the state $|e\rangle$ suffers a second-order Stark level shift, given by $-\hbar\Omega^2 n/\delta$, where Ω is the vacuum Rabi coupling between the atomic dipole

on the $e \rightarrow i$ transition and the cavity mode, and $\delta = \omega - |\omega_{eg}|$ is the frequency mismatch. It is assumed here for simplicity that the Rabi coupling is constant throughout the cavity. It is also assumed that δ is sufficiently large so that $\Omega^2 n/\delta^2 \ll 1$, and at the same time small compared to the difference in frequency between the $e \rightarrow i$ transition and all the other transitions in the Rydberg atom spectrum (specially the $e \rightarrow g$ one). In this case, only levels e and i are appreciably affected by the nonresonant atom-field coupling, which leaves the level g unperturbed.

If the field in the high-Q cavity is initially in the state $|\psi_0\rangle = \sum_n c_n|n\rangle$, the state of the atom-field system right after the atom exits that cavity is

$$|\Psi_1\rangle = \frac{1}{\sqrt{2}}\left(|e\rangle \sum_n c_n e^{-in\epsilon}|n\rangle + |g\rangle \sum_n c_n|n\rangle\right) \tag{110}$$

where $\epsilon = (\Omega^2/\delta)t_{\text{int}}$ is the one-photon phase shift that originates from the Stark energy shift. The atom crosses then the cavity $\mathbf{R}_2$, where it interacts with another $\pi/2$ pulse, which takes $|e\rangle$ again into $(|e\rangle + |g\rangle)/\sqrt{2}$ and $|g\rangle$ into $(-|e\rangle + |g\rangle)/\sqrt{2}$. The correlated atom-field state becomes:

$$|\Psi_2\rangle = \frac{1}{2}\left[|e\rangle \sum_n (e^{-in\epsilon} - 1)c_n|n\rangle + |g\rangle \sum_n (e^{in\epsilon} + 1)c_n|n\rangle\right] \tag{111}$$

After crossing $\mathbf{R}_2$, the atom is detected by letting it go through ionization plates, with an electric field in the form of an ascending ramp, so that the atom is ionized earlier if it is in state $|e\rangle$ and later if it is in state $|g\rangle$. If the atom is found in the state $|e\rangle$, then the atom-field system is described by the state

$$|\Psi_e\rangle = \frac{\sum_n e^{-in\epsilon/2} \sin(n\epsilon/2)c_n|n\rangle}{[\sum_n |c_n|^2 \sin^2(n\epsilon/2)]^{1/2}}|e\rangle \tag{112}$$

A similar expression applies if the atom is detected in state $|g\rangle$. Equation (112) shows that detection of the atom changes the field populations, even though the atom has not exchanged photons with the field (as the interaction is dispersive). The photon-number probability distribution previous to the passage of the atom is multiplied, after the detection, by a modulation factor that oscillates with n:

$$P_1^{(e)}(n) = \frac{\sin^2(n\epsilon/2)}{\sum_n P_0(n) \sin^2(n\epsilon/2)} P_0(n) \tag{113}$$

This modulation is displayed in Fig. 22. Depending on the choice for ϵ, detection of the first atom may dizimate several populations. For instance, for $\epsilon = \pi$, one gets $P_1^{(e)}(n) = 0$ for all the even values of n.

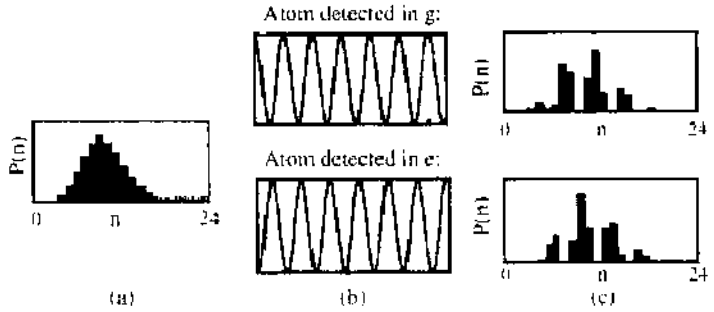


Figure 22. Modulation of the original photon-number distribution after the detection of the first atom. The initial distribution, displayed in (a), is Poissonian with $\bar{n} = 10$. It is multiplied by the oscillating fringe function $\sin^2(n\epsilon/2)$, if the atom is detected in the state e, or $\cos^2(n\epsilon/2)$, if the atom is detected in state g. These functions are represented as a function of n in (b). In the resulting distributions (c), photon numbers closest to the "dark fringes" have been decimated. Reprinted with permission from [87], M. Brune et al., *Phys. Rev. A* 45, 5193 (1992). © 1992, American Physical Society.

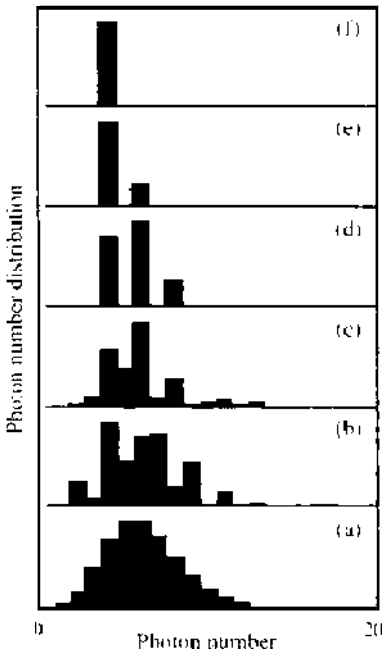


Figure 23. Typical QND sequence. The initial state is coherent with $\bar{n} = 5$, and its photon-number distribution is displayed in (a). Plots (b)–(f) correspond to the detection of 1, 3, 6, 10, and 15 atoms, respectively. In this realization, the field collapses, after detection of 15 atoms, in the $n = 3$ Fock state. Reprinted with permission from [87], M. Brune et al., *Phys. Rev. A* 45, 5193 (1992). © 1992, American Physical Society.

Detection of successive atoms, with changing velocities and therefore different values of ϵ, will dizimate other populations, until one is finally left with a single Fock state, which is stable under the transformation (113). Figure 23 illustrates a typical QND sequence. By repeating this experiment many times, starting with the same initial state, one finds each Fock state with a probability equal to the one in the initial distribution: one ends up therefore reconstructing the initial photon-number distribution.

This process can be thought of as a model of quantum measurement: the sequence of atoms can be seen as the macroscopic measuring device, and the process of dizimation as the gradual convergence of the state of the system toward an eigenstate of the observable being measured, in this case the photon number, to which the dispersive phase shift is proportional.

This process can be optimized by careful selection of the velocity of each atom, taking into account the result of the previous measurement. Thus, if the phase shift per photon associated with the first atom is π, then after detection of this atom only odd or even populations will remain. In case even populations remain, the velocity of the second atom is chosen so that the phase shift per photon is equal to $\pi/2$, then only populations with $n = 2k$, and k even or odd, will remain. Proceeding in this way, one can show that one reaches a Fock space after a number of atoms of the order of $\log_2 \Delta(n)$, where $\Delta(n)$ is the width of the photon-number distribution [108].

9. FULL CHARACTERIZATION OF THE STATE OF THE SYSTEM: MEASURING THE WIGNER DISTRIBUTION

One might wonder if it could be possible to get, from the experimental setup discussed before, a more complete information on the field in the cavity. This was shown to be indeed possible in Refs. [109, 110]: a slight modification of the experiment on decoherence leads to the reconstruction of the so-called Wigner distribution [111] of the field in the cavity, which provides a complete description of the quantum state of the field in phase space.

Phase-space probability distributions are very useful in classical statistical physics. Averages of relevant functions of the positions and momenta of the particles can be obtained by integrating these functions with those probability weights.

In quantum mechanics, similar averages are calculated by taking the trace of the product of the density operator that describes the system with the observable of interest. Heisenberg's

inequality forbids the existence in phase space of *bona fide* probability distributions, as one cannot determine simultaneously the position and the momentum of a particle. In spite of this, phase-space distributions may still play a useful role in quantum mechanics, allowing the calculation of the average of operator-valued functions of the position and momentum operators as classical-like integrals of c-number functions. These functions are associated to those operators through correspondence rules, which depend on a previously defined operator ordering.

From all phase space representations, the Wigner distribution is the most natural, when one looks for a quantum-mechanical analog of a classical probability distribution in phase space. This is a consequence of a beautiful result demonstrated by Bertrand and Bertrand [112], which will be presented here. The following account stays close to the ones in Refs. [112] and [113].

Let us look for a representation for which the *marginal distributions* coincide with the quadrature probability distributions:

$$\int dp W(q, p) = \langle q|\hat{\rho}|q\rangle, \qquad \int dq W(q, p) = \langle p|\hat{\rho}|p\rangle, \tag{114}$$

where $\hat{\rho}$ is the density operator for the system. One should note that from (114) it follows immediately the normalization property:

$$\int dp dq W(q, p) = 1 \tag{115}$$

Properties (114) must remain true if one rotates the axes in phase space, so that

$$\hat{q}_\theta = \hat{U}^\dagger(\theta)\hat{q}\hat{U}(\theta) = \hat{q}\cos\theta + \hat{p}\sin\theta \tag{116}$$

$$\hat{p}_\theta = \hat{U}^\dagger(\theta)\hat{p}\hat{U}(\theta) = -\hat{q}\sin\theta + \hat{p}\cos\theta \tag{117}$$

or, inversely,

$$\hat{q} = \hat{q}_\theta\cos\theta - \hat{p}_\theta\sin\theta \tag{118}$$

$$\hat{p} = \hat{q}_\theta\sin\theta + \hat{p}_\theta\cos\theta \tag{119}$$

Here $\hat{U}(\theta)$ is the rotation operator in phase space, given by Eq. (7). Thus:

$$P(q_\theta) = \int W(q_\theta\cos\theta - p_\theta\sin\theta, q_\theta\sin\theta + p_\theta\cos\theta)dp_\theta \tag{120}$$

where now

$$P(q_\theta) = \langle q|\hat{U}(\theta)\hat{\rho}\hat{U}^\dagger(\theta)|q\rangle \tag{121}$$

Expression (120), which yields the probability distribution for q_θ in terms of the function $W(q, p)$, is called a *Radon transform*. It was investigated in 1917 by the mathematician Johan Radon [114], who showed that, if one knows $P(q_\theta)$ for all angles θ, then one can uniquely recover $W(q, p)$, through the so-called *Radon inverse transform*. If one now identifies $P(q_\theta)$, given by Eq. (120), with the quantum expression given by Eq. (121), then it follows that Eqs. (120) and (121) uniquely determine the function $W(q, p)$, in terms of the density operator $\hat{\rho}$ of the system. The function $W(q, p)$ is in this case precisely the Wigner function of the system.

Before demonstrating this result, let us note that Radon's result is the mathematical basis of tomography. In fact, application of this procedure to medicine (Fig. 24) brought the Nobel Prize in Medicine to the physicist Allan Cormack and the electrical engineer Godfrey Hounsfield in 1979.

9.1. Reconstruction of the Wigner Function

It is shown now that the distribution $W(q, p)$ is uniquely determined from the knowledge of $P(q_\theta)$. For this end, let us introduce the *characteristic function* corresponding to $W(q, p)$,

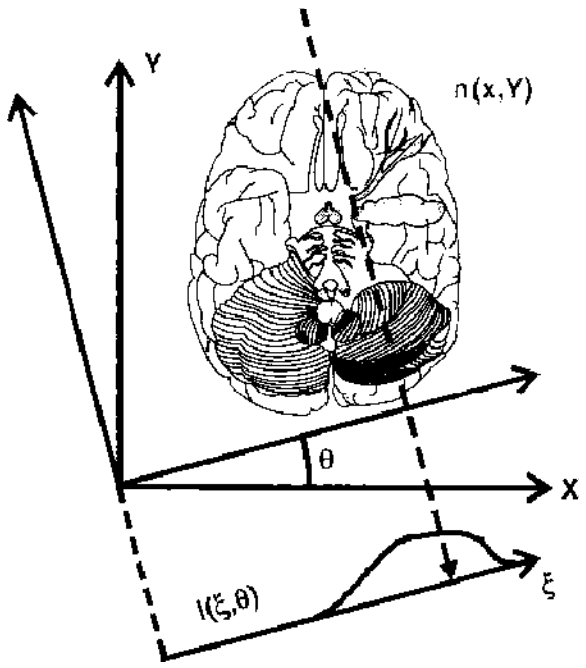


Figure 24. Medical tomography. Measurement of the x-ray absorption for all angles along a plane allows one to reconstruct the absorptive part of the refraction index for a slice of the organ under investigation.

which is just the Fourier transform of this distribution:

$$\tilde{W}(u, v) = \iint W(q, p) \exp(-iuq - ivp)\, dq\, dp \tag{122}$$

The characteristic function corresponding to $P(q_\theta)$ is introduced in a similar way:

$$\tilde{p}(\xi, \theta) = \int P(q_\theta) \exp(-i\xi q_\theta) dq_\theta \tag{123}$$

Inserting into Eq. (123) the expression for $P(q_\theta)$ as a function of W, given by Eq. (120), one gets:

$$\tilde{p}(\xi, \theta) = \iint W(q, p) \exp(-i\xi q_\theta) dq_\theta\, dp_\theta$$

where $q = q_\theta \cos\theta - p_\theta \sin\theta$ and $p = q_\theta \sin\theta + p_\theta \cos\theta$, and therefore $q_\theta = q\cos\theta + p\sin\theta$.

Changing the integration variables in $\tilde{p}(\xi, \theta)$, so that $(q_\theta, p_\theta) \to (q, p)$, one gets:

$$\tilde{p}(\xi, \theta) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} W(q, p) \exp[-i\xi(q\cos\theta + p\sin\theta)] dq\, dp \tag{124}$$

Therefore, $\tilde{p}(\xi, \theta)$ is the Fourier transform of $W(q, p)$ in polar coordinates:

$$\tilde{p}(\xi, \theta) = \tilde{W}(\xi\cos\theta, \xi\sin\theta)$$

which as we have seen is the characteristic function corresponding to $W(q, p)$. This implies that, from $P(q_\theta)$, one can calculate $\tilde{p}(\xi, \theta)$, and from this function one can calculate $W(q, p)$.

This demonstrates the tomographic reconstruction of $W(q, p)$. In order to connect this distribution to the density operator of the system, one uses Eq. (121):

$$\tilde{p}(\xi, \theta) = \int_{-\infty}^{+\infty} \langle q_\theta | \hat{\rho} | q_\theta \rangle e^{-i\xi q_\theta} dq_\theta = \int_{-\infty}^{+\infty} \langle q_\theta | \hat{\rho} e^{i\xi \hat{q}_\theta} | q_\theta \rangle dq_\theta = \mathrm{Tr}[\hat{\rho} e^{-i\xi \hat{q}_\theta}] \tag{125}$$

But $\xi \hat{q}_\theta = \hat{q}\xi\cos\theta + \hat{p}\xi\sin\theta$. Therefore, setting $u = \xi\cos\theta$, $v = \xi\sin\theta$, one gets, for the characteristic function,

$$\tilde{W}(u, v) = \mathrm{Tr}[\hat{\rho} e^{-iu\hat{q} - iv\hat{p}}] \tag{126}$$

Another form for $W(q, p)$ can be obtained in the following way. We rewrite the characteristic function $\widetilde{W}(u, v)$ as:

$$\widetilde{W}(u, v) = \int_{-\infty}^{+\infty} \langle q|\hat{\rho}e^{-iu\hat{q}-iv\hat{p}}|q\rangle dq = e^{iuv/2}\int_{-\infty}^{+\infty} \langle q|\hat{\rho}e^{-iu\hat{q}}|q+v\rangle dq$$
$$= \int_{-\infty}^{+\infty} e^{-iux}\left\langle x - \frac{v}{2}|\hat{\rho}|x + \frac{v}{2}\right\rangle dx \tag{127}$$

where we have used that

$$\exp(-iu\hat{q} - iv\hat{p}) = \exp\left(\frac{-iuv}{2}\right)\exp(-iu\hat{q})\exp(-iv\hat{p})$$

and we have set $q = x - v/2$. Taking the Fourier transform of $\widetilde{W}(u, v)$ given by Eq. (127), one gets the following expression for the distribution $W(q, p)$:

$$W(q, p) = \frac{1}{\pi}\int_{-\infty}^{+\infty} e^{-2ipx}\langle q + x|\hat{\rho}|q - x\rangle dx \tag{128}$$

which, except for a normalization constant, is the famous expression written down by Wigner [111] in his article "On the Quantum Correction for Thermodynamic Equilibrium," published in 1932. Wigner used this quasi-probability distribution in phase space as a convenient way of calculating quantum corrections to classical statistical mechanics. He wrote in his paper that the expression in Eq. (128) "was chosen from all possible expressions, because it seems to be the simplest." He added a quite intriguing footnote: "This expression was found by L. Szilard and the author some years ago for another purpose." One has shown here that the Wigner distribution has in fact a quite distinctive feature: it is the only distribution in phase space that yields the correct marginal distributions for any quadrature!

The tomographic procedure has a simple interpretation for a harmonic oscillator. From Eq. (10), it is clear that in this case measuring the quadratures for all angles is equivalent to measuring the position of the harmonic oscillator for all times from 0 to $2\pi/\omega$. This implies that the measurement of $|\psi(x, t)|^2$ for $0 < t \leq 2\pi/\omega$ allows one to reconstruct the state $\psi(x, t)$ of the harmonic oscillator.

A direct connection between the density matrix of the field and the tomographic probabilities was established by D'Ariano et al. [115]. Use of unbalanced homodyne detection to reconstruct the quantum state of the field was proposed in Ref. [116]. The use of tomographic techniques to reconstruct the vibrational state of trapped ions was proposed in Refs. [117] and [118].

The question about what is the minimum set of measurements needed to reconstruct the state of a system is actually a very old problem in quantum mechanics. In his article on quantum mechanics in the *Handbuch der Physik* in 1933 [119], Pauli stated that "the mathematical problem, as to whether for given functions $W(x)$ and $\widetilde{W}(p)$ [probability distributions in position and momentum space], the wave function ψ, if such a function exists, is always uniquely determined has still not been investigated in all its generality." One knows now the answer to this question: the probability distributions $W(x)$ and $\widetilde{W}(p)$ do not form a complete set in the tomographic sense and therefore are not sufficient to determine uniquely the quantum state of the system.

9.2. Expression of the Wigner Function in Terms of $\hat{a}$ and $\hat{a}^\dagger$

From Eq. (128), one may write

$$W(q, p) = \frac{1}{\pi}\int \langle q + x|\hat{\rho}|q - x\rangle e^{-2ipx}dx = \frac{1}{\pi}\int \langle x|e^{iq\hat{p}}\hat{\rho}e^{-iq\hat{p}}|-x\rangle e^{-2ipx}dx$$
$$= \frac{1}{\pi}\int \langle x|e^{-ip\hat{q}}e^{iq\hat{p}}\hat{\rho}e^{-iq\hat{p}}e^{ip\hat{q}}|-x\rangle dx = \frac{1}{\pi}\int \langle x|e^{-ip\hat{q}}e^{iq\hat{p}}\hat{\rho}e^{-iq\hat{p}}e^{ip\hat{q}}e^{i\pi\hat{a}^\dagger\hat{a}}|x\rangle dx$$
$$= \frac{1}{\pi}\mathrm{Tr}[\hat{D}^\dagger(\alpha, \alpha^*)\hat{\rho}\hat{D}(\alpha, \alpha^*)e^{i\pi\hat{a}^\dagger\hat{a}}] \tag{129}$$

where we have used the definition for the displacement operator $\hat{D}(\alpha, \alpha^*)$ and that $\exp(i\pi\hat{a}^\dagger\hat{a})$ is the parity operator.

The last expression defines the Wigner function in terms of the operators $\hat{a}$ and $\hat{a}^\dagger$. It shows that this distribution is proportional to the average of the displaced parity operator.

The Wigner function expressed in terms of $\hat{a}$ and $\hat{a}^\dagger$ involves usually a different normalization with respect to the one defined before: one must set $W \to 2\pi W$, so that

$$\int\left(\frac{d^2\alpha}{\pi}\right)W(\alpha, \alpha^*) = 1 \tag{130}$$

We have then [120]:

$$W(\alpha, \alpha^*) = 2\mathrm{Tr}[\hat{D}^\dagger(\alpha, \alpha^*)\hat{\rho}\hat{D}(\alpha, \alpha^*)e^{i\pi\hat{a}^\dagger\hat{a}}] \tag{131}$$

9.3. Properties of the Wigner Distribution

Thorough discussions of properties of the Wigner distribution can be found in Refs. [113] and [121]. Here only some of them are summarized.

It is easy to show that the Wigner function is real and bounded. If one adopts the normalization of Eq. (128), so that Eq. (115) holds, then Schwarz's inequality implies that $|W(q, p)| \leq 1/\pi$. If the Cahill–Glauber normalization in Eq. (131) is adopted instead, so that Eq. (130) is satisfied, one has

$$|W(\alpha, \alpha^*)| \leq 2 \tag{132}$$

Furthermore, let $W_\psi(q, p)$ be the Wigner function corresponding to the state $\psi(q)$, and $W_\phi(q, p)$ the Wigner function corresponding to the state $\phi(q)$, as given by Eq. (128). Then,

$$\left|\int dq\psi^*(q)\phi(q)\right|^2 = 2\pi\int dq\int dpW_\psi W_\phi$$

This relation has several consequences. First, setting $\psi(q) = \phi(q)$, one gets $\int dq\int dp[W(q, p)]^2 = 1/2\pi$. More generally, it is easy to show that

$$\mathrm{Tr}\,(\hat{\rho}^2) = 2\pi\iint dq\,dp[W(q, p)]^2 \tag{133}$$

and therefore $\iint dq\,dp[W(q, p)]^2 < 1/2\pi$ for a statistical mixture. It is also clear that

$$\int dq\int dpW_\psi W_\phi = 0 \quad \text{if } \langle\psi|\phi\rangle = 0 \tag{134}$$

which implies that W cannot be always positive. This may be thought as a consequence of the Heisenberg inequalities: as it is not possible to measure simultaneously q and p, one cannot have in quantum mechanics *bona fide* probability distributions in phase space. In fact, one can show that the only pure states leading to positive-definite Wigner functions correspond to Gaussian wave functions [122]. This is the case for coherent states, and also for squeezed states, which are obtained from coherent states by a scale transformation of two orthogonal quadrature axes, amounting to multiplying one of the quadratures by a constant factor and dividing the other quadrature by the same factor. One gets then a squeezed Gaussian as the Wigner distribution corresponding to a squeezed state.

One should note that the Husimi or Q distribution [123], often found in the literature, and defined by

$$Q(\alpha, \alpha^*) = \frac{1}{\pi}\langle\alpha|\hat{\rho}|\alpha\rangle \tag{135}$$

is always positive, but does not lead to the correct marginal distributions. It is easy to show that

$$Q(\alpha, \alpha^*) = \frac{2}{\pi^2}\int d^2\lambda e^{-2|\lambda-\alpha|^2}W(\lambda, \lambda^*) \tag{136}$$

so that the Husimi distribution is obtained by smoothing out the Wigner distribution. Although this may bring some mathematical convenience, it makes it harder to uncover quantum features, from the experimental point-of-view.

Equation (136) shows that the Husimi distribution can be obtained from the Wigner function. The converse is also true, as can be seen from the convolution character of Eq. (136): the Fourier transform of the Husimi distribution (that is, the corresponding characteristic function) is equal to the Fourier transform of the Wigner function, as given by Eq. (122), multiplied by $\exp[-(u^2+v^2)/4]$.

9.4. Averages of Operators

As shown by Moyal in 1949 [124], the Wigner distribution can be used to calculate averages of symmetric operator functions of q and p, as classical-like integrals in phase space. Thus, for instance,

$$\mathrm{Tr}\,(\hat{\rho}\{\hat{q}^2\hat{p}\}_{\mathrm{sim}}) = \mathrm{Tr}\,[\hat{\rho}(\hat{q}^2\hat{p}+\hat{q}\hat{p}\hat{q}+\hat{p}\hat{q}^2)/3] = \int dq\,dp W(qp)q^2p \tag{137}$$

The association of a symmetrized quantum operator to a classical function is called *Weyl correspondence*.

This property of the Wigner function can be shown by considering the two equivalent expressions for the characteristic function $\widetilde{W}(u, v)$,

$$\widetilde{W}(u, v) = \mathrm{Tr}\,[\hat{\rho}e^{-iu\hat{q}-iv\hat{p}}] \tag{138}$$

$$\widetilde{W}(u, v) = \iint W(q, p)\exp(-iuq - ivp)dq\,dp \tag{139}$$

from which one gets

$$\mathrm{Tr}\,[\hat{\rho}(\mu\hat{q}+\nu\hat{p})^k] = i^k\frac{\partial^k}{\partial\xi^k}\widetilde{W}(\xi\mu, \xi\nu)\bigg|_{\xi=0} = \int_{-\infty}^{+\infty} W(q, p)(\mu q+\nu p)^k dq\,dp \tag{140}$$

Comparing powers of μ and ν, one gets

$$\mathrm{Tr}\,(\hat{\rho}\{\hat{q}^m\hat{p}^n\}_{\mathrm{sim}}) = \int_{-\infty}^{+\infty} W(q, p)q^m p^n dq\,dp$$

Of course, the same property holds for the Wigner function expressed in terms of $\hat{a}$ and $\hat{a}^\dagger$. Thus, for instance,

$$\mathrm{Tr}\,[\hat{\rho}(\hat{a}\hat{a}^\dagger+\hat{a}^\dagger\hat{a})/2] = \int\left(\frac{d^2\alpha}{\pi}\right)\alpha\alpha^* W(\alpha, \alpha^*) \tag{141}$$

Other distributions in phase space can be introduced, which allow writing as classical-like integrals averages of functions of the operators $\hat{a}$ and $\hat{a}^\dagger$ written in normal and antinormal order. Thus, for instance, the Husimi distribution can be shown to correspond to operators in antinormal order. These distributions will not be discussed here [125–130].

9.5. Examples of Wigner Functions

Wigner functions corresponding to special states of the electromagnetic field can be obtained from either Eq. (128) or Eq. (131). We adopt in this section the normalization corresponding to Eq. (128).

For the vacuum state, one has the Gaussian

$$W_0(q, p) = \frac{1}{\pi}e^{-q^2-p^2} \tag{142}$$

The Wigner function for a coherent state can be easily obtained by applying a displacement to the above Gaussian:

$$W_{\alpha}(q, p) = \frac{1}{\pi} e^{-(q-q_0)^2-(p-p_0)^2} \tag{143}$$

Application of a scaling transformation to Eq. (142) yields the Wigner function for a squeezed vacuum, plotted in Fig. 25a:

$$W_{\xi}(q, p) = \frac{1}{\pi} \exp(-e^{2t}q^2 - e^{-2t}p^2) \tag{144}$$

For a one-photon Fock state $|1\rangle$, one gets:

$$W_1(q, p) = \frac{1}{\pi} e^{-q^2-p^2}(2q^2 + 2p^2 - 1) \tag{145}$$

This function vanishes for $\sqrt{q^2 + p^2} = 1/\sqrt{2}$ and is negative at the origin of phase space. This negative value reminds us of the highly nonclassical nature of a Fock state. For higher photon numbers, the Wigner function displays more oscillations, the number of zeros coinciding with n. Figure 25b displays the Wigner function corresponding to a Fock state with $n = 3$.

Of special interest for our discussion on the classical limit of quantum mechanics is the state formed by superimposing two coherent states $|\alpha_0\rangle$ and $|-\alpha_0\rangle$ (setting $\alpha_0 = q_0$ real for simplicity):

$$|\psi\rangle = \mathcal{N}[|\alpha_0\rangle + |-\alpha_0\rangle] \tag{146}$$

where $\mathcal{N}$ is a normalization constant, given by

$$\mathcal{N} = [2(1 + \exp(-2|\alpha_0|^2))]^{-1/2} \tag{147}$$

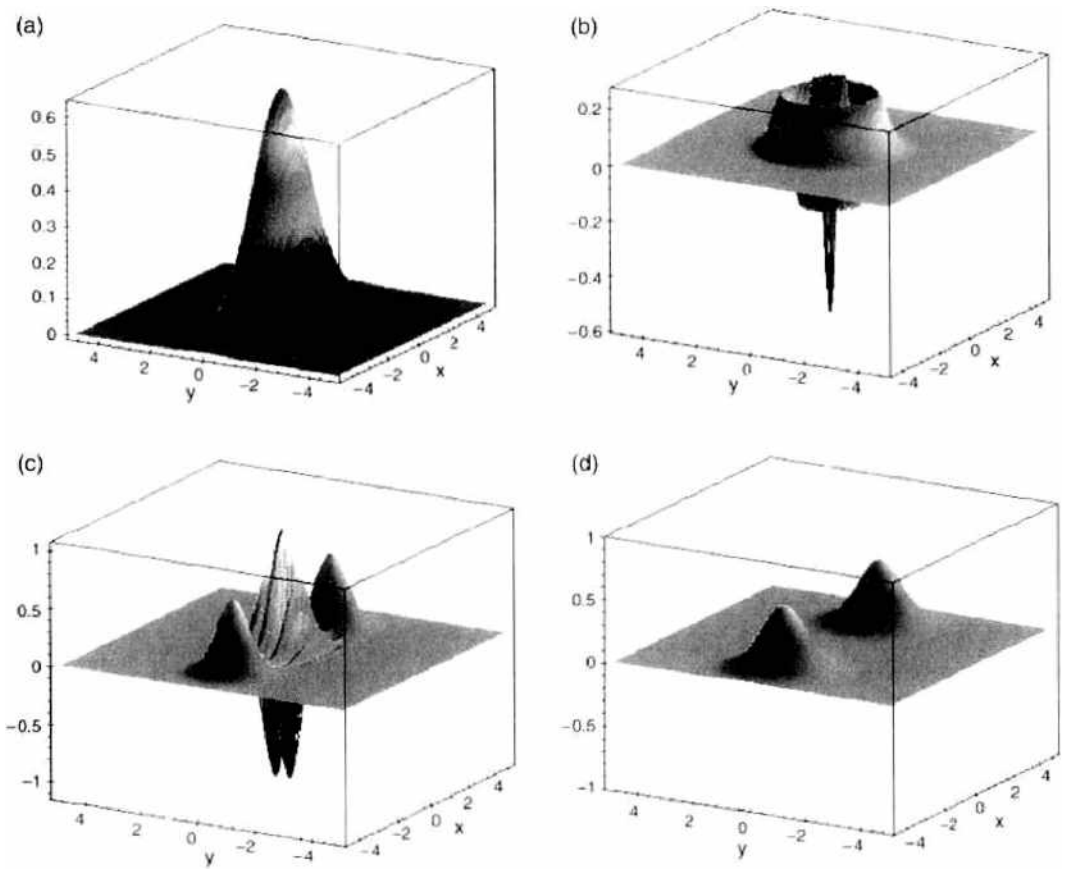


Figure 25. Examples of Wigner distributions. (a) Squeezed state; (b) Fock state with $n = 3$; (c) superposition of two coherent states, $|\psi\rangle \propto |\alpha_0\rangle - |-\alpha_0\rangle$, with $\alpha_0 = 3$; (d) statistical mixture $(|\alpha_0\rangle\langle\alpha_0| + |-\alpha_0\rangle\langle-\alpha_0|)/2$, also with $\alpha_0 = 3$.

The corresponding wave function in configuration space is given by

$$\psi(q) = \mathcal{N}\left[e^{-(q-q_0)^2/2} + e^{-(q+q_0)^2/2}\right]$$

whereas the Wigner function is

$$W(q, p) = \left(\frac{\mathcal{N}}{\pi}\right)\left[e^{-(q-q_0)^2-p^2} + e^{-(q+q_0)^2-p^2} + 2e^{-q^2-p^2}\cos(2pq_0)\right] \tag{148}$$

This function is displayed in Fig. 25c.

It is interesting to compare this Wigner function with the one corresponding to a statistical mixture of the same coherent states, with equal weights:

$$\hat{\rho} = \frac{1}{2}(|\alpha_0\rangle\langle\alpha_0| + |-\alpha_0\rangle\langle-\alpha_0|) \tag{149}$$

for which

$$W(q, p) = \left(\frac{1}{2\pi}\right)\left[e^{-(q-q_0)^2-p^2} + e^{-(q+q_0)^2-p^2}\right] \tag{150}$$

This function is displayed in Fig. 25d.

Although the expression in Eq. (150) is just the sum of two Gaussians, corresponding to the coherent states $|\alpha_0\rangle$ and $|-\alpha_0\rangle$, respectively, the one in Eq. (148) displays interference fringes around the origin of phase space, which is a clear signature of the coherence between the two states $|\alpha_0\rangle$ and $|-\alpha_0\rangle$ in Eq. (146). Therefore, the measurement of the Wigner function of the electromagnetic field would be a clear-cut way of distinguishing between a coherent superposition and a mixture of the two coherent states. As shown in Section 7.2, these coherent states may be interpreted, within the framework of recent experiments in cavity QED, as pointers of a measuring apparatus. The mechanism by which a state like the one in Eq. (146) loses its coherence, approaching the state in Eq. (149), is thus very relevant for the quantum theory of measurement.

9.6. Measurement of the Wigner Function

We discuss two methods that have been used to reconstruct Wigner functions of electromagnetic fields in this Section.

9.6.1. Quantum Tomography

The inverse Radon transform suggests that the Wigner function of an electromagnetic field can be reconstructed by determining the probability distribution for the generalized quadrature q_θ, $P(q_\theta)$, through homodyne detection [131]. As discussed in the section on homodyne measurement, the angle θ is changed by varying the phase difference between the local oscillator and the measured field. For each θ, the resulting intensity is measured many times, so as to build enough statistics, thus leading to the measurement of $P(q_\theta)$. As this angle is continuously changed between 0 and 2π, the changing distribution $P(q_\theta)$ mimics the time-dependent probability distribution for a harmonic oscillator.

The first measurements were made in 1993 by Smithey et al. [132]. In view of the low detection efficiency in those experiments, the detected distribution was actually a smoothed version of the Wigner function, closely related to the Husimi distribution. A much better result was achieved by Mlynek's group in 1995 [133, 134], clearly displaying a highly compressed Gaussian, corresponding to the experimentally obtained Wigner function of a squeezed state of light emerging from an optical parametric oscillator. A procedure closely related to the homodyne detection method was used to reconstruct the vibrational state of a molecule by T. J. Dunn et al. [135].

The results obtained by Mlynek's group are displayed in Fig. 26. The first column exhibits the noisy current, resulting from the homodyning procedure, measured for each angle θ (in the actual experiment, this angle is swept by 2π in approximately 200 ms; the horizontal axis represents the sweeping time). The second column stands for the probability distribution

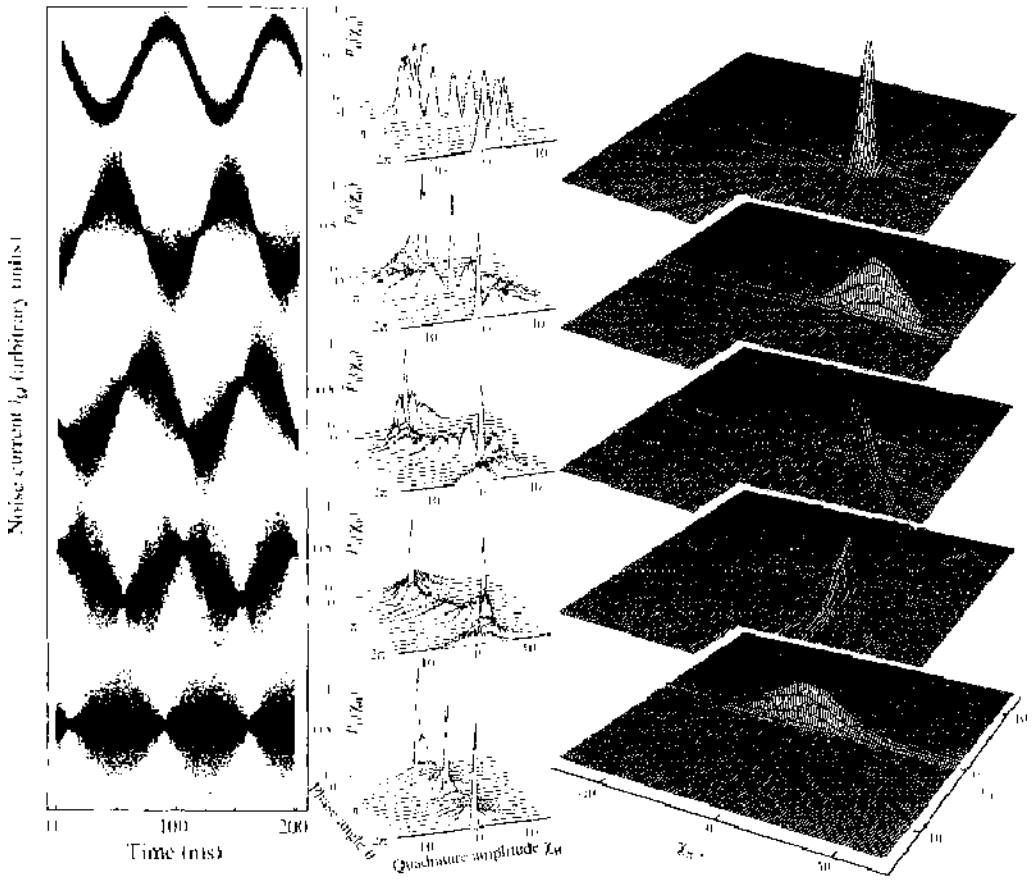


Figure 26. Noise traces in the current (left), quadrature distribution (center), and reconstructed Wigner functions (right) of several generated quantum states. From the top: coherent state, phase-squeezed state, state squeezed in the 48° quadrature, amplitude-squeezed state, squeezed vacuum state. Reprinted with permission from [134], G. Breitenbach et al., *Nature* 387, 471 (1997). © 1997, Nature Publishing Group.

of the quadrature as a function of θ. As mentioned before, it could also be interpreted as the time-dependent behavior of the probability distribution for a harmonic oscillator, with period approximately equal to 200 ms. Finally, the third column exhibits the corresponding Wigner distributions, obtained by applying the inverse Radon transform.

9.6.2. Measurement of Displaced Populations

The Wigner function can also be obtained by measuring the populations of displaced states. Indeed, from Eq. (131), one has:

$$W(\alpha,\alpha^*) = 2\mathrm{Tr}\left[\hat{\rho}\hat{D}(\alpha,\alpha^*)e^{i\pi\hat{a}^\dagger\hat{a}}\hat{D}^{-1}(\alpha,\alpha^*)\right] = 2\sum_n \langle n|\hat{D}^{-1}(\alpha,\alpha^*)\hat{\rho}\hat{D}(\alpha,\alpha^*)e^{i\pi\hat{a}^\dagger\hat{a}}|n\rangle$$

$$= 2\sum_n (-1)^n \langle n|\hat{D}^{-1}(\alpha,\alpha^*)\hat{\rho}\hat{D}(\alpha,\alpha^*)|n\rangle = 2\sum_n (-1)^n P_n(-\alpha,-\alpha^*) \qquad (151)$$

where $P_n(-\alpha,-\alpha^*)$ is the probability of finding n photons in the field after the field is displaced by $-\alpha$. This displacement can be implemented for instance, for a field in a cavity, by injecting a coherent state into the cavity, through the coupling of the cavity with a microwave generator (if the frequency of the mode in the cavity is in the microwave range). In order to determine the Wigner function by this method, one must measure then the population of the displaced states, which can be done for instance by applying the procedure described in Section 8.

This method was used by Wineland's group at NIST to measure the Wigner function of vibrational states of a trapped $^9Be^+$ ion [136]. The relevant level scheme is shown in Fig. 27. States $|\downarrow\rangle$ and $|\uparrow\rangle$ correspond to two ground-state hyperfine sublevels ($^2S_{1/2}$, with $F = 2$, $m_F = -2$ and $F = 1$, $m_F = -1$, respectively), separated by $\hbar\omega_{HF}$. The ion is trapped in a

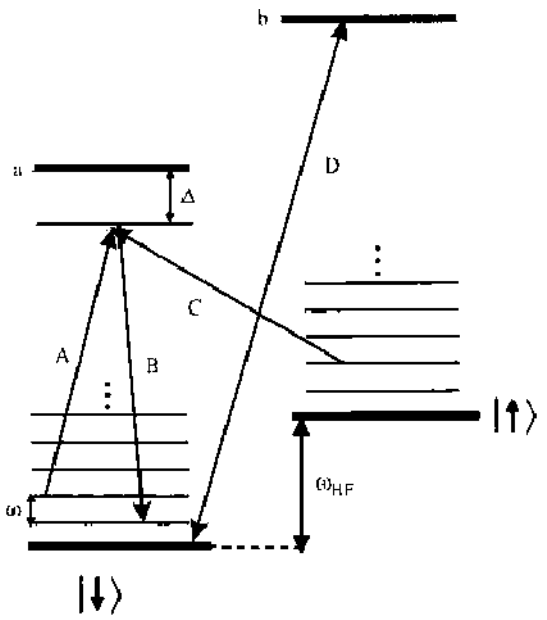


Figure 27. Measurement of the Wigner function for a trapped ion, by Leibfried et al. Displacement of the vibrational state associated with the internal state $|\downarrow\rangle$ (a hyperfine structure sublevel) is achieved by applying fields A and B, which induce transitions between neighboring vibrational states corresponding to the electronic level $|\downarrow\rangle$, without changing the ion's internal state. Population of displaced states is measured by inducing a resonant exchange during a time t between states $|\downarrow, n\rangle$ and $|\uparrow, n+1\rangle$ with fields B and C (turning off field A). Probability of finding the atom in $|\downarrow\rangle$ as a function of t, determined by exciting it to level b (with a high fluorescence yield), leads to information on population of displaced vibrational states.

harmonic potential, and the vibrational levels associated with each electronic state $|\downarrow\rangle$ and $|\uparrow\rangle$ are also sketched in Fig. 27.

Initially, the ion is in the internal state $|\downarrow\rangle$. The displacement of the vibrational state in phase space is obtained by inducing a Raman transition between neighboring vibrational states, when the internal state of the ion is $|\downarrow\rangle$. This is accomplished by applying the two fields A and B illustrated in Fig. 27, with a frequency difference equal to the vibrational frequency ω. Beam B is circularly polarized (σ_-), and does not couple $|\uparrow\rangle$ to any virtual $^2P_{1/2}$ state, so that only the motional state correlated with $|\downarrow\rangle$ is displaced. These fields do not lead to transitions between electronic levels of the trapped ion, as they are detuned with respect to the possible electronic transitions, and therefore they affect only the center-of-mass motion. The action of the two fields can thus be modeled by an effective Hamiltonian of the form $H \propto (\hat{a} + \hat{a}^\dagger)$, where $\hat{a}$ and $\hat{a}^\dagger$ are harmonic oscillator lowering and raising operators. The evolution operator corresponding to this Hamiltonian is precisely the displacement operator, therefore the Raman process induces a displacement of the original state in phase space.

A resonant exchange between states $|\downarrow\rangle|n\rangle$ and $|\uparrow\rangle|n+1\rangle$ is then induced for a time t, with fields B and C (turning off field A). For each time t and each displacement α the population $P_\downarrow(t, \alpha)$ of the $|\downarrow\rangle$ state is measured. A fourth level b is used for detecting the electronic state of the ion (and also for Doppler precooling): a circularly polarized pulse D resonant with the $|\downarrow\rangle \leftrightarrow |b\rangle$ transition leads to a fluorescence signal if the ion is in $|\downarrow\rangle$, and the absence of fluorescence implies that the ion is in $|\uparrow\rangle$, which is not coupled to level b due to angular momentum selection rules (the detection efficiency for this process is close to 100%). The internal state at $t = 0$ being always equal to $|\downarrow\rangle$, the signal averaged over many measurements is

$$P_\downarrow(t, \alpha) = \frac{1}{2}\left[1 + \sum_{n=0}^{\infty} P_n(\alpha)\cos(2g\sqrt{n+1}\tau)e^{-\gamma_n \tau}\right] \tag{152}$$

where $\tau = gt$ is a dimensionless time, g is the coupling constant between the ion and the external resonant fields, γ_n the experimentally determined decay constants (in units of the coupling constant g), and $P_n(\alpha) = \langle n|\widehat{D}^\dagger(\alpha)\hat{\rho}\widehat{D}(\alpha)|n\rangle$ is the population distribution of

the displaced state. By fitting the measured $P_\downarrow(t, \alpha)$ with the above expansion, one can determine $P_n(\alpha)$ [136], and from $P_n(\alpha)$ one determines the Wigner function through Eq. (151).

Figure 28 displays the Wigner function measured for the center-of-mass state of a trapped ion in the $n = 1$ eigenstate of the harmonic-oscillator trapping potential. The corresponding density matrix, in the harmonic oscillator basis, is also displayed in the same figure, confirming that the only state with appreciable population is the $n = 1$ eigenstate.

Similar techniques have been proposed for the reconstruction of the Wigner function of a running-wave field through photon-counting [137–139], by mixing the field to be measured with a coherent field through a beam splitter. This is equivalent to displacing the field in phase space.

The reconstruction of the Wigner function of an intracavity field, by measuring the photon-number through the atoms that have interacted with the field, was proposed in Ref. [140].

The determination of the coefficients $P_n(\alpha)$ in the expansion given by Eq. (152) can also be made, when dissipation is neglected, by using the Fresnel representation of the Wigner function [141]. One uses then the integral relation

$$\frac{2}{\pi\sqrt{i}}\int_0^\infty d\tau \exp\left(\frac{i\tau^2}{\pi}\right)\cos(2\sqrt{n\tau}) = (-1)^n \tag{153}$$

so that, in the absence of dissipation, one gets from Eq. (152):

$$\frac{2}{\pi\sqrt{i}}\int_0^\infty d\tau \exp\left(\frac{i\tau^2}{\pi}\right)\left[P_\downarrow(t, -\alpha) - \frac{1}{2}\right] = \frac{1}{2}\sum_{n=0}^{\infty}(-1)^n P_n(-\alpha) = \frac{1}{4}W(\alpha, \alpha^*) \tag{154}$$

The Wigner function is thus determined, at each phase space point α, by a weighted time integral of the atomic dynamics due to the applied fields, with the vibrational state displaced by $-\alpha$, the weight function being the Fresnel phase factor $\exp(i\tau^2/\pi)$.

One should note, however, that the infinite-time integration in Eq. (153) is impossible to achieve in practice, as it would involve knowing $P_\downarrow(t, -\alpha)$ for all interaction times. In Ref. [141], the authors discuss the degree of approximation obtained when the integration time remains finite. Due to the oscillatory character of the kernel, the convergence is slow: for an upper limit of integration $\tau = 4\pi$, an error of about 10% is obtained, in the determination of the Wigner function of an ion in the first excited vibrational level.

It is clear that the above-described methods are highly indirect. It will be shown in the following, however, that it is possible to conceive a much more direct method for measuring

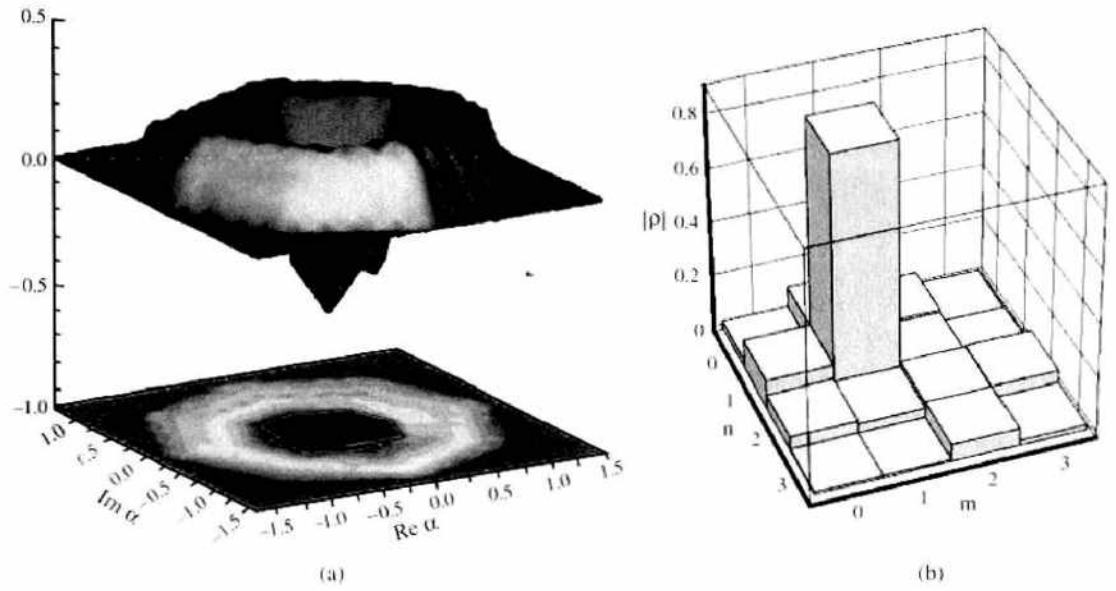


Figure 28. (a) Experimentally determined Wigner function corresponding to the first excited vibrational state of a trapped ion, as measured by Wineland et al. in NIST; (b) corresponding density matrix. Reprinted with permission from [135], D. Leibfried et al., *Phys. Rev. Lett.* 77, 4281 (1996). © 1996 American Physical Society.

the Wigner function at any point in phase space, for either an electromagnetic field in a cavity or a trapped ion.

10. DIRECT MEASUREMENT OF THE WIGNER FUNCTION

Once the proper state of the field is produced in the cavity, how would one be able to measure it? As shown in Refs. [109, 110], it is actually possible to measure the Wigner function of the field by a relatively simple scheme, which provides directly the value of the Wigner function at any point of phase space. This is in contrast with the tomographic procedure, or the method based on the measurement of populations, which yield the Wigner function only after some integration or summation. Furthermore, and also in contrast with those methods, the present scheme is not sensitive to detection efficiency, as long as one atom is detected within a time shorter than the decoherence time. A similar procedure can be applied to the reconstruction of the vibrational state of a trapped ion [109], and also in some cases to molecules [142]. We will discuss here only the application to the electromagnetic field.

10.1. Experimental Scheme

The basic experimental scheme for measuring the Wigner function [109, 110] coincides with the one used to produce the "Schrödinger cat"-like state, illustrated in Fig. 18. A high-Q superconducting cavity **C** is placed between two low-Q cavities ($\mathbf{R}_1$ and $\mathbf{R}_2$ in Fig. 18). The cavities $\mathbf{R}_1$ and $\mathbf{R}_2$ are connected to the same microwave generator, the field in $\mathbf{R}_2$ being dephased by η with respect to the field in $\mathbf{R}_1$. Another microwave source is connected to **C**, allowing the injection of a coherent state in this cavity, so that the density operator $\hat{\rho}$ of the field to be measured is transformed into $\hat{\rho}' = \hat{D}(\alpha, \alpha^*)\hat{\rho}\hat{D}^{-1}(\alpha, \alpha^*)$. This system is crossed by a velocity-selected atomic beam, such that an atomic transition $e \leftrightarrow g$ is resonant with the fields in $\mathbf{R}_1$ and $\mathbf{R}_2$, while another transition $e \leftrightarrow i$ is quasi-resonant (detuning δ) with the field in **C**, so that the atom interacts dispersively with this field if it is in state e, whereas no interaction takes place in **C** if the atom is in state g. The relevant level scheme is shown in Fig. 19. Just before $\mathbf{R}_1$, the atoms are promoted to the highly excited circular Rydberg state $|e\rangle$ (typical principal quantum numbers of the order of 50, corresponding to lifetimes of the order of some milliseconds). As each atom crosses the low-Q cavities, it sees a $\pi/2$ pulse, so that $|e\rangle \rightarrow [|e\rangle + \exp(i\eta)|g\rangle]/\sqrt{2}$, and $|g\rangle \rightarrow [-\exp(-i\eta)|e\rangle + |g\rangle]/\sqrt{2}$, with $\eta = 0$ in $\mathbf{R}_1$. If the atom is in state e when crossing **C**, there is an energy shift of the atom-field system (Stark shift), which dephases the field, after an effective interaction time t_{int} between the atom and the cavity mode. The one-photon phase shift is given by $\phi = (\Omega_0^2/4\delta)t_{\text{int}}$, where the resonant Rabi frequency Ω_0 measures the coupling between the atom and the cavity mode.

The atom is detected and the experiment is repeated many times, for each amplitude and phase of the injected field α, starting from the same initial state of the field $\hat{\rho}$. In this way, the probabilities P_e and P_g of detecting the probe atom in states e or g are determined.

Then [110],

$$P_g - P_e = \Re e\{e^{i\eta}\mathrm{Tr}[\hat{D}(\alpha, \alpha^*)\hat{\rho}\hat{D}^{-1}(\alpha, \alpha^*)e^{i\phi a^\dagger a}]\} \tag{155}$$

Setting $\eta = 0$ and $\phi = \pi$, we can see from (131) that

$$P_g - P_e = W\frac{(-\alpha, -\alpha^*)}{2} \tag{156}$$

Therefore, the difference between the two probabilities yields a direct measurement of the Wigner function!

We show now a simple derivation of this result. Let $\hat{\mathcal{T}}_e(\phi) = \exp(i\phi\hat{a}^\dagger\hat{a})$ be the phase shift operator associated with level e. The field is displaced so that $\hat{\rho} \rightarrow \hat{\rho}'$, and the atom, prepared in state e, crosses $\mathbf{R}_1$, **C**, and $\mathbf{R}_2$. The entangled atom-field state becomes:

$$\hat{\rho}_{\text{atom+field}} = |e\rangle\langle e| \otimes (\hat{\mathcal{T}}_e - e^{i\eta}\hat{1})\hat{\rho}'(\hat{\mathcal{T}}_e^\dagger - e^{-i\eta}\hat{1}) + |g\rangle\langle g| \otimes (\hat{1} + e^{i\eta}\hat{\mathcal{T}}_e)\hat{\rho}'(\hat{1} + e^{-i\eta}\hat{\mathcal{T}}_e^\dagger)$$
$$+ \text{terms non-diagonal in atomic space} \tag{157}$$

Calculating from this expression the probability $P_e = \mathrm{Tr}_{\text{atom+field}}[|e\rangle\langle e|\hat{\rho}_{\text{atom+field}}]$, and analogously for P_g, one gets immediately Eq. (156).

One may notice that the experimental procedure discussed above amounts to implementing experimentally on the state to be reconstructed the two operations explicitly represented in Eq. (131): the displacement operation (implemented through the injection of the coherent microwave field) and the parity operation (implemented through the conditional π-phase shift). In particular, the distribution in Eq. (156) clearly satisfies (132), as $|P_g - P_e| \leq 2$.

An important feature of this scheme is the insensitivity to the detection efficiency of the atomic counters (of the order of $40 \pm 15\%$ in recent experiments [87]). Indeed, if an atom is not detected after interacting with the cavity mode, the next atom will find a field described by the reduced density operator obtained from the entangled atom-field density matrix by tracing out the atomic states: $\hat{\rho}' \rightarrow \hat{\rho}'' = 1/2(\hat{\rho}' + \hat{J}_e\hat{\rho}'\hat{J}_e^{\dagger})$, where $\hat{J}_e(\phi) = \exp(i\phi\hat{a}^{\dagger}\hat{a})$ is the phase shift operator associated with level e. The value of $P_g - P_e$ for this second atom is then easily shown to reduce to Eq. (156).

The measurement accuracy does depend, however, on the detector's selectivity, that is, the ability to distinguish between the two atomic states. Another possible source of error is the velocity spread of the atomic beam, which would produce an uncertainty in the angle ϕ and in the angles of rotation in $\mathbf{R}_1$ and $\mathbf{R}_2$. For a 1% velocity spread and for average photon numbers of the order of 10, one can show that the distortion is at most equal to 0.04, in the relevant region of phase space, so that the measured distribution is practically indistinguishable from the true one. In fact, the insensitivity of the proposed scheme to the detection efficiency allows a passive selection of atomic velocity (only the atom which goes through the detectors at the right time after excitation is detected), which can be made with high precision.

One should note that this method allows the measurement of the Wigner function at each time t, as for each realization it involves the measurement of a single atom. It allows therefore the monitoring of the decoherence process "in real-time." This should be contrasted with the photon-counting measurement, as described in Section 8, which involves the measurement of many atoms in each realization.

It is interesting, in this respect, to compare the procedure described above with the one suggested by Davidovich et al. [94], with the objective of observing the decoherence of a Schrödinger cat-like state. In that reference, as discussed in Section 7.2, it was proposed that the decoherence of the state $|\pm\rangle = (|\alpha\rangle \pm |-\alpha\rangle)/N_{\pm}$ could be observed by measuring the joint probability of detecting in states $|e\rangle$ or $|g\rangle$ a pair of atoms, both prepared in the state $|e\rangle$ initially, and sent through the system depicted in Fig. 18. The atomic configuration considered in that reference coincides with the one adopted here. Detection of the first atom prepares the coherent superposition of coherent states, as described above. Detection of the second atom probes the state produced in **C**. Because no field was injected into the cavity between the two atoms, it is clear now that the experiment proposed in Ref. [94] amounts to a measurement of the Wigner function at the origin, which is nonzero for the pure state $|\pm\rangle$, as shown in Fig. 25c, vanishes after the decoherence time (shorter than the intensity decay time by the factor $2|\alpha|^2$), as shown in Fig. 25d, and increases again as dissipation takes place, bringing the field to the vacuum state.

In the experiment realized by Brune et al. [87], both $|e\rangle$ and $|g\rangle$ lead to dephasings (in opposite directions) of the field in **C**. In this case, it is easy to show that the Wigner function is again recovered, as long as the one-photon phase shift is $\phi = \pi/2$ (with opposite signs for e and g), and a dephasing $\eta = \pi/2$ is applied to the second Ramsey zone [109]. One gets then Eq. (156), with a minus sign added to the right-hand side. This condition was not satisfied, however, in the experiment reported in Ref. [87]: due to experimental limitations, the angle ϕ was actually smaller than $\pi/2$.

Getting a π phase shift per photon imposes stringent conditions on the experiment. The interaction time between the atom and the cavity field should be large enough, which implies using slow atoms, with a precisely controlled speed. Furthermore, the interaction time between the atom and the cavity field should be much smaller than the dampening time of the field in the cavity, and therefore a very good cavity is required.

10.2. Experimental Measurements of Negative Values for the Wigner Function

An easier task consists in measuring the value of the Wigner function at the origin of phase space when one knows beforehand that the field in the cavity contains at most one photon. In this case, one does not need to inject a field into the cavity, and the dispersive interaction leading to the phase shift of the field can be replaced by a resonant 2π interaction between levels e and i (see Fig. 14). As we have seen before, this interaction takes the atom from state e to state i and then back to state e, if there is one photon in the field, the state changing sign under this transformation: $|e\rangle|1\rangle_{\text{field}} \rightarrow -|e\rangle|1\rangle_{\text{field}}$. On the other hand, nothing happens if the atom is in state g or if there is no photon in the field. The conditional one-photon π phase shift is thus accomplished in this case with a resonant interaction, which requires an interaction time much shorter than the dispersive case. This idea was implemented in an experiment at Ecole Normale Supérieure, in Paris [143, 144]. The one-photon state was produced by sending an excited atom through the empty cavity, where the atom suffers a π transition, leaving one photon in the cavity, from which it exits in the state g. This was the first time a negative value was measured for the Wigner function of an electromagnetic field, namely the value at the origin of the Wigner function corresponding to a one-photon state [this distribution is shown in Fig. 29a]. The experimentally obtained value was -1.32, to be compared with the theoretical value -2, associated with the normalization defined by Eq. (130). The smaller experimental value is due to a contamination of the vacuum state, due to the possible decay of the photon in the cavity in the time interval between preparation and measurement.

The full Wigner function corresponding to a running-wave one-photon field was measured by Lvovsky et al. [145]. The single-photon Fock state was prepared using conditional measurements on photon pairs born in the process of parametric down-conversion: the homodyne detection of one of the photons of the pair was conditioned to the detection of the other photon, thus making sure that a single-photon was detected. In practice, experimental limitations cause an admixture of the vacuum to the measured state, so that the measured negative value of the Wigner function at the origin of phase space was -0.062, as opposed to the theoretical value $-2/\pi$, for the normalization of the Wigner function adopted by the authors.

10.3. Direct Measurement of the Wigner Function for a One-Photon State in a Cavity

More recently [60], the Paris group was able to measure the full Wigner function for a one-photon state in the cavity, using the technique proposed in Ref. [109]. The result is displayed in Fig. 29b, which exhibits a slice of the cylindrically symmetric distribution. From the Wigner function, it is possible to get the photon-number distribution, which is displayed

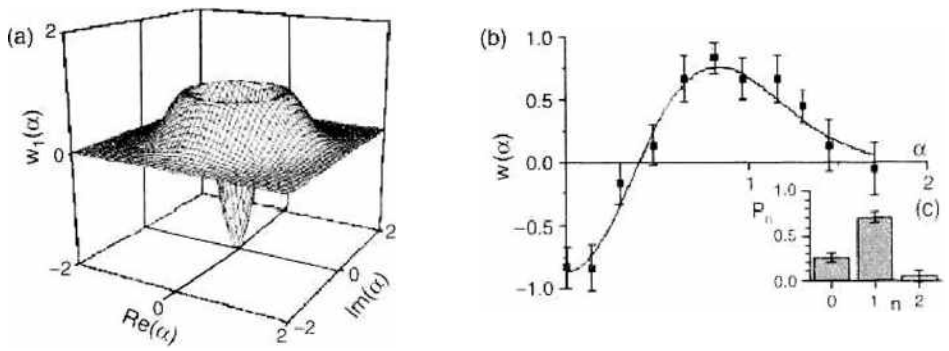


Figure 29. Wigner function for a one-photon state: (a) Distribution in phase space; (b) distribution measured in Ref. [60]; (c) corresponding photon-number distribution, showing that one does not have a pure one-photon state, due to imperfections in the preparation process, the possible decay of the photon in the cavity, and the contamination with thermal photons. Reprinted with permission from [60], P. Bertet et al., *Phys. Rev. Lett.* 89, 200402 (2002). © 2002, American Physical Society.

in Fig. 29c. This distribution shows that the state produced in the cavity is not a perfect one-photon state, which explains the fact that the value of the Wigner function at the origin of phase space is larger than -2, the value it should have for a one-photon state. An interesting feature of this measurement is that it probes a region of the phase space with area smaller than $\hbar$, which corresponds to the negative region of the Wigner function displayed in Fig. 29. It is thus an explicit demonstration of the fact that it is possible in principle to probe regions of phase space as small as one wants!

10.4. Measurement of the Wigner Function and Quantum Circuits

The above scheme for measuring the Wigner function of an electromagnetic field in a cavity amounts to a controlled operation, after the field is displaced. Indeed, the field phase shift depends on the atomic state, as explicitly shown in Eq. (157). Therefore, the direct measurement of the Wigner function can be described in the following way: a $\pi/2$ rotation is applied to the atomic states, while a displacement is applied to the field, which undergoes thereafter a π phase shift conditioned to the atomic state. Finally, the atom undergoes another $\pi/2$ rotation and is then detected. This sequence can be represented by the circuit displayed in Fig. 30.

Quantum circuits analogous to the one in Fig. 30 were used to interpret tomography and spectroscopy as dual forms of quantum computation [146]. In Ref. [146], the above procedure was extended to discrete systems and applied to measure the Wigner function of a system of two qubits, corresponding to spin-1/2 nuclei, in a quantum computation involving liquid-state nuclear magnetic resonance. Further discussions and applications of the measurement of the Wigner function for discrete systems can be found in Refs. [147] and [148].

10.5. Measurement of the Characteristic Function

It is also possible to measure the characteristic function of the electromagnetic field in a cavity, by measuring atoms that have interacted resonantly with the field. Proposals have been presented in Refs. [149–151].

A simple procedure for directly getting the characteristic function involves the application of the quantum-switch concept presented in Ref. [93]. In this implementation, the cavity is connected to a microwave generator detuned with respect to the cavity mode. Because of this frequency mismatch, the field from the generator does not enter the cavity. However, if a two-level atom enters the cavity in the state $|e\rangle$, which interacts dispersively with the cavity mode, the single-atom refraction index effect will displace the frequency of field, so that now the field from the generator becomes resonant, and enters the cavity. This means that the field in the cavity is displaced conditioned to the state of the atom (if the atom is in state $|g\rangle$, nothing happens to the field). The magnitude of the displacement can be controlled by changing with Stark shifts the interaction time between atom and field, while the phase of the displacement is controlled by submitting the field coming from the microwave generator to a delay line.

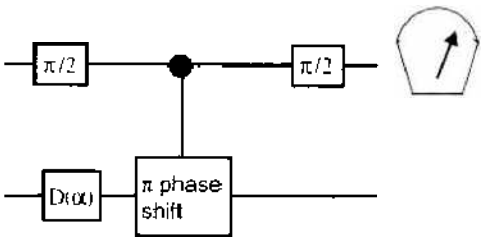


Figure 30. Circuit corresponding to the direct measurement of the Wigner function for an electromagnetic field in a cavity. The upper line corresponds to the atom, the lower one to the field. The atomic state undergoes a $\pi/2$ rotation before and after interacting dispersively with the field. This dispersive interaction can be thought as a phase shift of the field controlled by the atomic state. The field is displaced by α before undergoing this phase shift. Measurement of the atomic state leads to the determination of the Wigner function of the field at the point $-\alpha$ of phase space.

The experimental scheme coincides with the one used to measure the Wigner function, displayed in Fig. 18. Before crossing the cavity, the atom is placed in the state $(|e\rangle + |g\rangle)/\sqrt{2}$, in the first cavity $\mathbf{C}_1$. After the atom interacts with the cavity mode, and undergoes in cavity $\mathbf{C}_2$ the transformation $|e\rangle \rightarrow [|e\rangle + \exp(i\eta)|g\rangle]/\sqrt{2}$, and $|g\rangle \rightarrow [-\exp(-i\eta)|e\rangle + |g\rangle]/\sqrt{2}$ (where η is the relative phase between the classical fields in cavities $\mathbf{C}_1$ and $\mathbf{C}_2$), the entangled atom-field state can be written, similarly to Eq. (157):

$$\begin{aligned}\hat{\rho}_{\text{atom+field}} = {} & |e\rangle\langle e| \otimes [\widehat{D}(\alpha, \alpha^*) - e^{i\eta}\hat{1}]\hat{\rho}[\widehat{D}^{\dagger}(\alpha, \alpha^*) - e^{-i\eta}\hat{1}] \\ & + |g\rangle\langle g| \otimes [\hat{1} + e^{i\eta}\widehat{D}(\alpha, \alpha^*)]\hat{\rho}[\hat{1} + e^{-i\eta}\widehat{D}^{\dagger}(\alpha, \alpha^*)] \\ & + \text{terms non-diagonal in atomic space} \end{aligned} \tag{158}$$

From this expression, one gets

$$P_g - P_e = \Re e\{e^{i\eta}\mathrm{Tr}[\widehat{D}(\alpha, \alpha^*)\hat{\rho}]\} \tag{159}$$

From Eq. (126) and the definition of the displacement operator, one sees that the trace in the above expression is just the symmetrically ordered characteristic function, expressed in terms of the operators $\hat{a}$ and $\hat{a}^{\dagger}$. Therefore, letting

$$C(\alpha, \alpha^*) = \mathrm{Tr}[\widehat{D}(\alpha, \alpha^*)\hat{\rho}] \tag{160}$$

one can see that

$$P_g - P_e = \Re e[e^{i\eta}C(\alpha, \alpha^*)] \tag{161}$$

which means that both the real and the imaginary part of the characteristic function can be measured, by choosing η equal to zero and $\pi/2$, respectively.

The circuit corresponding to this method of measurement of the characteristic function is displayed in Fig. 31.

10.6. Measuring the Husimi Function

As we have seen before, a phase-space picture of the quantum state of the electromagnetic field is also provided by the Husimi function, defined by Eq. (135). Because

$$Q(\alpha, \alpha^*) = \frac{1}{\pi}\langle 0|\widehat{D}^{-1}(\alpha, \alpha^*)\hat{\rho}\widehat{D}(\alpha, \alpha^*)|0\rangle \tag{162}$$

where $\widehat{D}(\alpha, \alpha^*)$ is the displacement operator, one can see that the Husimi distribution can be determined by first displacing the electromagnetic field, as in the Wigner function measurement, and then measuring the population of the vacuum state. This can be done for instance

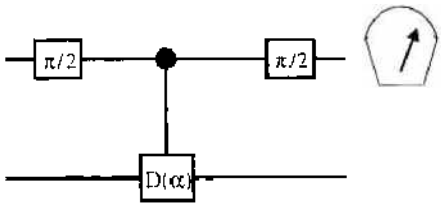


Figure 31. Circuit corresponding to the direct measurement of the characteristic function for an electromagnetic field in a cavity. The upper line corresponds to the atom, the lower one to the field. The atomic state undergoes a $\pi/2$ rotation before and after interacting dispersively with the field. The field produced by the microwave generator is out of resonance with the empty cavity mode and therefore does not enter into the empty cavity but becomes resonant due to the dispersive interaction between the atom in state $|e\rangle$ and the cavity mode, so that if the atom enters the cavity in state $|e\rangle$, a microwave field with complex amplitude α is injected into the cavity. The magnitude $|\alpha|$ of this field is adjusted by changing the interaction time, whereas the phase is changed by letting the microwave field go through a delay line. On the other hand, if the atom is in state $|g\rangle$, nothing happens to the field. This implies that the field in the cavity undergoes a displacement in phase space, conditioned to the atomic state. Measurement of the atomic state, for two different relative phases between the classical fields in cavities $\mathbf{C}_1$ and $\mathbf{C}_2$, leads to the determination of the characteristic function of the field at the point α of phase space.

by the ONB method discussed above. For trapped ions, the measurement of the Husimi function of the motional state through the measurement of the ground-state population was proposed in Ref. [152].

11. EFFECT OF DISSIPATION

11.1. Modeling the Interaction with the Reservoir

We discuss here the effect of dissipation, due to imperfections in the mirrors and diffraction losses. A simple model for dissipation is obtained by coupling the field oscillator (of frequency ω) to a bath of harmonic oscillators, which represent the modes of the reservoir. We consider here for simplicity a rotating-wave Hamiltonian, which may be written as

$$\hat{H} = \hbar\omega\hat{a}^\dagger\hat{a} + \sum_q \hbar\omega_q\hat{b}_q^\dagger\hat{b}_q + \hbar\sum_q(G_q\hat{a}^\dagger\hat{b}_q + G_q^*\hat{a}\hat{b}_q^\dagger) \tag{163}$$

where G_q are the coupling constants, and the bath oscillators have frequency ω_q.

From this Hamiltonian, and under the hypotheses of weak coupling and Markov approximation, one can derive the following master equation for the reduced density operator of the field oscillator [57]

$$\frac{d\hat{\rho}_f}{dt} = \Gamma\bar{n}\left(\hat{a}^\dagger\hat{\rho}_f\hat{a} - \frac{1}{2}\hat{a}\hat{a}^\dagger\hat{\rho}_f - \frac{1}{2}\hat{\rho}_f\hat{a}\hat{a}^\dagger\right) + \Gamma(1+\bar{n})\left(\hat{a}\hat{\rho}_f\hat{a}^\dagger - \frac{1}{2}\hat{a}^\dagger\hat{a}\hat{\rho}_f - \frac{1}{2}\hat{\rho}_f\hat{a}^\dagger\hat{a}\right) \tag{164}$$

where $\bar{n}$ is the average number of thermal photons at frequency ω, given by Planck's distribution, and $\Gamma = 1/t_{\text{cav}}$, where t_{cav} is the dampening time of the cavity mode.

This equation has the general form of the Lindblad equation, which describes the reduced dynamics of a system interacting with its environment [153–157]:

$$\frac{d\hat{\rho}}{dt} = \mathcal{J}\hat{\rho} \equiv \sum_i\left(\frac{\gamma_i}{2}\right)(2\hat{c}_i\hat{\rho}\hat{c}_i^\dagger - \hat{c}_i^\dagger\hat{c}_i\hat{\rho} - \hat{\rho}\hat{c}_i^\dagger\hat{c}_i) \tag{165}$$

where $\hat{\rho}$ is the reduced density operator of the system in the interaction picture, and we have omitted the unitary evolution term $-(i/\hbar)[\hat{H}, \hat{\rho}]$. The operators $\hat{c}_i$ are closely related to the system operators present in the interaction Hamiltonian and γ_i measures the strength of the system-environment coupling.

By solving Eq. (164), with the initial state of the mode in the cavity being a coherent superposition of two coherent states, one arrives at the results mentioned in Sections 7.2 and 7.3. Rather than doing this, we adopt a procedure that leads directly to the time-dependent Wigner function corresponding to the cavity mode.

11.2. Heisenberg Equations of Motion

The method here exposed, and which results from unpublished work of the author with V. M. Kenkre, involves solving the Heisenberg equations of motion for the cavity mode and bath operators. It is straightforward to write down the evolution equation for the Heisenberg operators $\hat{a}(t)$, $\hat{b}_q(t)$, and their adjoints. Because the resulting equations are linear, they may be solved by the Laplace transform method.

The explicit solution for $\hat{a}(t)$ is given by

$$\hat{a}(t) = \hat{a}(0)\xi(t) + \sum_q \eta_q(t)\hat{b}_q(0) \tag{166}$$

The Laplace transforms of the c-number functions $\xi(t)$ and $\eta_q(t)$ appearing in (166) are given respectively by

$$\tilde{\xi}(s) = [s + i\omega + \tilde{\phi}(s)]^{-1} \tag{167}$$

$$\tilde{\eta}_q(s) = -iG_q(s + i\omega_q)^{-1}[s + i\omega + \tilde{\phi}(s)]^{-1} \tag{168}$$

The function ϕ that appears in both (167) and (168) reflects the nature of the coupling and is given in the time domain by

$$\phi(t) = \sum_q |G_q|^2 e^{-i\omega_q t} = \int_0^\infty d\nu\ \mathcal{G}(\nu) e^{-i\nu t} \tag{169}$$

In the second equality in (169) we have introduced the quantity $\mathcal{G}(\nu)$, which equals the product of the coupling $|G_q|^2$, assumed to be a function of the frequency alone, and the density of states $\sum_q \delta(\nu - \omega_q)$ of the b-oscillators.

Because the total number of oscillator excitations in both oscillators, $\hat{a}^\dagger \hat{a} + \sum_q \hat{b}_q^\dagger \hat{b}_q$, commutes with the Hamiltonian and is therefore an invariant, it follows that

$$|\xi(t)|^2 + \sum_q |\eta_q(t)|^2 = 1 \tag{170}$$

11.3. Time-Dependent Wigner Function

In order to get now the time-dependent Wigner function, we express it in terms of the corresponding characteristic function:

$$W(\alpha, \alpha^*, t) = \frac{1}{\pi} \int d^2 z\ e^{\alpha z^* - \alpha^* z} C_W(z, z^*, t) \tag{171}$$

The characteristic function, which is the expectation value of the displacement operator, is given by Eq. (138), expressed in terms of the operators $\hat{a}$ and $\hat{a}^\dagger$. One gets

$$C_W(z, z^*) = Tr\left[\hat{\rho}(0) e^{z\hat{a}^\dagger(t) - z^* \hat{a}(t)}\right] \tag{172}$$

where $\hat{\rho}(0)$ denotes the initial density matrix. This expression may be evaluated from Eq. (166):

$$C(z, z^*) = \mathrm{Tr}\left\{\hat{\rho}(0) \exp[z\xi^*(t)\hat{a}^\dagger - z^*\xi(t)\hat{a}] \exp\left[\sum_q (z\eta_q^*(t)\hat{b}_q^\dagger - z^*\eta_q(t)\hat{b}_q)\right]\right\} \tag{173}$$

This result is obtained by using the Baker–Hausdorff identity in the standard manner.

We will take the initial state of the oscillator to be a Schrödinger-cat-like state, and the reservoir to be in thermal equilibrium initially, so that the initial density matrix of the entire system is an outer product of the thermal density matrix for the b-oscillators at temperature T and the Schrödinger-cat state characterized by α_0 for the a-oscillator:

$$\hat{\rho}(0) = \frac{1}{\mathcal{N}}(|\alpha_0\rangle + |-\alpha_0\rangle)(\langle\alpha_0| + \langle -\alpha_0|) \otimes e^{-\beta \sum_q \hbar\omega_q \hat{b}_q^\dagger \hat{b}_q} \left[\mathrm{Tr}_b\ e^{-\beta \sum_q \hbar\omega_q \hat{b}_q^\dagger \hat{b}_q}\right]^{-1} \tag{174}$$

Here β is $1/k_B T$, k_B is the Boltzmann constant, the trace over the b-oscillators is denoted by Tr_b, and $\mathcal{N} = 2[1 + \exp(-2|\alpha_0|^2)]$ is a normalization constant. We denote the thermal average of the bath operators by $\langle \ldots \rangle_{th}$ and use the identity [158]

$$\left\langle e^{s\hat{b}_q^\dagger + r\hat{b}_q} \right\rangle_{th} = e^{sr(\bar{n}_q + 1/2)} \tag{175}$$

where s and r are any c-numbers and $\bar{n}$ is the Bose–Einstein distribution function:

$$\bar{n}_q = (e^{\hbar\omega_q / k_B T} - 1)^{-1} \tag{176}$$

Evaluation of the characteristic function in Eq. (173) is then straightforward, and its substitution in Eq. (171) and computation of the integrals leads immediately, upon using Eq. (170), to the explicit expression for the Wigner function for all times and arbitrary temperatures:

$$W(\alpha, \alpha^*) = \frac{1}{\mathcal{N}K(t)} \left\{ \exp\left[-\frac{|\alpha - \xi(t)\alpha_0|^2}{K(t)}\right] + \exp\left[-\frac{|\alpha + \xi(t)\alpha_0|^2}{K(t)}\right] + 2F(\alpha, \alpha^*, t) \right\} \tag{177}$$

The fringe function $F(\alpha, \alpha^*, t)$ is given by

$$F(\alpha, \alpha^*, t) = \exp(-2|\alpha_0|^2) \exp\left[\frac{|\alpha|^2 - |\xi|^2|\alpha_0|^2}{K(t)}\right] \cos\left\{2\frac{\mathrm{Im}[\alpha\alpha_0^*\xi^*(t)]}{K(t)}\right\} \tag{178}$$

while the width function $K(t)$, which reduces to $1/2$ for zero temperatures, is generally larger and a function of the temperature:

$$K(t) = \frac{1}{2} + \sum_q \frac{|\eta_q(t)|^2}{e^{\hbar\omega_q/k_BT} - 1} \tag{179}$$

Equations (177), (178), and (179) show that the initial Wigner function consists of two Gaussians centered around $\pm\alpha_0$, and interference fringes in between. As time evolves, the centers of the Gaussians move, and the visibility of the interference fringes also changes. The function $\xi(t)$ governs both the motion of the Gaussians and the change in fringe visibility.

The Markovian limit corresponds to setting $\phi(t) = \Gamma\delta(t)$, which implies that

$$\xi(t) = e^{-i\Omega t} e^{-\Gamma t} \tag{180}$$

More generally, one could have a frequency shift as well, in the Markovian limit, arising from the fact that the integration in Eq. (169) is from zero to infinity.

For zero temperature, the centers of the two Gaussians decay exponentially with a lifetime given by $2/\Gamma$, while the visibility of the interference fringes decays initially with the lifetime $1/|\alpha_0|^2\Gamma$, which is the decoherence time in this case. After this initial decay of the coherence, which for $|\alpha_0| \gg 1$ is much faster than the decay rate of the energy of the system, the state of the system is very close to a statistical mixture of the two states $|\alpha_0\rangle$ and $|-\alpha_0\rangle$.

For temperatures different from zero, one has, in addition, a variation in time of the width of the Gaussians and of the fringe function, which corresponds to the thermalization of the initial state of the system. This function may be calculated by making another approximation, in addition to the Markov one: one assumes that one may extract from the sum in Eq. (179) the temperature-dependent factor, setting $\omega_q \to \omega$ [159]. From Eq. (170), it follows that $\sum_q |\eta_q(t)|^2 = 1 - |\xi(t)|^2$, and therefore

$$K(t) \approx \frac{1}{2} + \frac{1 - |\xi(t)|^2|}{e^{\hbar\omega/k_BT} - 1} \tag{181}$$

Because $|\xi(0)| = 1$ and $\lim_{t\to\infty} |\xi(t)| \to 0$, this expression explicitly displays the evolution of the width of the Gaussian contributions to the Wigner function, from the initial value $1/2$ to the final thermal expression $1/2 + \bar{n}$, where $\bar{n}$ is, as in Eq. (164), the number of thermal photons at the cavity-mode frequency ω, given by Planck's distribution.

12. WIGNER FUNCTION AND ENTANGLEMENT

The Wigner function may also be useful to characterize entanglement between two or more particles, or modes of the electromagnetic field. In fact, it was shown by Banaszek and Wódkiewicz [160] that Bell-type inequalities can be established between four points of the two-mode Wigner function corresponding to the pure entangled state $(|1\rangle|0\rangle - |0\rangle|1\rangle)/\sqrt{2}$, which involves Fock states of the electromagnetic field. Further relations between the Wigner and Husimi functions and nonlocality were explored by Banaszek et al. [139, 161, 162]. In special, it was shown by these authors that a non-negative Wigner function may have non-classical features, involving the violation of Bell-type inequalities. Experimental verification was obtained through the homodyne measurement of an optical Einstein–Podolsky–Rosen (EPR) state, produced in the spontaneous parametric down-conversion of an ultrashort classical pump pulse [163]. This experiment was based on measuring interference between the light in the EPR state and weak light pulses in a coherent state. In this case, the corresponding Wigner function is non-negative, but a Bell-type inequality was shown to be violated.

More generally, it is possible to show that the Wigner function may provide a quantitative measurement of the entanglement between two Fock states of the electromagnetic field [164].

The two-mode Wigner function can be written as

$$W(\alpha,\beta) = tr\{\rho\hat{\Pi}(\alpha,\beta)\} \tag{182'}$$

where $\hat{\Pi}(\alpha,\beta)$ is the product of the displaced parity operator for each mode (A and B).

$$\hat{\Pi}(\alpha,\beta) = D(\alpha)(-1)^{\hat{n}_A}D^{\dagger}(\alpha)D(\beta)(-1)^{\hat{n}_B}D^{\dagger}(\beta)$$

This expression shows that the value of the Wigner function at any point $\{\alpha,\beta\}$ in phase space may be measured by first displacing each mode with a complex amplitude (α and β respectively), and then measuring the sum of the parities for the resulting fields. These modes may be in a same cavity and have different frequencies, or may be distributed in two cavities and have the same or different frequencies. It is easy to generalize the procedure discussed before to the direct measurement of the Wigner function corresponding to these modes: after displacing the field in each mode (by different complex amplitudes), one lets the same atom interact dispersively with the two modes, so as to realize the corresponding parity operators. Measurement of the atomic population difference after this interaction leads to the multimode Wigner function (182).

Let us consider a two-mode pure state of the form:

$$|\Psi_{AB}\rangle = a|0_A\rangle|0_B\rangle + b|0_A\rangle|1_B\rangle + c|1_A\rangle|0_B\rangle + d|1_A\rangle|1_B\rangle \tag{183}$$

where a, b, c, d are arbitrary complex numbers restrained by the normalization condition $\langle\Psi_{AB}|\Psi_{AB}\rangle = 1$, and we have restricted ourselves to the state space generated by the Fock states with $n = 0, 1$ in each mode, A and B.

The quantity of entanglement is defined as [165]

$$E\{\rho_{AB}\} = -2tr\{\hat{\rho}_A \log \hat{\rho}_A\} \tag{184}$$

where $\hat{\rho}_A$ is the reduced density operator corresponding to mode A.

For state (183), $\hat{\rho}_A$ is given by

$$\hat{\rho}_A = tr_B\{\hat{\rho}_{AB}\} = \begin{pmatrix} |a|^2 + |b|^2 & ac^* + bd^* \\ a^*c + b^*d & |c|^2 + |d|^2 \end{pmatrix} \tag{185}$$

The von Neumann entropy corresponding to this density operator may be expressed in terms of the eigenvalues $\lambda_{\pm}$ of $\hat{\rho}_A$:

$$E\{\hat{\rho}_{AB}\} = -2\{\lambda_+ \log \lambda_+ + \lambda_- \log \lambda_-\} \tag{186}$$

These eigenvalues are the roots of the equation $\lambda^2 - \lambda + n_1 + n_2 = 0$, where

$$n_1 = |a|^2|c|^2 + |b|^2|d|^2 - ac^*b^*d - a^*cbd^* \tag{187}$$

$$n_2 = (|a|^2 - |b|^2)(|d|^2 - |c|^2) \tag{188}$$

It will be shown now that n_1 and n_2 can be obtained from four points of the two-mode Wigner function corresponding to the state given by Eq. (183):

$$W(\alpha,\beta) = \langle\Psi_{AB}|\hat{\Pi}(\alpha,\beta)|\Psi_{AB}\rangle \tag{189}$$

It is easy to show that

$$\frac{W(\alpha,0) + W(-\alpha,0)}{2e^{-2|\alpha|^2}} = |a|^2 - |b|^2 + (4|\alpha|^2 - 1)(|c|^2 - |d|^2) \tag{190}$$

Evaluation of this expression for two different values of $|\alpha|$ is enough to determine the population differences $|d|^2 - |c|^2$ and $|a|^2 - |b|^2$, and therefore n_2. Therefore, two pairs of symmetric points lying on different circumferences in the plane $\beta = 0$ determine n_2.

On the other hand,

$$n_1 = \frac{1}{4}\left\{\left[\frac{W(|\alpha|,0) - W(|\alpha|e^{i\pi},0)}{2|\alpha|e^{-2|\alpha|^2}}\right]^2 + \left[\frac{W(|\alpha'|e^{i\pi/2},0) - W(|\alpha'|e^{i3\pi/2},0)}{2|\alpha'|e^{-2|\alpha'|^2}}\right]^2\right\} \tag{191}$$

More generally, any set of points of the two-mode Wigner function containing the vertices of a rhombus located in the plane $\beta = 0$, and centered at $\alpha = 0$, allows one to calculate the eigenvalues of the reduced density matrix for mode A, with the only restriction that the diagonals of this rhombus must be different ($|\alpha'| \neq |\alpha|$). Once the eigenvalues of $\hat{\rho}_1$ are obtained, through these four points, one can fully determine the quantity of entanglement for the proposed system. Obviously the same conditions apply to mode B.

For a three-mode entangled state, it is also possible to relate its entanglement with properties of the Wigner function. In particular, it is possible to show that the value of the Wigner function at the origin of phase space provides a test for the quantum character of a GHZ state [166].

13. CONTROL OF DECOHERENCE

We have seen that decoherence is an important mechanism for the establishment of the properties of the classical world. It is also a major obstacle for the realization of quantum computers. For this reason, several ideas have been presented for overcoming decoherence and protecting quantum information.

13.1. Quantum Error Correction

A possible approach is, in analogy with classical computation, to encode the information redundantly, through *error correcting codes* [167–171]. Thus, for correcting flip errors (in which the bit 0 gets transformed into the bit 1, and vice versa), one may encode the state $|\psi\rangle = a|0\rangle + b|1\rangle$ into the state $|\Psi\rangle = a|000\rangle + b|111\rangle$. The circuit that implements this encoding is displayed in Fig. 32.

A flip error could then be detected in the following way, if one assumes that the probability of having two flip errors is very small. One measures the binary sum of the values of the first and third bits, $x \oplus z$, and also of the second and third bits, $y \oplus z$. For the states $|000\rangle$ and $|111\rangle$, these sums are equal to zero, but if one of the bits flips, then at least one of the sums becomes equal to one. Then $S = (y \oplus z, x \oplus z)$ indicates in binary notation the position (1,2, or 3) of the bit that has flipped. The pair of bits $(y \oplus z, x \oplus z)$ constitutes a *syndrome* that allows the identification and the correction of the error. For instance, if $a|000\rangle + b|111\rangle \rightarrow a|100\rangle + b|011\rangle$, then $S = (0, 1)$, implying that the first bit must be corrected.

The measurement of the syndrome can be implemented through the circuits displayed in Fig. 33. Note that these circuits lead to the measurement of the sums $x \oplus z$ and $y \oplus z$, but not of the individual qubits, which are unaffected by the measurement. They can be corrected therefore after the syndrome is known.

This procedure does not correct phase errors, that transform for instance the qubit $a|0\rangle + b|1\rangle$ into the qubit $a|0\rangle - b|1\rangle$. In order to also correct these errors, Shor [167] proposed an

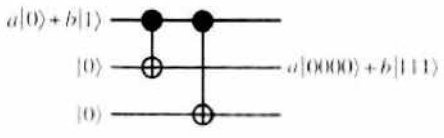


Figure 32. Circuit for encoding an initial state. The state $|\psi\rangle = a|0\rangle + b|1\rangle$ is encoded into the state $|\Psi\rangle = a|000\rangle + b|111\rangle$.

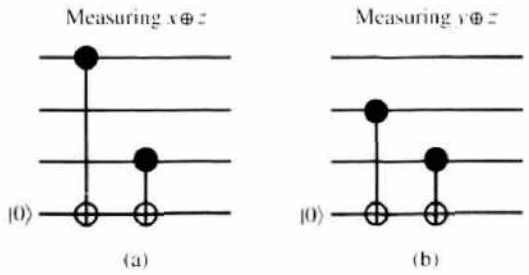


Figure 33. Circuits for identifying flip errors. The circuit shown in (a) provides the sum $x \oplus z$, where x stands for the upper bit, and z for the third one from up to down. The lower bit is used to register the sum. The circuit shown in (b) provides, in the same way, the sum $y \oplus z$.

encoding involving nine qubits:

$$\begin{aligned} |0\rangle &\rightarrow |\tilde{0}\rangle = (|000\rangle + |111\rangle)(|000\rangle + |111\rangle)(|000\rangle + |111\rangle)/2^{3/2} \\ |1\rangle &\rightarrow |\tilde{1}\rangle = (|000\rangle - |111\rangle)(|000\rangle - |111\rangle)(|000\rangle - |111\rangle)/2^{3/2} \end{aligned} \tag{192}$$

A flip error can be corrected in the same was as before, by addressing two of the qubits inside one of the three groups in the above encoding. On the other hand, a phase error can be detected by comparing the relative phases of the states within two of the three groups. This can be done by applying to this state products of spin-flip operators σ_x corresponding to six of the qubits.

A simpler scheme for general quantum error correction, involving only five qubits, was proposed in Ref. [168].

13.2. Decoherence-Free Subspaces

A different scheme for protection against decoherence involves the concept of decoherence-free subspaces. They arise whenever all the qubits are coupled to the same environment [172–175]. An example is provided by the following interaction Hamiltonian, which may describe the interaction of N two-level atoms (or N spins) with a common reservoir:

$$\hat{H}_I = \sum_k (g_k \hat{S}^+ \hat{b}_k + g_k^* \hat{S}^- \hat{b}_k^\dagger) \tag{193}$$

where

$$\hat{S}^+ = \sum_{i=1}^{N} \hat{\sigma}_i^+ \tag{194}$$

is a collective spin operator, and $\hat{\sigma}_i^+$ is the spin-raising operator for atom i. We note that if $N = 2$, then the two-atom singlet state $(|+-\rangle - |-+\rangle)/\sqrt{2}$, belonging to the subspace with total angular momentum $J = 0$, is not affected by the interaction with the environment. For $N = 4$, the subspace corresponding to $J = 0$ has multiplicity equal to two, and therefore it is possible to find in this subspace two basis vectors, which would be associated to two states of a decoherence-free qubit. The dimensionality of the decoherence-free subspace gets larger as N increases.

In practice, the conditions for application of this method may not be easy to achieve, as in general the coupling with the environment depends on the position of the qubit. Experimental demonstrations have been made for twin-photon beams [176] and trapped ions [177].

Subspaces may also be decoupled from the environment through dynamic techniques, involving time-dependent short-time modifications of the dynamics of the system [175, 178].

13.3. Reservoir Engineering

Another possible technique consists in engineering an artificial environment, which has a stronger interaction with the system of interest than the other environments, and which is built for the purpose of protecting a given state [179]. This is based on the idea of reservoir

engineering, first proposed by Cirac and Zoller [180], and closely related to the method of Hamiltonian engineering proposed by Matos Filho and Vogel [181]. One submits two internal electronic states of the ion to a nonresonant laser field, so that the upper level remains with negligible population all the time. It can be then adiabatically eliminated, leading to a net effect that may be characterized as a nonunitary evolution of the center-of-mass of the ion, described by an extra term in the center-of-mass master equation, of the form

$$\frac{d\hat{\rho}}{dt} = \mathcal{L}\hat{\rho} + \left(\frac{\Gamma_{\text{eng}}}{2}\right)(2\hat{D}\hat{\rho}\hat{D}^{\dagger} - \hat{D}^{\dagger}\hat{D}\hat{\rho} - \hat{\rho}\hat{D}^{\dagger}\hat{D}) \tag{195}$$

where the first term on the right-hand side stands for the Lindbladian associated with the environment, described by Eq. (165), and the operator $\hat{D}$ is chosen so that the state one wants to protect is the only steady-state of Eq. (195) without the environment term $\mathcal{L}\hat{\rho}$. For $\Gamma_{\text{eng}} \gg \gamma_l$, where γ_l are the decay constants in the Lindbladian given by Eq. (165), the steady state of Eq. (195) will be very close to the state to be protected (if however the state is not unique, the term $\mathcal{L}\hat{\rho}$ could still induce transitions between the steady states). Besides, any state of the system will decay into the state chosen to be protected.

14. DECOHERENCE AND CLASSICALLY CHAOTIC SYSTEMS

The problem of understanding the classical world from quantum theory is subtle and especially challenging when dealing with classically chaotic systems.

Indeed, the definition of classical chaos cannot directly be translated to quantum mechanics, as the exponential sensitivity to initial conditions, used to define classical chaos, relies on the concept of individual trajectories in phase space, which is absent in the quantum formalism. The use of classical phase space distributions, instead of trajectories, helps to circumvent this problem, as they can readily be compared with quasi-probability distributions, like the Wigner distribution, defined for the corresponding quantum system.

14.1. The Ehrenfest Time

One expects, however, that the dynamics of the quantum and the corresponding classical system should differ, after some time, even if the initial distributions coincide. This time, often called Ehrenfest time or breaking time, while large for integrable systems, can be very short for chaotic systems. Indeed, in this case it has been shown [182, 183] to be proportional to the logarithm of the inverse of an effective Planck constant, $\hbar_{\text{eff}}$, which is the ratio between Planck's constant and a typical action of the system. For integrable systems, on the other hand, it scales as an inverse power of $\hbar_{\text{eff}}$. In fact, quantum corrections become important when the distribution is able to explore the nonlinearities of the potential, which for chaotic systems occurs in a logarithmic time-scale due to the exponentially fast stretching of the distribution, imposed by chaotic dynamics.

In the macroscopic limit, namely $\hbar_{\text{eff}} \to 0$, one gets an infinite breaking time $\tau_{\hbar}$. Nevertheless, for any physical system, $\hbar_{\text{eff}}$ is not zero and therefore $\tau_{\hbar}$ has a finite value, which can be short, even for macroscopic systems. Indeed, it has been argued that, due to the shortness of the separation time, components of the solar system should exhibit quantum features, which is in contradiction with observation [184–187].

Reconciliation of quantum and classical predictions in this case is provided by the irreversible coupling of the system with an environment, which leads to the elimination of the quantum signatures, so that quantum and classical evolutions remain alike [188–198].

14.2. The Kicked Harmonic Oscillator

An example is provided by the kicked harmonic oscillator, described by the Hamiltonian:

$$\hat{H} = \frac{\hat{p}^2}{2m} + \frac{m\nu^2\hat{x}^2}{2} + \frac{m\nu K}{k^2}\cos(k\hat{x})\sum_{n}^{\infty}\delta(t - n\tau) \tag{196}$$

where ν is the oscillator frequency, τ the interval between two consecutive kicks, and K their dimensionless amplitude. The kicking potential is position dependent with a periodicity

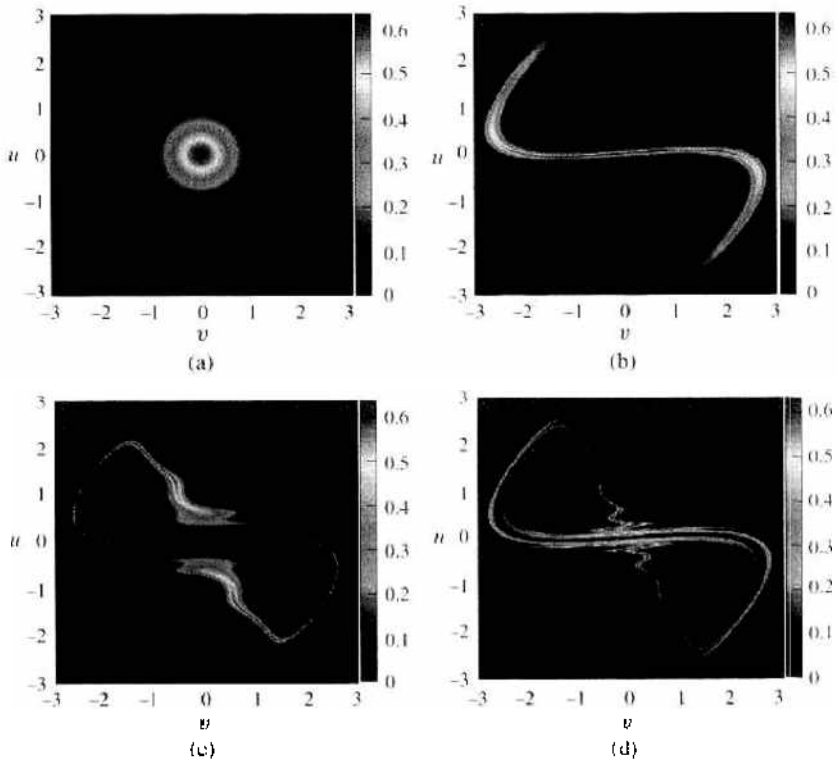


Figure 34. Classical probability distribution for the kicked harmonic oscillator, after 0 (a), 3 (b), 6 (c) and 9 (d) kicks.

given by the wave vector k. The same Hamiltonian represents the classical and the quantum systems.

The role of effective Planck constant in this case is played by the Lamb–Dicke parameter, defined as

$$\eta = k\sqrt{\frac{\hbar}{2m\nu}} \tag{197}$$

In the following, we describe the system in phase space through the renormalized dimensionless variables

$$\bar{v} = x\sqrt{\frac{m\nu}{2\hbar}} = \frac{kx}{2\sqrt{\eta}}, \qquad \bar{u} = \frac{p}{\sqrt{2m\nu\hbar}} = \frac{kp}{2m\nu\sqrt{\eta}} \tag{198}$$

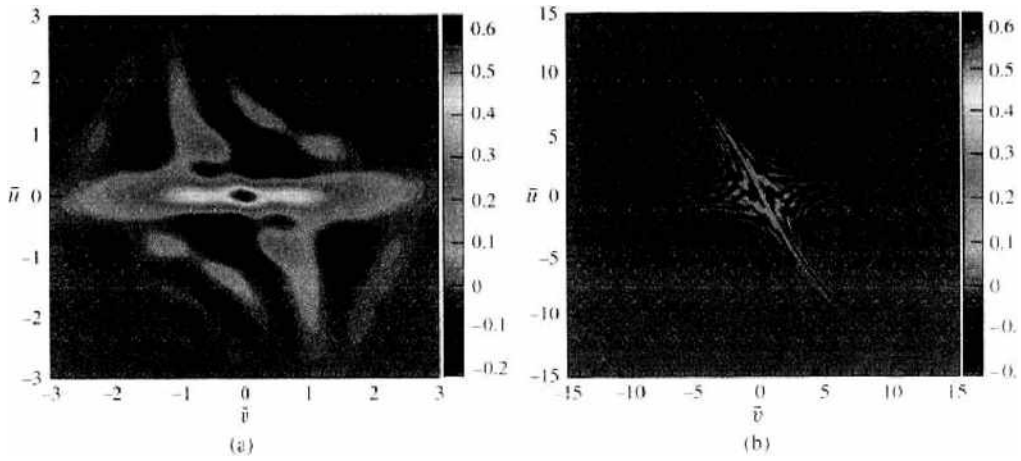


Figure 35. Wigner distribution after 9 kicks for $\eta = 0.5$ (a) and $\eta = 0.1$ (b). In both cases, the Wigner function presents negative values, but as the Lamb–Dicke parameter is decreased (closer to the classical limit), there is a better correspondence with the overall classical structure shown in Fig. 34d.

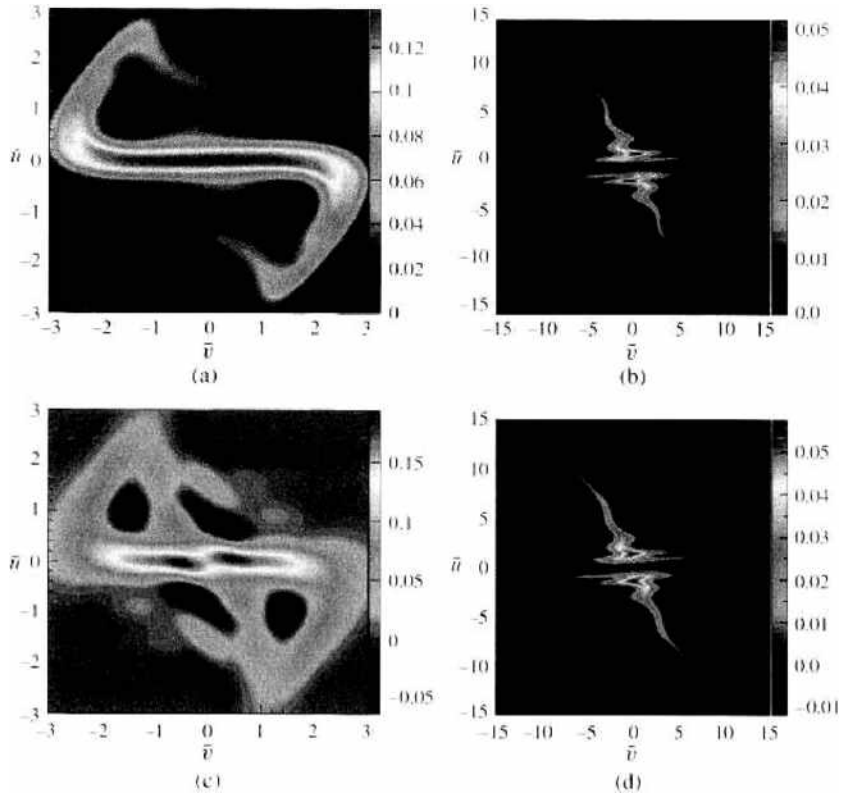


Figure 36. Classical (top) and Wigner (bottom) distributions for $K = 2.0$, $D = \gamma\tau/2 = 0.01047$, and $\eta = 0.5$ (left) or $\eta = 0.1$ (right). Diffusion leads to a better quantum-classical correspondence as compared to Fig. 35. For $\eta = 0.1$, this correspondence is quite impressive, whereas for $\eta = 0.5$, differences still remain. Diffusion also prevents the appearance of small scale structures on classical dynamics (top).

Diffusion may be introduced via a diffusive term in a Fokker–Planck equation, satisfied by both the classical and the quantum (Wigner) distributions, which in terms of these renormalized variables $\bar{u}$ and $\bar{v}$ can be written as:

$$\frac{\partial W}{\partial t} = \frac{\gamma}{4}\left(\frac{\partial^2}{\partial \bar{v}^2} + \frac{\partial^2}{\partial \bar{u}^2}\right) W \tag{199}$$

Figure 34 displays the evolution of the classical distribution with the number of kicks, in the absence of diffusion, and for an initial state that coincides with the ground state of the quantum harmonic oscillator. For the classical distribution, the plot is made in terms of the variables u and v, which are obtained from $\bar{u}$ and $\bar{v}$ by setting $\eta = 1$. For other values of η, the renormalized variables $\bar{u}$ and $\bar{v}$ are used, so that the initial state has always the same width. In all cases, we take $\nu\tau = \pi/3$ and $K = 2$. Figure 35 displays the corresponding quantum Wigner distribution after 9 kicks, also in the absence of diffusion, and starting with the same initial distribution. The situation when diffusion is present is displayed in Fig. 36. Comparison between these two figures clearly demonstrates the important role played by the interaction with the environment in the classical limit of quantum mechanics [198].

15. CONCLUSIONS

In this article, it was shown that techniques used in the field of quantum optics are helpful to demonstrate several peculiar features of quantum physics, of relevance to the area of quantum information, and to the discussion of the quantum-classical transition. They allow in particular the monitoring of the decoherence process, which not only is at the heart of the quantum theory of measurement, but is a major obstacle to the realization of quantum computers. Its understanding and control is therefore an important aim in current efforts to build these devices. It is interesting that a problem of such practical importance should be so intertwined with fundamental questions of physics.

The demonstration that the decoherence time decreases with the size of the system does not mean, however, that the problem of the classical limit has been solved. In fact, coherence does not really disappear, and it is still present in entangled states of the cavity field and the rest of the universe. A detailed treatment of this problem would thus involve the consideration of the quantum character of the universe. According to Murray Gell-Man and Jim Hartle, "quantum mechanics is better and more fundamentally understood within the context of quantum cosmology" [199].

Even though fundamental problems related to the classical limit of quantum mechanics and the quantum theory of measurement remain to be solved, I think it is fair to say that quantum optics has helped us to understand and observe some important pieces of this puzzle.

Recent developments in the area of cavity quantum electrodynamics involve the investigation of mesoscopic systems (Cooper-pair boxes) interacting with the quantized electromagnetic field in a cavity [200]. They may provide new directions for the study of entanglement and quantum measurement physics with macroscopic objects.

ACKNOWLEDGMENTS

This work was partially supported by PRONEX (Programa de Apoio a Núcleos de Excelência), the Brazilian Millennium Institute for Quantum Information, CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico), FAPERJ (Fundação de Amaparo à Pesquisa do Estado do Rio de Janeiro), CAPES (Coordenação de Aperfeiçoamento de Pessoal de Ensino Superior), and FUJB (Fundação Universitária José Bonifácio). It is a pleasure to acknowledge the collaboration on the subjects covered by this article with M. Brune, A. R. R. Carvalho, S. Haroche, M. França Santos, L. G. Lutterbach, R. L. de Matos Filho, P. Milman, M. Orszag, J. M. Raimond, F. Toscano, and N. Zagury.

REFERENCES

1. C. H. Bennett and D. P. DiVincenzo, *Nature* 404, 247 (2000).
2. D. Bouwmeester, A. Ekert, and A. Zeilinger (Eds.), "The Physics of Quantum Information." Springer, Berlin, 2000.
3. M. A. Nielsen and I. Chuang, "Quantum Computation and Quantum Information." Cambridge University Press, Cambridge, U.K., 2000.
4. J. Preskill, "Quantum Information and Computation." Lecture notes available at http://www.theory.caltech.edu/~preskill/ph229 (1998).
5. R. Omnès, "The Interpretation of Quantum Mechanics." Princeton University Press, Princeton, NJ, 1994.
6. D. Giulini, E. Joos, C. Kiefer, J. Kupsch, I.-O. Stamatescu, and H. D. Zeh, "Decoherence and the Appearance of a Classical World in Quantum Theory." Springer, Berlin, 1996.
7. W. H. Zurek, *Rev. Mod. Phys.* 75, 715 (2003).
8. Letter from Albert Einstein to Max Born in 1954, cited by E. Joos, in Ref. 6, p. 54.
9. E. Schrödinger, *Naturwissenschaften* 23, 807 (1935); 23, 823 (1935); 23, 844 (1935). English translation by J. D. Trimmer, *Proc. Am. Phys. Soc.* 124, 3235 (1980).
10. J. Von Neumann, "Die Mathematische Grundlagen der Quantenmechanik." Springer-Verlag, Berlin, 1932; English translation by R. T. Beyer: "Mathematical Foundations of Quantum Mechanics." Princeton University Press, Princeton, NJ, 1955.
11. J. A. Wheeler and W. H. Zurek (Eds.) "Quantum Theory and Measurement." Princeton University Press, Princeton, NJ, 1983; W. H. Zurek, *Phys. Today* 44, 36 (1991).
12. E. Wigner, in "The Scientist Speculates," (I. J. Good, Ed.), p. 284, William Heinemann, London, 1962; and also in "Symmetries and Reflections," p. 171. Indiana University Press, Bloomington, IN, 1967. See also E. Wigner, *Am. J. Phys.* 31, 6 (1963).
13. G. Ludwig, in "Werner Heisenberg und die Physik unserer Zeit." Friedrich Vieweg und Sohn, Braunschweig, 1961.
14. K. Hepp, *Helv. Phys. Acta* 45, 237 (1972).
15. J. S. Bell, *Helv. Phys. Acta* 48, 93 (1975).
16. K. Gottfried, "Quantum Mechanics," Sec. IV. Benjamin, Reading, MA, 1966.
17. H. D. Zeh, *Found. Phys.* 1, 69 (1970).
18. W. H. Zurek, *Phys. Rev. D* 24, 1516 (1981); 26, 1862 (1982).
19. H. Dekker, *Phys. Rev. A* 16, 2116 (1977).
20. A. O. Caldeira and A. J. Leggett, *Physica (Amsterdam)* 121A, 587 (1983); *Phys. Rev. A* 31, 1059 (1985).
21. E. Joos and H. D. Zeh, *Z. Phys. B* 59, 223 (1985).

22. G. J. Milburn and C. A. Holmes, *Phys. Rev. Lett.* 56, 2237 (1986).
23. F. Haake and D. Walls, *Phys. Rev. A* 36, 730 (1987).
24. W. G. Unruh and W. H. Zurek, *Phys. Rev. D* 40, 1071 (1989).
25. W. H. Zurek, *Phys. Today* 44, No. 10, 36 (1991).
26. B. L. Hu, J. P. Paz, and Y. Zhang, *Phys. Rev. D* 45, 2843 (1992).
27. C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W. K. Wootters, *Phys. Rev. Lett.* 70, 1895 (1993).
28. L. Davidovich, N. Zagury, M. Brune, J. M. Raimond, and S. Haroche, *Phys. Rev. A* 50, R895 (1994).
29. D. Bouwmeester, J.-W. Pan, K. Mattle, M. Eible, H. Weinfurter, and A. Zeilinger, *Nature* 390, 575 (1997).
30. D. Boschi, S. Branca, F. De Martini, L. Hardy, and S. Popescu, *Phys. Rev. Lett.* 80, 1121 (1998).
31. A. Furusawa, J. L. Sørensen, S. L. Braunstein, C. A. Fuchs, H. J. Kimble, and E. S. Polzik, *Science* 282, 706 (1998).
32. M. A. Nielsen, E. Knill, and R. Laflamme, *Nature* 396, 52 (1998).
33. E. Lombardi, F. Sciarrino, S. Popescu, and F. De Martini, *Phys. Rev. Lett.* 88, 070402 (2002).
34. A. Dolińska, B. C. Buchler, W. P. Bowen, T. C. Ralph, and P. K. Lam, *Phys. Rev. A* 68, 052308 (2003).
35. I. Marcikic, H. de Riedmatten, W. Tittel, H. Zbinden, and N. Gisin, *Nature* 421, 509 (2003).
36. H. de Riedmatten, I. Marcikic, W. Tittel, H. Zbinden, D. Collins, and N. Gisin, *Phys. Rev. Lett.* 92, 047904 (2004).
37. D. Fattal, E. Diamanti, K. Inoue, and Y. Yamamoto, *Phys. Rev. Lett.* 92, 037904 (2004).
38. C. J. Hood, T. W. Lynn, A. C. Doherty, A. S. Parkins, and H. J. Kimble, *Science* 287, 1447 (2000).
39. P. Goy, J. M. Raimond, M. Gross, and S. Haroche, *Phys. Rev. Lett.* 50, 1903 (1983).
40. S. Haroche, M. Brune, J. M. Raimond, *Europhys. Lett.* 14, 19 (1991).
41. D. Meschede, H. Walther, and G. Muller, *Phys. Rev. Lett.* 54, 551 (1985).
42. G. Rempe, F. Schmidt-Kaler, and H. Walther, *Phys. Rev. Lett.* 64, 2783 (1990).
43. S. Haroche, in "Fundamental Systems in Quantum Optics" (J. Dalibard, J. M. Raimond, and J. Zinn-Justin, Eds.), Proc. Les Houches Summer School, Session LIII, Elsevier, Amsterdam, 1992.
44. P. Berman (Ed.), "Cavity Quantum Electrodynamics," Academic Press, New York, 1994.
45. R. G. Hulet and D. Kleppner, *Phys. Rev. Lett.* 51, 1430 (1983).
46. A. Nussenzveig, J. Hare, A. M. Steinberg, L. Moi, M. Gross, and S. Haroche, *Euro. Phys. Lett.* 14, 755 (1991).
47. C. Cohen-Tannoudji, B. Diu, and F. Laloë, "Quantum Mechanics," Wiley and Hermann, Paris, 1977.
48. R. J. Glauber, *Phys. Rev.* 131, 2766 (1963).
49. R. Loudon and P. Knight, *J. Mod. Opt.* 34, 709 (1987).
50. H. P. Yuen and J. H. Shapiro, *IEEE Trans. IT* 26, 78 (1980).
51. C. M. Caves, *Phys. Rev. D* 23, 1693 (1981).
52. B. L. Schumaker, *Opt. Lett.* 9, 189 (1984).
53. E. T. Jaynes and F. W. Cummings, *Proc. IEEE* 51, 89 (1963).
54. J. J. Kim, K. M. Fonseca Romero, A. M. Horiguti, L. Davidovich, M. C. Nemes, and A. F. R. de Toledo Piza, *Phys. Rev. Lett.* 82, 4737 (1999).
55. M. Brune, E. Hagley, J. Dreyer, X. Maître, A. Maali, C. Wunderlich, J. M. Raimond, and S. Haroche, *Phys. Rev. Lett.* 77, 4887 (1996).
56. R. P. Feynman, F. L. Vernon, Jr., and R. W. Hellwarth, *J. Appl. Phys.* 28, 49 (1957).
57. G. Grynberg, J. Dupont-Roc, and C. Cohen-Tannoudji, "Atom-Photon Interactions: Basic Processes and Applications," Wiley, New York, 1992.
58. C. Cohen-Tannoudji, in "Optical Pumping and Interactions of Atoms with the Electromagnetic Field," Vol. 2 of Cargese Lectures in Physics (M. Levy, Ed.), p. 347, Gordon and Breach, New York, 1968.
59. S. Haroche, Doctoral Thesis, Paris, 1971, published in *Ann. Phys. (Paris)* 6, 189 and 327 (1971).
60. P. Bertet, A. Auffeves, P. Maioli, S. Osnaghi, T. Meunier, M. Brune, J. M. Raimond, and S. Haroche, *Phys. Rev. Lett.* 89, 200402 (2002).
61. M. Greiner, O. Mandel, T. Esslinger, T. W. Hänsch, and Immanuel Bloch, *Science* 415, 39 (2002).
62. D. Jaksch, C. Bruder, J. I. Cirac, C. W. Gardiner, and P. Zoller, *Phys. Rev. Lett.* 81, 3108 (1998).
63. M. P. A. Fisher, P. B. Weichman, G. Grinstein, and D. S. Fisher, *Phys. Rev. B* 40, 546 (1989).
64. S. Inouye, M. R. Andrews, J. Stenger, H.-J. Miesner, D. M. Stamper-Kurn, and W. Ketterle, *Nature* 392, 151 (1998).
65. E. A. Donley, N. R. Claussen, S. L. Cornish, J. L. Roberts, E. A. Cornell, and C. E. Wieman, *Nature* 412, 295 (2001).
66. D. Jaksch, H.-J. Briegel, J. I. Cirac, C. W. Gardiner, and P. Zoller, *Phys. Rev. Lett.* 82, 1975 (1999).
67. E. Jané, G. Vidal, W. Dür, P. Zoller, and J. I. Cirac, *Quantum Information and Computation* 3, 15 (2003).
68. S. Haroche, M. Brune, and J. M. Raimond, *Europhys. Lett.* 14, 19 (1991).
69. B. G. Englert, J. Schwinger, A. O. Barut, and M. O. Scully, *Europhys. Lett.* 14, 25 (1991).
70. A. C. Doherty, A. S. Parkins, S. M. Tan, and D. F. Walls, *Phys. Rev. A* 56, 833 (1997).
71. A. Rauschenbeutel, G. Nogues, S. Osnaghi, P. Bertet, M. Brune, J. M. Raimond, and S. Haroche, *Science* 288, 2024 (2000).
72. D. M. Greenberger, M. A. Horne, and A. Zeilinger, in "Bell's Theorem, Quantum Theory and Conceptions of the Universe" (M. Kafatos, Ed.), p. 69, Kluver Academic, Dordrecht, 1989.
73. D. M. Greenberger, M. A. Horne, A. Shimony and A. Zeilinger, *Am. J. Phys.* 58, 1131 (1990).
74. David Mermin, *Phys. Today* 43, 9 (1990).
75. D. Bouwmeester, J.-W. Pan, M. Daniell, H. Weinfurter, and A. Zeilinger, *Phys. Rev. Lett.* 82, 1345 (1999).

76. C. A. Sackett, D. Kielpinski, B. E. King, C. Langer, V. Meyer, C. J. Myatt, M. Rowe, Q. A. Turchette, W. M. Itano, D. J. Wineland, and C. Monroe, *Nature* 404, 256 (2000).
77. P. Benioff, *Phys. Rev. Lett.* 48, 1581 (1982).
78. R. Feynman, *Optics News* 11, 11 (1985).
79. D. Deutsch, *Proc. R. Soc. London* A400, 97 (1985).
80. D. P. DiVincenzo, *Phys. Rev. A* 51, 1015 (1995).
81. A. Barenco, C. H. Bennett, R. Cleve, D. P. DiVincenzo, N. Margolus, P. Shor, T. Sleator, J. Smolin, and H. Weinfurter, *Phys. Rev. A* 52, 3457 (1995).
82. A. Rauschenbeutel, G. Nögues, S. Osnaghi, P. Bertet, M. Brune, J. M. Raimond, and S. Haroche, *Phys. Rev. Lett.* 83, 5166 (1999).
83. W. K. Wootters and W. H. Zurek, *Nature (London)* 299, 802 (1982).
84. D. Dieks, *Phys. Lett. A* 92, 271 (1982).
85. M. Brune, S. Haroche, J. M. Raimond, L. Davidovich, and N. Zagury, *Phys. Rev. A* 45, 5193 (1992).
86. N. F. Ramsey, "Molecular Beams." Oxford University Press, New York, 1985.
87. M. Brune, P. Nussenzveig, F. Schmidt-Kaler, F. Bernardot, A. Maali, J. M. Raimond, and S. Haroche, *Phys. Rev. Lett.* 76, 1800 (1996).
88. A. Einstein, B. Podolski, and N. Rosen, *Phys. Rev.* 47, 777 (1935).
89. J. S. Bell, *Physics (Long Island City* 1, 195 (1964).
90. S. J. Freedman and J. S. Clauser, *Phys. Rev. Lett.* 28, 938 (1972).
91. A. Aspect, J. Dalibard, and G. Roger, *Phys. Rev. Lett.* 49, 1804 (1982).
92. C. Monroe, D. M. Meekhof, B. E. King, and D. J. Wineland, *Science* 272, 1131 (1996).
93. L. Davidovich, A. Maali, M. Brune, J. M. Raimond, and S. Haroche, *Phys. Rev. Lett.* 71, 2360 (1993).
94. L. Davidovich, M. Brune, J. M. Raimond, and S. Haroche, *Phys. Rev. A* 53, 1295 (1996).
95. B.-G. Englert, N. Sterpi, and H. Walther, *Optics Commun.* 100, 526 (1993).
96. C. J. Myatt, B. E. King, Q. A. Turchette, C. A. Sackett, D. Kielpinski, W. M. Itano, C. Monroe, and D. J. Wineland, *Nature* 403, 269 (2000).
97. Q. A. Turchette, C. J. Myatt, B. E. King, C. A. Sackett, D. Kielpinski, W. M. Itano, C. Monroe, and D. J. Wineland, *Phys. Rev. A* 62, 053807 (2000).
98. V. B. Braginsky, Yu. I. Vorontsov, and F. Ya. Khalili, *Zh. Eksp. Teor. Fiz.* 73, 1340 (1977) [*Sov. Phys. JETP* 46 (1977)].
99. W. G. Unruh, *Phys. Rev. D* 18, 1764 (1978).
100. C. M. Caves, *Phys. Rev. Lett.* 45, 75 (1980).
101. N. Imoto, H. A. Haus, and Y. Yamamoto, *Phys. Rev. A* 32, 2287 (1985).
102. Y. Yamamoto, S. Machida, and O. Nilsson, *Phys. Rev. A* 34, 4025 (1986).
103. R. M. Shelby, M. D. Levenson, S. H. Perlmutter, R. G. DeVoe, and D. F. Walls, *Phys. Rev. Lett.* 57, 691 (1986).
104. P. Grangier, R. E. Slusher, B. Yurke, and A. La Porta, 1987, *Phys. Rev. Lett.* 59, 2153.
105. Y. Yamamoto, S. Machida, N. Imoto, M. Kitagawa, and G. Björk, *J. Opt. Soc. Am.* B4, 1645 (1987).
106. C. M. Caves, *Opt. Lett.* 12, 971 (1987).
107. M. Brune, S. Haroche, V. Lefèvre, J. M. Raimond, and N. Zagury, *Phys. Rev. Lett.* 65, 976 (1990).
108. S. Haroche, M. Brune, and J. M. Raimond, *J. Physique II*, 659 (1992).
109. L. G. Lutterbach and L. Davidovich, *Phys. Rev. Lett.* 78, 2547 (1997).
110. L. G. Lutterbach and L. Davidovich, *Optics Express* 3, 147 (1998).
111. E. Wigner, *Phys. Rev.* 40, 749 (1932).
112. J. Bertrand and P. Bertrand, *Found. Phys.* 17, 397 (1987).
113. U. Leonhardt, "Measuring the Quantum State of Light." Cambridge University Press, Cambridge, U.K., 1997.
114. J. Radon, *Mathematisch-Physiche Klass* 69, 262 (1917).
115. G. M. D'Ariano, U. Leonhardt, and H. Paul, *Phys. Rev. A* 52, 1801 (1995).
116. S. Wallentowitz and W. Vogel, *Phys. Rev. A* 53, 4528 (1996).
117. S. Wallentowitz and W. Vogel, *Phys. Rev. Lett.* 75, 2932 (1995).
118. C. D'Helon and G. Milburn, *Phys. Rev. A* 54, R25 (1996).
119. W. Pauli, in "Handbuch der Physik" (H. Geiger and K. Scheel, Eds.), p. 83. Springer, Berlin, 1933. English translation: W. Pauli, "General Principles of Quantum Mechanics." Springer, Berlin, 1980.
120. K. E. Cahill and R. J. Glauber, *Phys. Rev.* 177, 1857 (1969); *Phys. Rev.* 177, 1882 (1969).
121. M. Hillery, R. F. O'Connell, M. O. Scully, and E. P. Wigner, *Phys. Rep.* 106, 121 (1984); see also W. P. Schleich, "Quantum Optics in Phase Space." Wiley, New York (2001).
122. R. L. Hudson, *Rep. Math. Phys.* 6, 249 (1974).
123. K. Husimi, *Proc. Phys. Mat. Jpn.* 22, 264 (1940).
124. J. E. Moyal, *Proc. Cambridge Philos. Soc.* 45, 99 (1949).
125. W. H. Louisell, "Quantum Statistical Properties of Radiation." Wiley, New York, 1973.
126. H. M. Nussenzveig, "Introduction to Quantum Optics." Gordon and Breach, New York, 1973.
127. C. W. Gardiner, "Quantum Noise." Springer, Berlin, 1991.
128. L. Mandel and E. Wolf, "Optical Coherence and Quantum Optics." Cambridge University Press, Cambridge, U.K., 1995.
129. M. Orszag, "Quantum Optics." Springer, Berlin, 1999.
130. H. J. Carmichael, "Statistical Methods in Quantum Optics 1: Master Equations and Fokker-Planck Equations." Springer, Berlin, 1999.
131. K. Vogel and H. Risken, *Phys. Rev. A* 40, 2847 (1989).

132. D. F. Smirnov, M. Beck, M. G. Raymer, and A. Faridani, *Phys. Rev. Lett.* 70, 1244 (1993).
133. G. Breitenbach, T. Muller, S. F. Pereira, J.-Ph. Poizat, S. Schiller, and J. Mlynek, *J. Opt. Soc. Am. B* 12, 2304 (1995).
134. G. Breitenbach, S. Schiller, and J. Mlynek, *Nature* 387, 471 (1997).
135. T. J. Dunn, I. A. Walmsley, and S. Mukamel, *Phys. Rev. Lett.* 74, 884 (1995).
136. D. Leibfried, D. M. Meekhof, B. E. King, C. Monroe, W. M. Itano, and D. J. Wineland, *Phys. Rev. Lett.* 77, 4281 (1996); see also *Physics Today* 51, 22 (1998).
137. K. Banaszek and K. Wodkiewicz, *Phys. Rev. Lett.* 76, 4344 (1996).
138. K. Banaszek, C. Radzewicz, K. Wódkiewicz, and J. S. Krasinski, *Phys. Rev. A* 60, 674 (1999).
139. K. Banaszek, A. Dragan, K. Wódkiewicz, and C. Radzewicz, *Phys. Rev. A* 66, 043803 (2002).
140. S. Mancini, P. Tombesi, and V. I. Man'ko, *Europhys. Lett.* 37, 79 (1997).
141. P. Lougovski, E. Solano, Z. M. Zhang, H. Walther, H. Mack, and W. P. Schleich, *Phys. Rev. Lett.* 91, 010401 (2003).
142. L. Davidovich, M. Orszag, and N. Zagury, *Phys. Rev. A* 57, 2544 (1998).
143. G. Nogues, A. Rauschenbeutel, S. Osnaghi, M. Brune, J. M. Raimond, and S. Haroche, *Nature* 400, 239 (1999).
144. G. Nogues, A. Rauschenbeutel, S. Osnaghi, P. Bertet, M. Brune, J. M. Raimond, S. Haroche, L. G. Lutterbach, and L. Davidovich, *Phys. Rev. A* 62, 054101 (2000).
145. A. I. Lvovsky, H. Hansen, T. Aichele, O. Benson, J. Mlynek, and S. Schiller, *Phys. Rev. Lett.* 87, 050402-1 (2001).
146. C. Miquel, J. P. Paz, M. Saraceno, E. Knill, R. Laflamme, and C. Negrevergne, *Nature* 418, 59 (2002).
147. C. Miquel, J. P. Paz, and M. Saraceno, *Phys. Rev. A* 65, 062309 (2002).
148. J. P. Paz, A. J. Roncaglia, and M. Saraceno, *Phys. Rev. A* 69, 032312 (2004).
149. M. Wilkens and P. Meystre, *Phys. Rev. A* 43, 3832 (1991).
150. S. M. Dutra and P. L. Knight, *Phys. Rev. A* 49, 1506 (1994).
151. M. S. Kim, G. Antesberger, C. T. Bodendorf, and H. Walther, *Phys. Rev. A* 58, R65 (1998).
152. J. F. Poyatos, R. Walser, J. I. Cirac, P. Zoller, and R. Blatt, *Phys. Rev. A* 53, R1966 (1996).
153. G. Lindblad, *Math. Phys.* 48, 119 (1976).
154. R. Alicki and K. Lendi, "Quantum Dynamical Semigroups and Applications," Lecture Notes in Physics, Springer-Verlag, Berlin, 1987, No. 286.
155. P. Pechukas, *Phys. Rev. Lett.* 73, 1060 (1994).
156. R. Alicki, *Phys. Rev. Lett.* 75, 3020 (1995).
157. P. Pechukas, *Phys. Rev. Lett.* 75, 3021 (1995).
158. This identity is attributed to F. Bloch, see A. Messiah, "Quantum Mechanics," Ch. XII, Wiley, New York, 1968.
159. M. S. Kim and V. Buzek, *Phys. Rev. A* 46, 4239 (1992).
160. K. Banaszek and K. Wódkiewicz, *Phys. Rev. A* 58, 4345 (1998).
161. K. Banaszek and K. Wódkiewicz, *Phys. Rev. Lett.* 82, 2009 (1999).
162. K. Banaszek and K. Wódkiewicz, *Acta Phys. Slovaca* 49, 491 (1999).
163. A. Kuzmich, I. A. Walmsley, and L. Mandel, *Phys. Rev. A* 64, 063804 (2001).
164. M. França Santos and L. Davidovich, (unpublished).
165. C. H. Bennett, H. J. Bernstein, S. Popescu, and B. Schumacher, *Phys. Rev. A* 53, 2046 (1996).
166. M. França Santos, L. G. Lutterbach, and L. Davidovich, *J. Opt. B: Quantum Semiclass. Opt.* 3, S55 (2001).
167. P. W. Shor, *Phys. Rev. A* 52, 2493 (1995).
168. R. Laflamme, C. Miquel, J. P. Paz, and W. H. Zurek, *Phys. Rev. Lett.* 77, 198 (1996).
169. A. Ekert and C. Macchiavello, *Phys. Rev. Lett.* 77, 2585 (1996).
170. D. Gottesman, *Phys. Rev. A* 54, 1862 (1996).
171. A. R. Calderbank, E. M. Rains, P. M. Shor, and N. J. Sloane, *Phys. Rev. Lett.* 78, 405 (1997).
172. P. Zanardi and M. Rasetti, *Phys. Rev. Lett.* 79, 3306 (1997).
173. P. Zanardi, *Phys. Rev. A* 57, 3276 (1998).
174. L.-M. Duan and G.-C. Guo, *Phys. Rev. A* 57, 737 (1998).
175. P. Zanardi, *Phys. Rev. A* 63, 012301 (2001).
176. P. G. Kwiat, A. J. Berglund, J. B. Altepeter, and A. G. White, *Science* 290, 498 (2000).
177. D. Kielpinski, V. Meyer, M. A. Rowe, C. A. Sackett, W. M. Itano, C. Monroe, and D. J. Wineland, *Science* 291, 1013 (2001).
178. L. Viola en S. Lloyd, *Phys. Rev. A* 58, 2733 (1998).
179. A. R. R. Carvalho, P. Milman, R. L. de Matos Filho, and L. Davidovich, *Phys. Rev. Lett.* 86, 4988 (2001).
180. J. I. Cirac and P. Zoller, *Phys. Rev. Lett.* 74, 4091 (1995).
181. R. L. de Matos Filho and W. Vogel, *Phys. Rev. Lett.* 76, 608 (1996).
182. G. P. Berman and G. M. Zaslavsky, *Physica A* 91, 450 (1978).
183. M. V. Berry, N. L. Balazs, M. Tabor, and A. Voros, *Ann. Phys.* 122, 26 (1979).
184. W. H. Zurek and J. P. Paz, *Phys. Rev. Lett.* 75, 351 (1995).
185. W. H. Zurek, *Phys. Scripta* T76 186 (1998).
186. W. H. Zurek, *Acta Phys. Polon.* B29, 3689 (1998).
187. M. Berry, in "Quantum Mechanics: Scientific Perspectives on Divine Action" (K. W.-M. Robert, J. Russell, P. Clayton, and J. Polkinghorne, Eds.), p. 41, Vatican Observatory - CTNS Publications, 2001.
188. E. Ott, T. M. Antonsen, and J. D. Hanson, *Phys. Rev. Lett.* 53, 2187 (1984).
189. T. Dittrich and R. Graham, *Ann. Phys.* 200, 363 (1990).

190. D. Cohen, *J. Phys. A* 27, 4805 (1994).
191. W. H. Zurek and J. P. Paz, *Phys. Rev. Lett.* 72, 2508 (1994).
192. A. R. Kolovsky, *Phys. Rev. Lett.* 76, 340 (1996).
193. S. Habib, K. Shizume, and W. H. Zurek, *Phys. Rev. Lett.* 80, 4361 (1998).
194. A. K. Pattanayak, *Phys. Rev. Lett.* 88, 4526 (1999).
195. Z. P. Karkuszewski, J. Zakrzewski, and W. H. Zurek, *Phys. Rev. A* 65, 042113 (2002).
196. A. K. Pattanayak, B. Sundaram, and B. D. Greenbaum, *Phys. Rev. Lett.* 90, 014103 (2003).
197. A. Iomin and G. M. Zaslavsky, *Phys. Rev. E* 67, 027203 (2003).
198. A. R. R. Carvalho, R. L. de Matos Filho, and L. Davidovich, *Phys. Rev. E* 70, 026211 (2004).
199. M. Gell-Mann and J. B. Hartle, in "Complexity, Entropy, and the Physics of Information" (W. H. Zurek, Ed.), p. 425. Addison-Wesley, Reading, MA, 1990.
200. A. Blais, R.-S. Huang, A. Wallraff, S. M. Girvin, and R. J. Schoelkopf, *Phys. Rev. A* 69, 062320 (2004); A. Wallraff, D. I. Schuster, A. Blais, L. Frunzio, R.-S. Huang, J. Majer, S. Kumar, S. M. Girvin, and R. J. Schoelkopf, *Nature* 431, 162 (2004).

CHAPTER 7

Optical Properties of Semiconductor Nanostructures: Decoherence versus Quantum Control

Ulrich Hohenester
Institut für Physik, Theoretische Physik, Karl–Franzens–Universität Graz, Graz, Austria

CONTENTS

ISBN: 1-58883-049-7

Handbook of Theoretical and Computational Nanotechnology
Edited by Michael Rieth and Wolfram Schommers
Volume 7: Pages (395–461)

1. INTRODUCTION

Although there is a lot of quantum physics at the nanoscale, one often has to work hard to observe it. In this review, we shall discuss how this can be done for semiconductor quantum dots. These are small islands of lower bandgap material embedded in a surrounding matrix of higher bandgap material. For properly chosen dot and material parameters, carriers become confined in all three spatial directions within the low-bandgap islands on a typical length scale of tens of nanometers. This three-dimensional confinement results in atomic-like carrier states with discrete energy levels. In contrast to atoms, quantum dots are not identical but differ in size and material composition, which results in large inhomogeneous broadenings that usually spoil the direct observation of the atomic-like properties. Optics allows overcoming of this deficiency by means of single-dot or coherence spectroscopy. Once this is accomplished, we fully enter into the quantum world: the optical spectra are governed by sharp and ultranarrow emission peaks—indicating a strong suppression of environment couplings. When more carriers are added to the dot (e.g. by means of charging or non-linear photoexcitation), they mutually interact through Coulomb interactions, which gives rise to intriguing energy shifts of the few-particle states.

This has recently attracted strong interest as it is expected to have profound impact on opto-electronic or quantum-information device applications. A detailed theoretical understanding of such Coulomb-renormalized few-particle states is therefore of great physical interest and importance and will be provided in the first part of this paper. In a nutshell, we find that nature is gentle enough to not bother us too much with all the fine details of the semiconductor materials and the dot confinement, but rather allows for much simpler description schemes. The most simple one, which we shall frequently employ, is borrowed from quantum optics and describes the quantum-dot states in terms of generic few-level schemes. Once we understand the nature of the Coulomb-correlated few-particle states and how they couple to the light, we can start to look closer. More specifically, we shall show that the intrinsic broadenings of the emission peaks in the optical spectra give detailed information about the way the states are coupled to the environment. This will be discussed at the examples of photon and phonon scatterings. Optics can do more than just providing a highly flexible and convenient characterization tool: it can be used as a *control* that allows transfer of coherence from an external laser to the quantum-dot states and to hereby deliberately set the wavefunction of the quantum system. This is successfully exploited in the fields of quantum control and quantum computation, as will be discussed in detail in later parts of the paper.

The field of optics and quantum optics in semiconductor quantum dots has recently attracted researchers from different communities and has benefited from their respective scientific backgrounds. This is also reflected in this paper, where we review genuine solid-state models, such as the rigid-exciton or independent-boson ones, as well as quantum-chemistry schemes, such as configuration interactions or genetic algorithms, or quantum-optics methods, such as the unraveling of the master equation through "quantum jumps" or the adiabatic population transfer. The review is intended to give an introduction to the field and to provide the interested reader with the key references for further details. Throughout, I have tried to explain briefly all concepts and to make the manuscript as self-contained as possible. The paper has been organized as follows. In Section 2, we give a brief overview of the field and introduce the basic concepts. Section 3 is devoted to an analysis of the Coulomb-renormalized few-particle states and of the more simplified few-level schemes for their description. How these states can be probed optically is discussed in Section 4. The coherence and decoherence properties of quantum-dot states are addressed in Section 5, and we show how single-photon sources work. Finally, Sections 6 and 7 discuss quantum-control and quantum-computation applications. To keep the paper as simple as possible, we have postponed several of the computational details to the various appendices.

2. MOTIVATION AND OVERVIEW

2.1. Quantum Confinement

The hydrogen spectrum

$$\epsilon_n = -\frac{E_0}{n^2}, \quad n = 1, 2, \ldots \tag{1}$$

provides a prototypical example for quantized motion: only certain eigenstates characterized by the quantum number n (together with the angular quantum numbers ℓ and m_ℓ) are accessible to the system. Although the detailed form is due to the Coulomb potential exerted by the nucleus, Eq. (1) exhibits two generic features: first, the spectrum ϵ_n is discrete because the electron motion is confined in all three spatial directions; second, the Rydberg energy scale $E_0 = e^2/(2a_0)$ and Bohr length scale $a_0 = \hbar^2/(me^2)$ are determined by the natural constants describing the problem (i.e., the elementary charge e, the electron mass m, and Planck's constant $\hbar$). These phenomena of quantum confinement and natural units prevail for the completely different system of semiconductor quantum dots. These are semiconductor nanostructures where the carrier motion is confined in all three spatial directions [1–3]. Figure 1 sketches two possible types of quantum confinement: in the *weak confinement regime* of Fig. 1a, the carriers are localized at monolayer fluctuations in the thickness of a semiconductor quantum well; in the *strong confinement regime* of Fig. 1b, the carriers are confined within small islands of lower bandgap material embedded in a higher bandgap semiconductor. Although the specific physical properties of these systems can differ drastically, the dominant role of the three-dimensional quantum confinement establishes a common link that will allow us to treat them on the same footing. To highlight this common

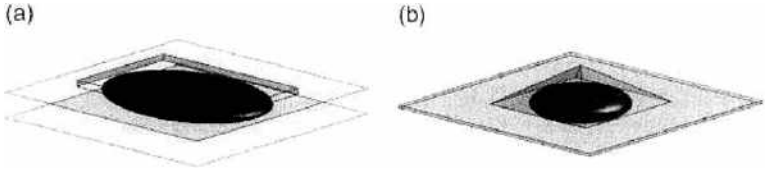


Figure 1. Schematic sketch of the (a) weak and (b) strong confinement regime. In the *weak confinement regime*, the carriers are usually confined at monolayer fluctuations in the width of a narrow quantum well, which form terraces of typical size $100 \times 100 \times 5$ nm^3 [4–8]. In the *strong confinement regime*, the carriers are confined within pyramidal or lens-shaped islands of lower-bandgap material, usually formed in strained layer epitaxy, with typical spatial extensions of $10 \times 10 \times 5$ nm^3 [2, 9–11].

perspective as well as the similarity to atoms, in the following we shall frequently refer to quantum dots as *artificial atoms*. In the generalized expressions for Rydberg and Bohr

$$E_s = e^2/(2\kappa_s a_s) \sim 5\,\text{meV} \quad \ldots\text{semiconductor Rydberg} \tag{2}$$

$$a_s = \hbar^2\kappa_s/(m_s e^2) \sim 10\,\text{nm} \quad \ldots\text{semiconductor Bohr} \tag{3}$$

we account for the strong dielectric screening in semiconductors, $\kappa_s \sim 10$, and the small electron and hole effective masses $m_s \sim 0.1\,m$ [12, 13]. Indeed, Eqs. (2) and (3) provide useful energy and length scales for artificial atoms: the carrier localization length ranges from 100 nm in the weak to about 10 nm in the strong confinement regime and the primary level splitting from 1 meV in the weak to several tens of meV in the strong confinement regime.

The artificial-atom picture can further be extended to optical excitations. Quite generally, when an undoped semiconductor is optically excited, an electron is promoted from a valence to a conduction band. In the usual language of semiconductor physics, this process is described as the creation of an *electron-hole pair* [12, 13]: the electron describes the excitation in the conduction band, and the hole accounts for the properties of the missing electron in the valence band. Conveniently, electron and hole are considered as independent particles with different effective masses, which mutually interact through the attractive Coulomb interaction. What happens when an electron-hole pair is excited inside a semiconductor quantum dot? Things strongly differ for the weak and strong confinement regime: in the first case, the electron and hole form a Coulomb-bound electron–hole complex—the so-called *exciton* [12]—whose center-of-mass motion becomes localized and quantized in presence of the quantum confinement; in the latter case confinement effects dominate over the Coulomb ones and give rise to electron–hole states with dominant single-particle character. However, in both cases the generic feature of quantum confinement gives rise to discrete, atomic-like absorption and emission-lines—and thus allows for the artificial-atom picture advocated above.

2.2. Scope of the Chapter

Quantum systems can usually not be measured directly. Rather, one has to perturb the system and measure indirectly how it reacts to the perturbation. This is schematically shown in Fig. 2 (shaded boxes): an external perturbation (e.g., a laser field) acts upon the quantum system and promotes it from the ground to an excited state; the excitation decays through environment coupling (e.g., photo emission) and finally a measurement is performed on

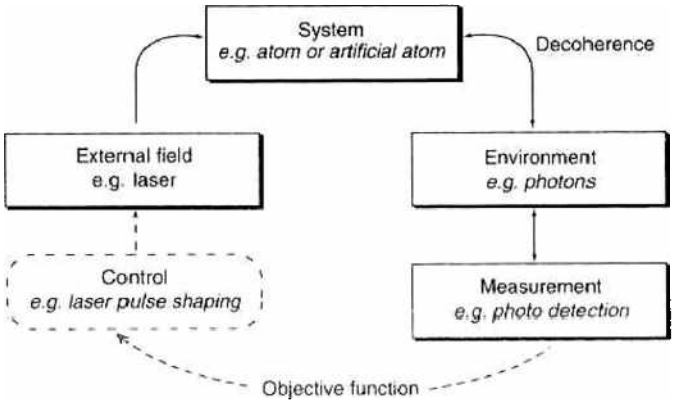


Figure 2. Schematic representation of optical spectroscopy (shaded boxes) and quantum control (dashed lines): an external field acts upon the system and promotes it from the ground to an excited state; the excitation decays through environment coupling (e.g., photo emission) and a measurement is performed indirectly on the environment (e.g., photo detection). In case of quantum control, the perturbation is tailored such that a given objective (e.g., the wish to channel the system from one state to another) is fulfilled in the best way. This is usually accomplished by starting with some initial guess for the external field, and to improve it by exploiting the outcome of the measurement. The arrows in the figure indicate the flow of information.

some part of the environment (e.g., through photo detection). As we shall see, this mutual interaction between system and environment—the environment influences the system and in turn becomes influenced by it (indicated by the arrows in Fig. 2, which show the flow of information)—plays a central role in the understanding of decoherence and the measurement process [14].

Optical spectroscopy provides one of the most flexible measurement tools, as it allows for a remote excitation and detection. It gives detailed information about the system and its environment. This is seen most clearly at the example of atomic spectroscopy, which played a major role in the development of quantum theory and quantum electrodynamics [15] and most recently has even been invoked in the search for nonconstant natural constants [16]. In a similar, although somewhat less fundamental manner, spectroscopy of artificial atoms allows for a detailed understanding of both electron–hole states (Section 4) and of the way these states couple to their environment (Section 5). On the other hand, such detailed understanding opens the challenging perspective to use the external fields in order to *control* the quantum system. More specifically, the coherence properties of the exciting laser are transfered to quantum coherence in the system, which allows deliberate setting of the state of the quantum system (see lower part of Fig. 2). Recent years have seen spectacular examples of such light-matter manipulations in atomic systems, for example, Bose–Einstein condensation or freezing of light (see, e.g., Chu [17] and references therein). This tremendous success also initiated great stimulus in the field of solid-state physics, as we shall discuss for artificial atoms in Section 6. More recently, the emerging fields of quantum computation [18–20] and quantum communication [21] have become another driving force in the field. They have raised the prospect that an almost perfect quantum control would allow for computation schemes that would outperform classical computation. In turn, a tremendous quest for suited quantum systems has started, ranging from photons over molecules, trapped ions and atomic ensembles, to semiconductor quantum dots. We will briefly review some proposals and experimental progress in Section 7.

2.3. Quantum Coherence

Quantum coherence is the key ingredient and workhorse of quantum control and quantum computation. To understand its essence, let us consider a generic two-level system with ground state $|0\rangle$ and excited state $|1\rangle$ (e.g., an artificial atom with one electron–hole pair absent or present). The most general wavefunction can be written in the form

$$\alpha|1\rangle + \beta|0\rangle \tag{4}$$

with α and β arbitrary complex numbers subject to the condition $|\alpha|^2 + |\beta|^2 = 1$. Throughout this paper, we shall prefer the slightly different description scheme of the *Bloch vector* picture [12, 15, 22]. Because the state (4) is unambiguously defined only up to an arbitrary phase factor—which can for instance be used to make α real—it can be characterized by three real numbers. A convenient representation is provided by the Bloch vector

$$\boldsymbol{u} = \begin{pmatrix} 2\Re e(\,\alpha^*\beta\,) \\ 2\Im m(\,\alpha^*\beta\,) \\ |\alpha|^2 - |\beta|^2 \end{pmatrix} \tag{5}$$

where the z-component accounts for the population inversion, which gives the probability for finding the system in either the upper or lower state, and the x and y components account for the phase relation between α and β (i.e., the *quantum coherence*). As we shall see, this coherence is at the heart of quantum computation and is responsible for such characteristic quantum features as interference or entanglement. For an isolated system whose dynamics is entirely *coherent* (i.e., completely governed by Schrödinger's equation), the norm of the Bloch vector is conserved. A pictorial description is provided by the *Bloch sphere* shown in Fig. 3, where in case of a coherent evolution $\boldsymbol{u}$ always stays on the surface of the sphere.

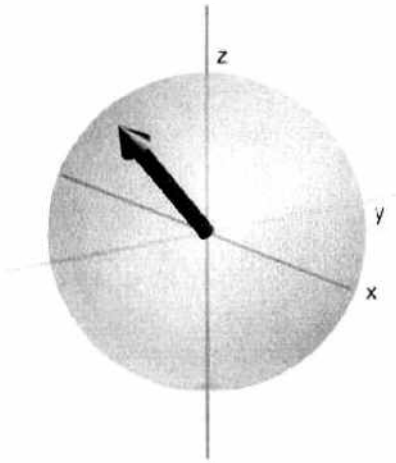


Figure 3. Schematic representation of the Bloch vector $\boldsymbol{u}$. The z component accounts for the population inversion and gives the probability for finding the system in either the upper or lower state. The x and y components account for *quantum coherence* (i.e., the phase relation between the upper and lower state), which is responsible for quantum-interference effects. For the *coherent* time evolution of an isolated quantum system, $\boldsymbol{u}$ stays at the surface of the Bloch sphere. For an incoherent time evolution in presence of environment couplings, $\boldsymbol{u}$ dips into the Bloch sphere: the system *decoheres*.

2.4. Decoherence

Isolated quantum systems are idealizations that cannot be realized in nature, as any quantum system interacts with its environment. In general, such environment couplings corrupt the quantum coherence and the system suffers *decoherence*. Strictly speaking, decoherence can no longer be described by Schrödinger's equation but calls for a more general density-matrix description, within which, as will be shown in Section 5, the x and y components of the Bloch vector are diminished—$\boldsymbol{u}$ dips into the Bloch sphere. The most simple description for the evolution of the Bloch vector in presence of environment couplings is given by [12, 15]

$$\dot{u}_1 = -\frac{u_1}{T_2}, \quad \dot{u}_2 = -\frac{u_2}{T_2}, \quad \dot{u}_3 = -\frac{u_3 + 1}{T_1} \tag{6}$$

where the first two equations account for the above-mentioned decoherence losses, and the last one for relaxation where the system is scattered from the excited to the ground state because of environment couplings. T_1 and T_2 are the relaxation and decoherence time, sometimes referred to as *longitudinal* and *transverse* relaxation times. They are conveniently calculated within the framework of Fermi's golden rule, where

$$\left(\frac{1}{T}\right) = 2\pi \int D(\omega) d\omega \, g^2 \, \delta(E_i - E_f - \omega) \tag{7}$$

accounts for the scattering from the initial state i to the final state f through creation of an environment excitation with energy ω, for example, photon; $D(\omega)$ is the density of states and g the matrix element associated to the interaction. In semiconductors of higher dimension, T_2 is always much shorter than T_1 [12, 27] because all elastic scatterings (i.e., processes where no energy is exchanged, such as impurity or defect scatterings) contribute to decoherence, whereas only inelastic scatterings (i.e., processes where energy is exchanged, such as phonon or mutual carrier scatterings) contribute to relaxation. On very general grounds, one expects that scatterings in artificial atoms become strongly suppressed and quantum coherence substantially enhanced: this is because in higher dimensional semiconductors, carriers can be scattered from a given initial state to a continuum of final states—and therefore couple to all environment modes ω—whereas in artificial atoms the atomic-like density of states only allows for a few selective scatterings with $\omega = E_i - E_f$. Indeed, in the coherence experiment of Bonadeo et al. [28] the authors showed for a quantum dot in the weak confinement regime that the broadening of the optical emission peaks is completely lifetime limited, that is, $T_2 \cong 2T_1 \sim 40$ ps—a remarkable finding in view of the extremely short

Table 1. Relations between T_1 and T_2 and typical scattering times for spontaneous photon emission and phonon-assisted dephasing [23–25]. In the last two columns, we report experimental values measured in the weak and strong confinement regime, respectively.

Interaction mechanism	Relation	Weak	Strong
Photon emission	$T_2 = 2T_1$	~40 ps [26]	ns [24]
Phonon dephasing	$T_1 \to \infty$	?	~5 ps [24]

sub-picosecond decoherence times in conventional semiconductor structures. Similar results were also reported for dots in the strong confinement regime [24]. However, there it turned out that at higher temperatures a decoherence channel dominates, which is completely ineffective in higher dimensional systems. An excited electron–hole pair inside a semiconductor provides a perturbation to the system and causes a slight deformation of the surrounding lattice. As will be discussed in Section 5.5, in many cases of interest this small deformation gives rise to decoherence but not relaxation. A T_2-estimate and some key references are given in Table 1.

2.5. Quantum Control

Decoherence in artificial atoms is much slower than in semiconductors of higher dimension because of the atomic-like density of states. Yet, it is substantially faster than in atoms where environment couplings can be strongly suppressed by working at ultrahigh vacuum—a procedure not possible for artificial atoms that are intimately incorporated in the surrounding solid-state environment. Let us consider for illustration a situation where a two-level system initially in its groundstate is excited by an external laser field tuned to the 0–1 transition. As will be shown in Section 4, the time evolution of the Bloch vector in presence of a driving field is of the form

$$\dot{\boldsymbol{u}} = -\Omega \hat{\boldsymbol{e}}_1 \times \boldsymbol{u} \tag{8}$$

with the Rabi frequency Ω determining the strength of the light-matter coupling and $\hat{\boldsymbol{e}}_1$ the unit vector along x. Figure 4a shows the trajectory of the Bloch vector that is rotated from the south pole $-\hat{\boldsymbol{e}}_3$ of the Bloch sphere through the north pole, until it returns after a certain time (given by the strength Ω of the laser) to the initial position $-\hat{\boldsymbol{e}}_3$. Because of the 2π-rotation of the Bloch vector, such pulses are called 2π-pulses. If the Bloch vector evolves in presence of environment coupling, the two-level system becomes entangled with the environmental degrees of freedom and suffers decoherence. This is shown in Fig. 4b for the phonon-assisted decoherence described above: while rotating over the Bloch sphere the length of $\boldsymbol{u}$ decreases, and the system does not return to its original position. Quite generally, in the process of decoherence it takes some time for the system to become entangled with its environment. If during this entanglement buildup the system is acted upon by an appropriately designed

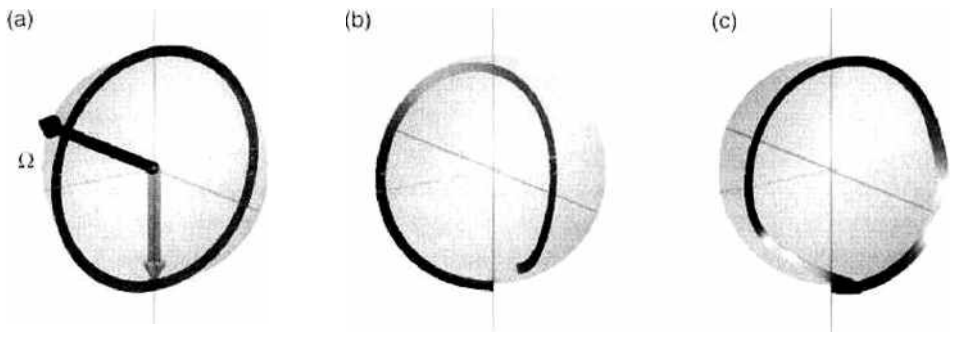


Figure 4. Trajectories of the Bloch vector $\boldsymbol{u}$ for a 2π-pulse and for (a) an isolated two-level system [$\boldsymbol{\Omega} = -\Omega\hat{\boldsymbol{e}}_1$ is the vector defined in Eq. (8)]; (b) a two-level system in presence of phonon-assisted dephasing (see Section 5.5) and a Gaussian pulse envelope: because of decoherence the length of $\boldsymbol{u}$ decreases [29, 30]; (c) same as (b) but for an optimal-control pulse envelope (for details, see Section 6.2).

Table 2. Summary of some important parameters for semiconductor quantum dots in the weak and strong confinement regime. The primary and fine-structure splittings refer to the level splittings associated to charge and spin degrees of freedom, respectively. (for details see text).

Property	Confinement	Group	Method	Result
Carrier localization length	Weak	Matsuda et al. [8]	Near-field microscopy	$100 \times 100 \times 5$ nm^3
Inhomogeneous broadening	Weak	Guest et al. [31]	Optical spectroscopy	1–5 meV
Primary level splitting	Weak	Guest et al. [31]	Local spectroscopy	0.1–1 meV
Fine-structure splitting	Weak	Tischler et al. [32]	Magnetoluminescence	10–100 μeV
Exciton decoherence time	Weak	Bonadeo et al. [26, 28]	Coherence spectroscopy	40 ps
Carrier localization length	Strong	Bimberg et al. [2]		$10 \times 10 \times 5$ nm^3
Inhomogeneous broadening	Strong	Bimberg et al. [2]		10–50 meV
Primary level splitting	Strong	Bimberg et al. [2]		10–100 meV
Fine-structure splitting	Strong	Bayer et al. [33]	Magnetoluminescence	10–100 μeV
Exciton decoherence time	Strong	Borri et al. [24]	Four-wave mixing	~ns[a]
Spin relaxation time	Strong	Lenihan et al. [34]	Coherence spectroscopy	750 ps
	Strong	Lenihan et al. [34]	Coherence spectroscopy	$\geq$40 ns
	Strong	Paillard et al. [35]	Luminescence	"Frozen carrier spins"

[a]There exists an additional phonon-assisted decoherence that is strongly temperature-dependent and can dominate at higher temperatures.

control, it becomes possible to channel back quantum coherence from the environment to the system and to suppress decoherence. This is shown in Fig. 4 for an optimized laser field—for details see Section 6.2—which drives $\boldsymbol{u}$ from the south pole through a sequence of excited states back to the initial position without suffering any decoherence losses. Alternatively, in presence of strong laser fields, the quantum dot states become renormalized, which can be exploited for efficient population transfers. Thus, quantum control allows to suppress or even overcome decoherence losses. In Section 6 we will discuss prototypical quantum-control applications and ways to combat decoherence in the solid state.

2.6. Properties of Artificial Atoms

The picture we have developed so far describes artificial atoms in terms of effective few-level schemes. They can be characterized by a few parameters, which can be either obtained from *ab initio* type calculations or can be inferred from experiment. Table 2 reports some of the relevant parameters for artificial atoms. For both the weak and strong confinement regime, the inhomogeneous broadening due to dot-size fluctuations is comparable to the primary level splittings themselves, and spectroscopy of single dots (Section 4) is compulsory to observe the detailed primary and fine-structure splittings. Decoherence and relaxation times for electron-hole states range from tens to hundreds of picoseconds, which is surprisingly long for the solid state. This is because of the atomic-like density of states and the resulting inhibition of mutual carrier scatterings and the strong suppression of phonon scatterings. Yet, when it comes to more sophisticated quantum-control or quantum-computation applications (Sections 6 and 7) such sub-nanosecond relaxation and decoherence appears to be quite limiting. A possible solution may be provided by spin excitations, with their long lifetimes because of weak solid-state couplings.

3. FEW-PARTICLE STATES

Electron-hole states in semiconductor quantum dots can be described at different levels of sophistication, ranging from *ab initio* type approaches over effective solid-state models to generic few-level schemes. All these approaches have their respective advantages and

disadvantages. For instance, *ab initio* type approaches provide results that can be quantitatively compared with experiment, but require a detailed knowledge of the confinement potential which is difficult to obtain in many cases of interest and often give only little insight into the general physical trends. On the other hand, few-level schemes grasp all the essential features of certain electron-hole states in a most simple manner, but the relevant parameters have to be obtained from either experiment or supplementary calculations. Depending on the physical problem under consideration, we shall thus chose between these different approaches. A description in terms of complementary models is not at all unique to artificial atoms, but has proven to be a particularly successful concept for many-electron atoms. These are highly complicated objects whose physical properties depend on such diverse effects as spin-orbit coupling, exchange interactions, or Coulomb correlations—and thus make first-principles calculations indispensable for quantitative predictions. On the other hand, in the understanding of the *aufbau* principle of the periodic table, it suffices to rely on just a few general rules, such as Pauli's principle, Hund's rules for open-shell atoms, and Coulomb correlation effects for transition metals. Finally, for quantum optics calculations, one usually invokes generic few-level schemes, for example, the celebrated Λ- and V-type ones, where all details of the relevant states are lumped into a few effective parameters. As we shall see, similar concepts can successfully be extended to semiconductor quantum dots. In the remainder of this section, we shall discuss how this is done.

Throughout, we assume that the carrier states in semiconductor quantum dots are described within a many-body framework such as density functional theory [36], and can be described by the effective single-particle Schrödinger equation ($\hbar = 1$ throughout)

$$\left(-\frac{\nabla^2}{2m} + U(\boldsymbol{r})\right)\psi(\boldsymbol{r}) = \epsilon\psi(\boldsymbol{r}) \tag{9}$$

In the parentheses on the left-hand side, the first term accounts for the kinetic energy, where m is the free electron mass, and the second one for the atomic-like potential of the crystal structure. For an ideal periodic solid-state structure, the eigenstates $\psi_{n\boldsymbol{k}}(\boldsymbol{r}) = u_{n\boldsymbol{k}}(\boldsymbol{r})\exp(i\boldsymbol{k}\boldsymbol{r})$ are given by the usual Bloch function u, with n the band index and $\boldsymbol{k}$ the wavevector, and the eigenenergies $\epsilon_{n\boldsymbol{k}}$ provide the semiconductor bandstructure [12, 13, 36]. How are things modified for semiconductor nanostructures? In this paper, we shall be concerned with quantum dots with spatial extensions of typically tens of nanometers in each direction, which consist of approximately 1 million atoms. This suggests that the detailed description of the atomic potential $U(\boldsymbol{r})$ of Eq. (9) is not needed and can safely be replaced by a more phenomenological description scheme. A particularly simple and successful one is provided by the *envelope-function approach* [12, 13], which assumes that the single-particle wavefunctions $\psi(\boldsymbol{r})$ are approximately given by the Bloch function u of the ideal lattice modulated by an envelope part $\phi(\boldsymbol{r})$ that accounts for the additional quantum confinement. In the following, we consider direct *III–V* semiconductors (e.g., GaAs or InAs), whose conduction and valence band extrema are located at $\boldsymbol{k} = 0$ and describe the band structure near the minima by means of effective masses $m_{e,h}$ for electrons and holes. Then,

$$\left(-\frac{\nabla^2}{2m_i} + U_i(\boldsymbol{r})\right)\phi_\lambda^i(\boldsymbol{r}) = \epsilon_\lambda^i\phi_\lambda^i(\boldsymbol{r}) \tag{10}$$

approximately accounts for the electron and hole states in presence of the confinement. Here, $U_{e,h}(\boldsymbol{r})$ is the effective confinement potential for electrons or holes, and m_i is the effective mass of electrons or holes that may depend on position (e.g., to account for the different semiconductor materials in the confinement of Fig. 1b). In the literature, numerous theoretical work—mostly based on the $k\cdot p$ [37–40] or empirical pseudopotential framework [41–43]—has been concerned with more sophisticated calculation schemes for single particle states. These studies have revealed a number of interesting peculiarities associated to effects such as piezoelectric fields, strain, or valence-band mixing, but have otherwise supported the results derived within the more simple-minded envelope-function and effective-mass description scheme for few-particle states in artificial atoms.

3.1. Excitons

3.1.1. Semiconductors of Higher Dimension

What happens for optical electron-hole excitations that experience in addition to the quantum confinement also Coulomb interactions? We first recall the description of a Coulomb-correlated electron–hole pair inside a bulk semiconductor. Within the envelope-function and effective mass approximations

$$H = -\sum_{i=e,h} \frac{\nabla_{r_i}^2}{2m_i} - \frac{e^2}{\kappa_s |\boldsymbol{r}_e - \boldsymbol{r}_h|} \tag{11}$$

is the Hamiltonian for the interacting electron–hole system, with κ_s the static dielectric constant of the bulk semiconductor. In the solution of Eq. (11), one usually introduces the center-of-mass and relative coordinates $\boldsymbol{R} = (m_e \boldsymbol{r}_e + m_h \boldsymbol{r}_h)/M$ and $\boldsymbol{\rho} = \boldsymbol{r}_e - \boldsymbol{r}_h$ [12], and decomposes $H = \mathscr{H} + h$ into the parts

$$\mathscr{H} = -\frac{\nabla_{\boldsymbol{R}}^2}{2M}, \qquad h = -\frac{\nabla_{\boldsymbol{\rho}}^2}{2\mu} - \frac{e^2}{\kappa_s |\boldsymbol{\rho}|} \tag{12}$$

with $M = m_e + m_h$ and $\mu = m_e m_h / M$. Correspondingly, the total wavefunction can be decomposed into parts $\Phi(\boldsymbol{R})$ and $\phi(\boldsymbol{\rho})$ associated to the center-of-mass and relative motion, respectively, whose solutions are provided by the Schrödinger equations

$$\mathscr{H}\Phi(\boldsymbol{R}) = \mathscr{E}\Phi(\boldsymbol{R}), \qquad h\phi(\boldsymbol{\rho}) = \epsilon\phi(\boldsymbol{\rho}) \tag{13}$$

Here, the first equation describes the motion of a free particle with mass M and the second one the motion of a particle with mass μ in a Coulomb potential $-e^2/\kappa_s|\boldsymbol{\rho}|$. The solutions of the latter equation are those of the hydrogen atom but for the modified Rydberg energy E_s and Bohr radius a_s of Eqs. (2) and (3). Similar results apply for the lower dimensional quantum wells and quantum wires, provided that $\phi(\rho)$ is replaced by the corresponding two- and one-dimensional wavefunction, respectively. For instance [12],

$$\phi_0(\boldsymbol{\rho}) \cong \frac{4}{a_s} \exp\left(-\frac{2\rho}{a_s}\right) \tag{14}$$

is the approximate groundstate wavefunction for a two-dimensional quantum well, whose energy is $\epsilon_0 = -4E_s$. For a quantum well of finite width, Eq. (14) only accounts for the in-plane part of the exciton wavefunction. If the quantum well is sufficiently narrow, the total wavefunction is approximately given by the product of (14) with the single-particle wavefunctions for electrons and holes along z—that is, those of a "particle in the box" [12]—and the exciton energy is the sum of ϵ_0 with the single-particle energies for the z-motion of electrons and holes [44–46].

3.1.2. Semiconductor Quantum Dots

How are the results of the previous section modified in presence of additional quantum confinements $U_e(\boldsymbol{r}_e)$ and $U_h(\boldsymbol{r}_h)$ for electrons and holes? In analogy to Eq. (11) we describe the interacting electron and hole subject to the quantum-dot confinement through the Hamiltonian

$$H = \sum_{i=e,h} \left[-\frac{\nabla_{r_i}^2}{2m_i} + U_i(\boldsymbol{r}_i) \right] - \frac{e^2}{\kappa_s |\boldsymbol{r}_e - \boldsymbol{r}_h|} \tag{15}$$

The first term on the right-hand side accounts for the motion of the carriers in presence of U_i. Because of the additional terms $U_i(\boldsymbol{r}_i)$, a separation into center-of-mass and relative motion is no longer possible. Provided that the potentials are sufficiently strong, the carrier motion becomes confined in all three spatial directions. Suppose that L is a characteristic confinement length. Then two limiting cases can be readily identified in Eq. (15): in case of *weak confinement* where $L \gg a_s$, the dynamics of the electron–hole pair is dominated by the

Coulomb attraction, and the confinement potentials $U_i(\boldsymbol{r}_i)$ only provide a weak perturbation; in the opposite case of *strong confinement* where $L \ll a_x$, confinement effects dominate, and the Coulomb part of Eq. (15) can be treated perturbatively. In the following we shall discuss both cases in slightly more detail.

3.1.2.1. Weak Confinement Regime We first consider the weak-confinement regime. A typical example is provided by monolayer interface fluctuations in the width of a semiconductor quantum well, as depicted in Fig. 1a, where the electron–hole pair becomes confined within the region of increased quantum-well thickness [4–6, 8, 31, 48–50]. If the resulting confinement length L is much larger than the Bohr radius a_x, the correlated electron–hole wavefunction factorizes into a center-of-mass and relative part, where, to a good degree of approximation, the relative part is given by the wavefunction of the quantum well ("rigid-exciton approximation" [46]). It then becomes possible to integrate over $\boldsymbol{\rho}$ and to recover an effective Schrödinger equation for the exciton center-of-mass motion (for details, see Appendix A)

$$\left(-\frac{\nabla_{\boldsymbol{R}}^2}{2M} + \overline{U}(\boldsymbol{R})\right)\Phi_\nu(\boldsymbol{R}) = \mathcal{E}_\nu \Phi_\nu(\boldsymbol{R}) \tag{16}$$

where $\overline{U}(\boldsymbol{R})$ is a potential obtained through convolution of $U_e(\boldsymbol{r}_e)$ and $U_h(\boldsymbol{r}_h)$ with the two-dimensional exciton wavefunction (14). Figure 5 shows for a prototypical square-like confinement the corresponding $\overline{U}(\boldsymbol{R})$, which only depicts small deviations from the rectangular shape. The corresponding wavefunctions and energies closely resemble those of a particle in a box. Figure 6 shows for the confinement potential depicted in Fig. 5 the square modulus of (a) the s-like groundstate, (b,c) the p-like excited states with nodes along x and y, and (d) the third excited state with two nodes along x. A word of caution is in order. Despite the single-particle character of the envelope-part $\Phi(\boldsymbol{R})$ of the exciton wavefunction, that of the total wavefunction $\Phi(\boldsymbol{R})\,\phi_0(\boldsymbol{r})$ is dominated by Coulomb correlations. This can easily be seen by comparing the length scale of single-particle states $L_n \sim L/n$ (n is the single-particle quantum number) with the excitonic Bohr radius a_x. To resolve spatially the variations of $\phi_0(\boldsymbol{\rho})$ on the length scale of a_x, we have to include states up to $L_n \sim a_x$. Hence, $n \sim L/a_x$ which, because of $L \gg a_x$ in the weak-confinement regime, is a large number.

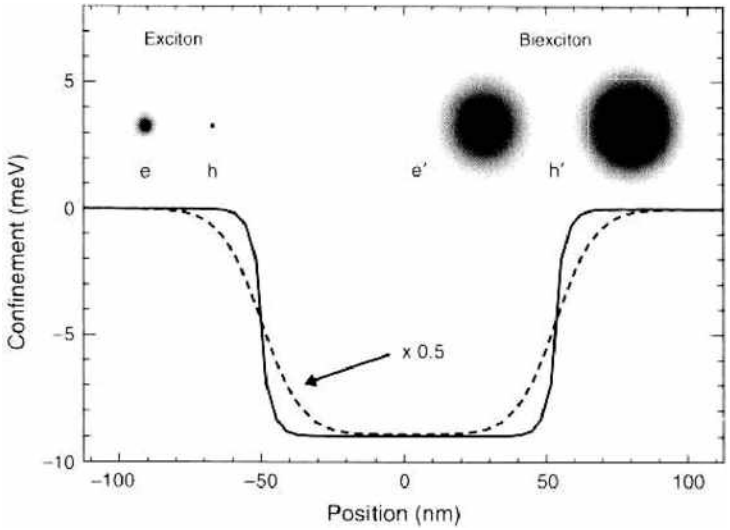


Figure 5. Confinement potential along x for the center-of-mass motion of excitons (solid line) and biexcitons (dashed line). The insets report the probability distributions for finding an (e) electron or (h) hole at a given distance from the center-of-mass coordinate $\boldsymbol{R}$; (e',h') same for biexcitons—the reduced probability at the center of (h') is attributed to the repulsive part χ of the trial wavefunction. In the calculations, we use material parameters representative for GaAs and assume an interface-fluctuation confinement of rectangular shape with dimensions 100×70 nm^2, and monolayer fluctuations of a 5 nm thick quantum well [47].

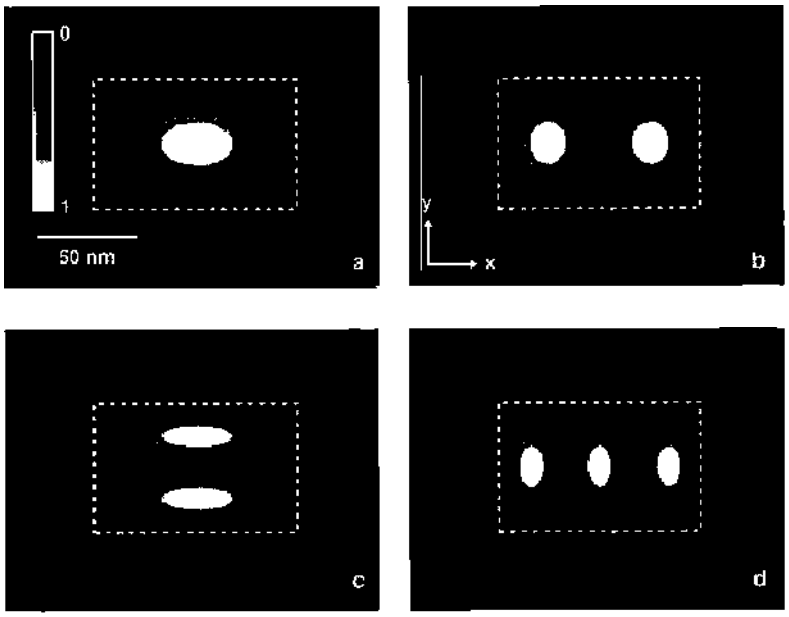


Figure 6. Contour map showing the square modulus of the wavefunction $\Phi_\lambda(\boldsymbol{R})$ for the center-of-mass motion of (a) the s-type groundstate, (b,c) the p-type first excited states with nodes along x and y, and (d) the third excited state with two nodes along x. We use material parameters of GaAs and the confinement potential depicted in Fig. 5.

3.1.2.2. Strong Confinement Regime Things are completely different in this strong-confinement regime where the confinement length is smaller than the excitonic Bohr radius a_x. This situation approximately corresponds to that of most types of self-assembled quantum dots [1–3] where carriers are confined in a region of typical size $10 \times 10 \times 5$ nm^3. To the lowest order of approximation, the groundstate Ψ_0 of the interacting electron–hole system is simply given by the product of electron and hole single-particle states of lowest energy (see also Fig. 7)

$$\Psi_0(\boldsymbol{r}_e, \boldsymbol{r}_h) \cong \phi_0^e(\boldsymbol{r}_e)\phi_0^h(\boldsymbol{r}_h) \tag{17}$$

$$E_0 \cong \epsilon_0^e + \epsilon_0^h - \int d\boldsymbol{r}_e d\boldsymbol{r}_h \frac{|\phi_0^e(\boldsymbol{r}_e)|^2 |\phi_0^h(\boldsymbol{r}_h)|^2}{\kappa_s |\boldsymbol{r}_e - \boldsymbol{r}_h|} \tag{18}$$

Here, the groundstate energy E_0 is the sum of the electron and hole single-particle energies reduced by the Coulomb attraction between the two carriers. Excited electron-hole states Ψ_x can be obtained in a similar manner by promoting the carriers to excited single-particle states. In many cases, the wavefunction *ansatz* of Eq. (17) is oversimplified. In particular, when the confinement length is comparable to the exciton Bohr radius a_x, the

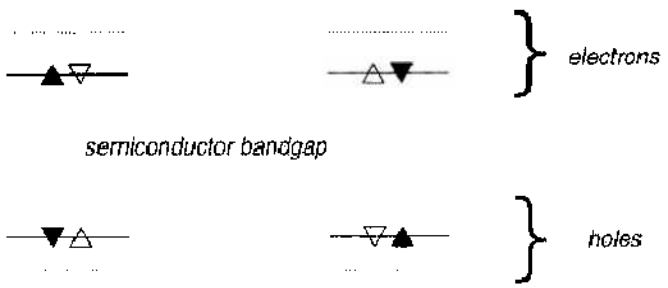


Figure 7. Schematic representation of the two possible groundstates of bright excitons in the strong confinement regime. The solid lines indicate the single-particle states of lowest energy and the dotted lines the first excited states. Excited exciton states can be obtained by promoting the electron or hole to excited single-particle states. The black triangles in the left and right panel indicate the different spin orientations of the electron and hole, as discussed in more detail in Section 3.1.3.

electron–hole wavefunction can no longer be written as a simple product (17) of two single-particle states. We shall now briefly discuss how an improved description can be obtained. To this end, we introduce the fermionic field operators $c_\mu^\dagger$ and $d_\nu^\dagger$ which, respectively, describe the creation of an electron in state μ or a hole in state ν (for details, see Appendix B.1). The electron–hole wavefunction of Eq. (17) can then be written as

$$c_{0_e}^\dagger \, d_{0_h}^\dagger \, |0\rangle \tag{19}$$

where $|0\rangle$ denotes the semiconductor vacuum (i.e., no electron-hole pairs present), and 0_e and 0_h denote the electron and hole single-particle states of lowest energy. Although Eq. (19) is an eigenstate of the single-particle Hamiltonian, it is only an approximate eigenstate of the Coulomb Hamiltonian. We now follow Hawrylak [51] and consider a simplified quantum-dot confinement with cylinder symmetry. The single-particle states can then be labeled by their angular momentum quantum numbers, where the groundstate 0 has s-type symmetry and the degenerate first excited states ± 1 have p-type symmetry. Because Coulomb interactions preserve the total angular momentum [3], the only electron–hole states coupled by Coulomb interactions to the groundstate are those indicated in Fig. 8. We have used that the angular momentum of the hole is opposite to that of the missing electron. Within the electron–hole basis $|0\rangle = |0_e, 0_h\rangle$ and $|\pm 1\rangle = |\pm 1_e, \pm 1_h\rangle$ the full Hamiltonian matrix is of the form

$$E_0 + \begin{pmatrix} 0 & V_{sp} & V_{sp} \\ V_{sp} & \Delta_p & V_{pp} \\ V_{sp} & V_{pp} & \Delta_p \end{pmatrix} \tag{20}$$

where E_0 is the energy (18) of the exciton groundstate, Δ_p the detuning of the first excited state in absence of Coulomb mixing, and V_{sp} and $V_{pp'}$ describe the Coulomb couplings between electrons and holes in the s and p shells (Fig. 8). The Coulomb renormalized eigenstates and energies can then be obtained by diagonalizing the matrix (20). Results of such configuration-interaction calculations will be presented in Section 4 (see Appendix B for more details).

3.1.3. Spin Structure

Besides the orbital degrees of freedom described by the envelope part of the wavefunction, the atomic part additionally introduces spin degrees of freedom. For *III–V* semiconductors, an exhaustive description of the band structure near the minima (at the so-called Γ point) is provided by an eight-band model [12, 52] containing the s-like conduction band states $|s, \pm\frac{1}{2}\rangle$ and the p-like valence band states $|\frac{3}{2}, \pm\frac{3}{2}\rangle$, $|\frac{3}{2}, \pm\frac{1}{2}\rangle$, and $|\frac{1}{2}, \pm\frac{1}{2}\rangle$ [note that these s- and p-states refer to the atomic orbitals and have nothing to do with those introduced in Eq. (20)]. In the problem of our present concern, four of the six valence band states can be approximately neglected: first, the $|\frac{1}{2}, \pm\frac{1}{2}\rangle$ ones, which are energetically split off by a few hundred meV because of spin-orbit interactions [12, 13, 52, 53]; second, the states

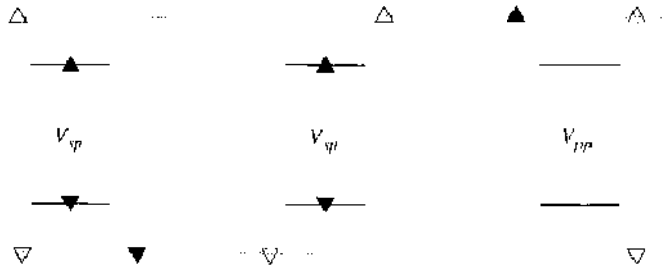


Figure 8. Schematic sketch of the configuration-interaction calculation of the Coulomb-correlated electron–hole states in the restricted single-particle basis of the ground states 0 and the first excited states -1 (left) and $+1$ (right); the numbers correspond to the angular momenta of electrons and holes. The figures show the allowed Coulomb transitions between different electron and hole states, indicated by the filled and open triangles.

$|\frac{3}{2}, \pm\frac{1}{2}\rangle$ associated to the *light-hole* band, which are energetically split off in case of a strong quantum confinement along the growth direction z (e.g., those shown in Fig. 1).

Thus, the atomic part of the electron and hole states of lowest energy is approximately given by the s-type conduction band states $|s, \pm\frac{1}{2}\rangle$ and the p-type states $|\frac{3}{2}, \pm\frac{3}{2}\rangle$ associated to the *heavy-hole* band. From these two electron and hole states, we can form four possible electron–hole states $|\pm\frac{1}{2}, \pm\frac{3}{2}\rangle$ and $|\pm\frac{1}{2}, \mp\frac{3}{2}\rangle$, where the entries account for the z-projection of the total angular momentum m_j for the electron and hole, respectively. A word of caution is in order: because the hole describes the properties of the *missing* electron in the valence band, its m_j value is opposite of that of the corresponding valence band state. For that reason, the usual optical selection rules $\Delta j = 0$ and $\Delta m_j = \pm 1$ [12] for the optical transitions under consideration translate to the matrix elements

$$\left\langle 0 \middle| er \middle| \frac{1}{2}, -\frac{3}{2} \right\rangle = \mu_0 \hat{e}_+, \quad \left\langle 0 \middle| er \middle| -\frac{1}{2}, \frac{3}{2} \right\rangle = \mu_0 \hat{e}_- \tag{21}$$

with μ_0 the optical dipole matrix element, $|0\rangle$ the semiconductor vacuum, and $\hat{e}_\pm$ the polarization vector for left- or right-handed circularly polarized light. Below, we shall refer to hole states with $m_j = \pm\frac{3}{2}$ as holes with spin-up or spin-down orientation and to exciton states $|\pm\frac{1}{2}, \mp\frac{3}{2}\rangle$ as excitons with spin-up or spin-down orientation. Thus, for optically allowed excitons, the spins of the electron and hole point into opposite directions, as indicated in Fig. 7 and Table 3. The degeneracy of the four exciton states of Table 3 is usually split. First, the bright and dark excitons are separated by a small amount $\delta \sim 10$–$100\ \mu$eV because of the *electron–hole exchange interaction* [32, 33, 43, 54]. This is a genuine solid-state effect that accounts for the fact that an electron promoted from the valence to the conduction band no longer experiences the exchange interaction with itself, and we thus have to correct for this missing interaction in the band structure description. It is a repulsive interaction that is only present for electrons and holes with opposite spin orientations. Additionally, in case of an asymmetric dot confinement, the exciton eigenstates can be computed from the phenomenological Hamiltonian [33]

$$H_{\text{exchange}} = \frac{1}{2}\begin{pmatrix} \delta & \delta' & 0 & 0 \\ \delta' & \delta & 0 & 0 \\ 0 & 0 & -\delta & \delta'' \\ 0 & 0 & \delta'' & -\delta \end{pmatrix} \tag{22}$$

where δ' and δ'' are small constants accounting for the asymmetry of the dot confinement. The corresponding eigenstates are linear combinations of the exciton states of Table 3, for example, $(|\frac{1}{2}, -\frac{3}{2}\rangle \pm |-\frac{1}{2}, \frac{3}{2}\rangle)/\sqrt{2}$ for the optically allowed excitons, which are linearly polarized along x and y. If a magnetic field is applied along the growth direction z, the two bright exciton states become energetically further split. Alternatively, if in the Voigt geometry a magnetic field B_x is applied along x, the two bright exciton states become mixed [33]; we will use this fact later in the discussion of possible exciton-based quantum computation schemes.

Table 3. Spin structure of electron–hole states.

Electron–hole state	Optical coupling	Polarization	notation
$\|+\frac{1}{2}, -\frac{3}{2}\rangle$	bright	$\hat{e}_-$	$\begin{smallmatrix}\blacktriangle\triangledown\\ \blacktriangledown\end{smallmatrix}$
$\|-\frac{1}{2}, +\frac{3}{2}\rangle$	bright	$\hat{e}_+$	$\begin{smallmatrix}\triangle\blacktriangledown\\ \blacktriangle\end{smallmatrix}$
$\|+\frac{1}{2}, +\frac{3}{2}\rangle$	dark		$\begin{smallmatrix}\blacktriangle\triangledown\\ \blacktriangle\end{smallmatrix}$
$\|-\frac{1}{2}, -\frac{3}{2}\rangle$	dark		$\begin{smallmatrix}\triangle\blacktriangledown\\ \blacktriangledown\end{smallmatrix}$

The first column reports the z-components of the angular momenta for the electron and hole, the second column indicates whether the exciton can be optically excited (bright) or not (dark), the third column shows the polarization vector of the transition, and the last column gives the short-hand notation used in this paper; the upper triangles indicate whether the electron spin points upward (▲▽) or downward (△▼) and the lower triangles give the corresponding information about the hole spin.

3.2. Biexcitons

In semiconductors of higher dimension, a few other Coulomb-bound electron–hole complexes exist: for instance, the negatively charged exciton, which consists of one hole and two electrons with opposite spin orientations, and the biexciton, which consists of two electron–hole pairs with opposite spin orientations. In both cases, the binding energy is of the order a few meV and is attributed to genuine Coulomb correlations: in the negatively charged exciton, the carriers arrange such that the hole is preferentially located in-between the two electrons and thus effectively screens the repulsive electron–electron interaction; similarly, in the biexciton, the four carriers arrange in a configuration reminiscent of the H_2 molecule, where the two heavier particles—the holes—are located at a fixed distance, and the lighter electrons are delocalized over the whole few-particle complex and are responsible for the binding (see insets of Fig. 5). In the literature, a number of variational wavefunction *ansätze* are known for the biexciton description, for example, that of Kleinman [44]

$$\bar{\phi}_0(\boldsymbol{r}_e, \boldsymbol{r}_h, \boldsymbol{r}_{e'}, \boldsymbol{r}_{h'}) = \exp[-(s_e + s_{e'})/2] \cosh[\beta(t_e - t_{e'})]\, \chi(r_{hh'}) \tag{23}$$

with $s_e = r_{eh} + r_{eh'}$, $t_e = r_{eh} - r_{eh'}$, and r_{ij} the distance between particles i and j. The first two terms on the right-hand side account for the attractive electron-hole interactions and $\chi(r_{hh'})$ for the repulsive hole–hole one (β is a variational parameter). In the inset of Fig. 5, we plot the probability distribution for the electron and hole as computed from Eq. (23); in comparison to the exciton, the biexciton is much more delocalized, and correspondingly the biexciton binding is much weaker [44, 55].

3.2.1. Weak Confinement Regime

Suppose that the biexciton is subject to an additional quantum confinement (e.g., induced by the interface fluctuations depicted in Fig. 1a). If the characteristic confinement length L_z is larger than the excitonic Bohr radius a_x and the extension of the biexciton, one can, in analogy to excitons, introduce a "rigid-biexciton" approximation: here, the biexciton wavefunction (23) of the ideal quantum well is modulated by an envelope function that depends on the center-of-mass coordinate of the biexciton. An effective confinement for the biexciton can be obtained through appropriate convolution of $U_i(\boldsymbol{r}_i)$ (for details, see Appendix A), which is shown in Fig. 5 for a representative interface fluctuation potential. Because of the larger extension of the biexciton wavefunction, the effective potential exhibits a larger degree of confinement and correspondingly the biexciton wavefunction of Fig. 9 is more localized. We will return to this point in the discussion of local optical spectroscopy in Section 4.6.

3.2.2. Strong Confinement Regime

In the strong confinement regime, the "binding" of few-particle complexes is not due to Coulomb correlations but to the quantum confinement, whereas Coulomb interactions only introduce minor energy renormalizations. It thus becomes possible to confine various few-particle electron–hole complexes that are unstable in semiconductors of higher dimension.

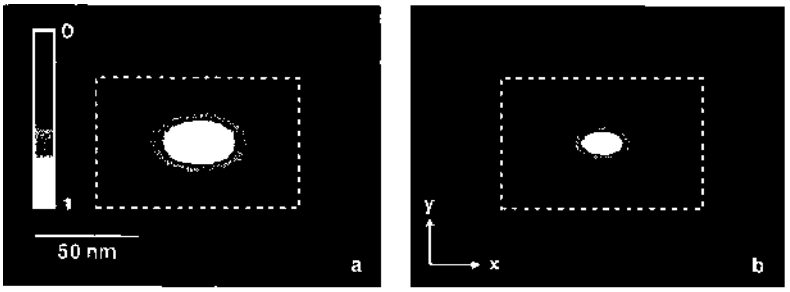


Figure 9. Contour plot of the square modulus of (a) the exciton and (b) the biexciton groundstate. The confinement potential is depicted in Fig. 5 and computational details are presented in Appendix A. Because of the larger spatial extension of the biexciton—see insets (e',h') of Fig. 5—the center-of-mass motion of the biexciton becomes more confined [47].

We start our discussion with the few-particle complex consisting of two electrons and holes. In analogy to higher dimensional semiconductors, we shall refer to this complex as a *biexciton*, keeping in mind that the binding is due to the strong quantum confinement rather than Coulomb correlations. To the lowest order of approximation, the biexciton groundstate $\overline{\Psi}_0$ in the strong confinement regime is given by the product of two excitons (17) with opposite spin orientations (Fig. 10)

$$\overline{\Psi}_0(r_e, r_h, r_{e'}, r_{h'}) \cong \Psi_0(r_e, r_h)\Psi_0(r_{e'}, r_{h'}) \tag{24}$$

$$\overline{E}_0 \cong 2E_0 + \langle \overline{\Psi}_0 | H_{ee'} + H_{hh'} + H_{eh'} + H_{e'h} | \overline{\Psi}_0 \rangle \tag{25}$$

The second term on the right-hand side of Eq. (25) accounts for the repulsive and attractive Coulomb interactions not included in the exciton groundstate energy E_0. If electron and hole single-particle states have the same spatial extension, the repulsive contributions $H_{ee'}$ and $H_{hh'}$ are exactly canceled by the attractive contributions H_{eh} and $H_{e'h}$ and the biexciton energy is just twice the exciton energy (i.e., there is no binding energy for the two neutral excitons). In general, this description is too simplified. If the electrons and holes arrange in a more favorable configuration, such as the H_2 one in the weak confinement regime, the Coulomb energy can be reduced. Within the framework of configuration interactions outlined in Appendix B, such correlation effects imply that the biexciton wavefunction no longer is a single product of two states but acquires additional components from excited states. A rough estimate for the magnitude of such correlation effects is given in first-order perturbation theory by $\langle V \rangle^2/(\Delta\epsilon)$, with $\langle V \rangle$ the average gain of Coulomb energy (typically a few meV) and $\Delta\epsilon$ the splitting of single-particle states (typically a few tens of meV). In general, it turns out to be convenient to parameterize the biexciton energy through

$$\overline{E}_0 = 2E_0 - \Delta \tag{26}$$

where Δ is the biexciton binding energy. Its value is usually positive and somewhat smaller than the corresponding quantum-well value but can sometimes even acquire negative values ("biexciton antibinding" [56]). We shall find that the Coulomb renormalization Δ has the important consequence that the biexciton transition is at a different frequency than the exciton one, which will allow us to distinguish the two states in incoherent and coherent spectroscopy.

3.3. Other Few-Particle Complexes

Besides the exciton and biexciton states, quantum dots in the strong confinement regime can host a number of other few-particle complexes. Depending on whether they are neutral (i.e., consist of an equal number of electrons and holes) or charged, we shall refer to them as

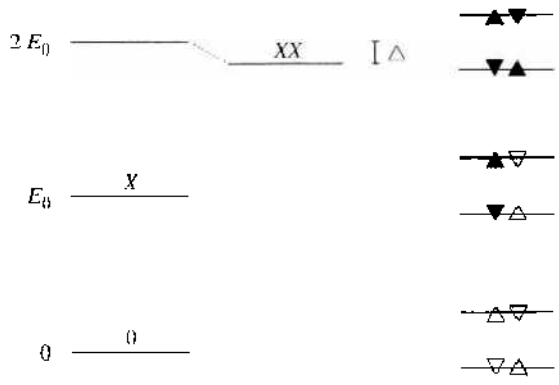


Figure 10. Schematic sketch of the biexciton groundstate, which consists of two electron–hole pairs with opposite spin orientations. Because of noncompensating Coulomb interactions and/or Coulomb correlation effects, the energy of the biexciton is modified by a small amount Δ. In the figure, X and XX refer to the exciton and biexciton groundstate, respectively.

multiexcitons or *multicharged excitons*. Because electron-hole pairs are neutral objects, quantum dots can be populated by a relatively large number of pairs ranging from six [57] to several tens [58, 59] depending on the dot confinement. In experiments, such multiexciton population is usually achieved as follows: a pump pulse creates electron-hole pairs in continuum states (e.g., wetting layer) in the vicinity of the quantum dot, and some of the carriers become captured in the dot; because of the fast subsequent carrier relaxation (Section 5), the few-particle system relaxes to its state of lowest energy, and finally the electron and hole recombine by emitting a photon. Thus, in a steady-state experiment, information about the few-particle carrier states can be obtained by varying the pump intensity and monitoring the luminescence from the quantum dot [57, 60–62]. Results of such multiexciton spectroscopy experiments will briefly be presented in the next section. Experimentally, it is also possible to create electron–hole complexes with an unequal number of electrons and holes. Figure 11 shows how this can be done [63–65]: a quantum dot is placed within a *n-i* field-effect structure; when an external gate voltage is applied, the energy of the electron groundstate drops below the Fermi energy of the *n*-type reservoir and an electron tunnels from the reservoir to the dot, where further charging is prohibited because of the *Coulomb blockade* (i.e., because of the strong Coulomb repulsion between electrons in the dot); when the dot is optically excited (e.g. by the same mechanism of off-resonant excitation and carrier capture described above), one can create charged excitons. A further increase of the gate voltage allows to promote more electrons from the reservoir to the dot, and to hereby create multicharged excitons with up to two surplus electrons.

In [66], a quantum dot was placed in a *n-i-p* structure, which allowed to create in the same sample either negatively (more electrons than holes) or positively (more holes than electrons) charged excitons by varying the applied gate voltage. A different approach was pursued by Hartmann et al. [67], where charging was achieved by unintentional background doping and the mechanism of photo depletion, which allowed to charge quantum dots with up to five surplus electrons. Luminescence spectra of such multicharged excitons will be presented in Section 4.4.

3.4. Coupled Dots

We conclude this section with a brief discussion of coupled quantum dots. In analogy to *artificial atoms*, we may refer to coupled dots as *artificial molecules*. Coupling is an inherent feature of any high-density quantum dot ensemble, as, for example, needed for most

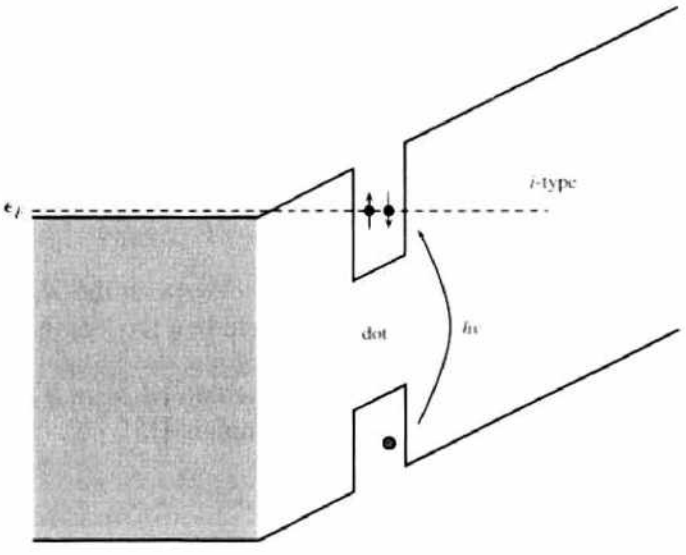


Figure 11. This figure schematically sketches the creation of charged or multicharged excitons. A quantum dot is placed inside a field-effect structure. By applying an external gate voltage, it becomes possible to transfer electrons one by one from the nearby *n*-type reservoir to the dot. When the sample is optically excited, an additional electron–hole pair is created (i.e., a charged or multi-charged exciton is formed).

optoelectronic applications [2]. On the other hand, it is essential to the design of (quantum) information devices, for example quantum dot cellular automata [68] or quantum-dot implementations of quantum computation (Section 7).

Artificial molecules formed by two or more coupled dots are extremely interesting also from the fundamental point of view, as the interdot coupling can be tuned far out of the regimes accessible in natural molecules, and the relative importance of single-particle tunneling and Coulomb interactions can be varied in a controlled way. The interacting few-electron states in a double dot were studied theoretically [69–71] and experimentally by tunneling and capacitance experiments [72–79], and correlations were found to induce coherence effects and novel ground-state phases depending on the interdot coupling regime. For self-organized dots, stacking was demonstrated [80], and the exciton splitting in a single artificial molecule was observed and explained in terms of single-particle level filling of delocalized bonding and antibonding electron and hole states [81–83]. When a few photoexcited particles are present, Coulomb coupling between electrons and holes adds to the homopolar electron–electron and hole–hole couplings. In addition, single-particle tunneling and kinetic energies are affected by the different spatial extension of electrons and holes, and the correlated ground and excited states are governed by the competition of these effects [69, 84–86]. A particularly simple parameterization of single-exciton and biexciton states in coupled dots is given by the Hubbard-type Hamiltonian [87]

$$H = E_0 \sum_{\sigma} (\hat{n}_{L\sigma} + \hat{n}_{R\sigma}) - t \sum_{\sigma} (b^{\dagger}_{L\sigma} b_{R\sigma} + b^{\dagger}_{R\sigma} b_{L\sigma}) - \Delta \sum_{\ell=L,R} \hat{n}_{\ell\uparrow} \hat{n}_{\ell\downarrow} \tag{27}$$

with $b^{\dagger}_{\ell\sigma}$ the creation operator for excitons with spin orientation $\sigma = \pm$ in the right or left dot, $\hat{n}_{\ell\sigma} = b^{\dagger}_{\ell\sigma} b_{\ell\sigma}$ the exciton number operator, t the tunneling matrix element, and Δ the biexciton binding. Indeed, the Hamiltonian (27) accounts properly for the formation of bonding and antibonding exciton states and the fact that in a biexciton state the two electron–hole pairs preferentially stay together to benefit from the biexciton binding Δ [69, 84]. We will return to coupled dots in the discussion of quantum control (Section 6) and quantum computation (Section 7).

4. OPTICAL SPECTROSCOPY

In the last section, we discussed the properties of electron–hole states in semiconductor quantum dots. We shall now show how these states couple to the light and can be probed optically. Our starting point is given by Eq. (9), which describes the propagation of one electron subject to the additional quantum confinement U. Quite generally, the light field is described by the vector potential $\mathcal{A}$, and the light-matter coupling is obtained by replacing the momentum operator $\boldsymbol{p} = -i\nabla$ with $\boldsymbol{p} - (q/c)\mathcal{A}$ [13, 15], where $q = -e$ is the charge of the electron and c the speed of light. The light-matter coupling then follows from

$$\frac{(\boldsymbol{p} + \frac{e}{c}\mathcal{A})^2}{2m} + U(\boldsymbol{r}) = H_0 + \frac{e}{mc}\mathcal{A}\boldsymbol{p} + \frac{e^2}{2mc^2}\mathcal{A}^2 \tag{28}$$

where we have used the Coulomb gauge $\nabla\mathcal{A} = 0$ [88] to arrive at the $\mathcal{A}\boldsymbol{p}$ term. In many cases of interest, the spatial dependence of $\mathcal{A}$ can be neglected on the length scale of the quantum states, that is, in the *far-field* limit—recall that the length scales of light and matter are given by micrometers and nanometers, respectively—and we can perform a gauge transformation to replace the $\mathcal{A}\boldsymbol{p}$ term by the well-known dipole coupling [15]

$$H_{\mathrm{ap}} = \frac{e}{mc}\mathcal{A}\boldsymbol{p} \cong e\boldsymbol{r}\mathcal{E} \tag{29}$$

The relation between the vector potential and the electric field is given by $\mathcal{E} = (1/c)\partial\mathcal{A}/\partial t$. In this far-field limit, we can also safely neglect the $\mathcal{A}^2$ term. This is because the matrix elements $\langle 0| \pm \frac{1}{2}, \mp\frac{3}{2}\rangle$ between the atomic states introduced in Section 3.1.3 vanish owing to the orthogonality of conduction and valence band states. Equation (29) is suited for both

classical and quantum light fields. In the first case, $\mathcal{E}$ is treated as a c-number, in the latter case, the electric field of photons reads [15, 22, 89, 90]

$$\mathcal{E} \cong i \sum_{k\lambda} \left(\frac{2\pi\omega_k}{\kappa_s} \right)^{1/2} (\hat{e}_{k\lambda} a_{k\lambda} - \hat{e}^*_{k\lambda} a^\dagger_{k\lambda}) \tag{30}$$

Here, $\boldsymbol{k}$ and λ are the photon wavevector and polarization, respectively, $\omega_k = ck/n_s$ is the light frequency, $n_s = \sqrt{\kappa_s}$ the semiconductor refractive index, $\hat{e}_{k\lambda}$ the photon polarization vector, and $a_{k\lambda}$ denotes the usual bosonic field operator. The free photon field is described by the Hamiltonian $H_0^\gamma = \sum_{k\lambda} \omega_k \, a^\dagger_{k\lambda} a_{k\lambda}$.

4.1. Optical Dipole Moments

Optical selection rules were already introduced in Section 3.1.3, where we showed that light with appropriate polarization λ (i.e., for light propagation along z either circular polarization for symmetric dots or linear polarization for asymmetric ones) can induce electron-hole transitions. We shall now show how things are modified when additionally the envelope part of the carrier wavefunctions is considered. In second quantization (see Appendix B), the light-matter coupling of Eq. (29) reads

$$H_{\rm op} \cong \sum_\lambda \int dr \, (\mu_0 \hat{e}_\lambda \, \psi^h_{\bar{\lambda}}(r) \psi^e_\lambda(\boldsymbol{r}) + \text{h.c.}) \, \mathcal{E} \tag{31}$$

where $\bar{\lambda}$ is the polarization mode orthogonal to λ. Because of the envelope-function approximation, the dipole operator $e\boldsymbol{r}$ has been completely absorbed in the bulk moment μ_0 [12, 13]. The first term in parentheses of Eq. (31) accounts for the destruction of an electron–hole pair, and the second one for its creation. Similarly, in an all-electron picture, the two terms can be described as the transfer of an electron from the conduction to the valence band or *vice versa*. This single-particle nature of optical excitations translates to the requirement that electron and hole are destroyed or created at the *same* position $\boldsymbol{r}$. Equation (31) usually comes together with the so-called *rotating-wave approximation* [12, 15]. Consider the light-matter coupling (31) in the interaction picture according to the Hamiltonian of the unperturbed system: the first term, which accounts for the annihilation of an electron–hole pair, then approximately oscillates with $e^{-i\omega_0 t}$ and the second term with $e^{i\omega_0 t}$, where ω_0 is a frequency of the order of the semiconductor band gap. Importantly, ω_0 sets the largest energy scale (eV) of the problem, whereas all exciton or few-particle level splittings are substantially smaller. If we accordingly separate $\mathcal{E}$ into terms oscillating with approximately $e^{\pm i\omega_0 t}$, we encounter in the light-matter coupling of Eq. (31) two possible combinations of exponentials: first, those with $e^{\pm i(\omega_0 - \omega_0)t}$, which have a slow time dependence and have to be retained; second, those with $e^{\pm i(\omega_0 + \omega_0)t}$, which oscillate with twice the frequency of the band gap. In the spirit of the random-phase approximation, the latter off-resonant terms do not induce transitions and can thus be neglected. Then, the light-matter coupling of Eq. (31) becomes

$$H_{\rm op} \cong \frac{1}{2} (\mathcal{P} \mathcal{E}^{(-)} + \mathcal{P}^\dagger \mathcal{E}^{(+)}) \tag{32}$$

where $\mathcal{E}^{(\pm)} \propto e^{\mp i\omega_0 t}$ solely evolves with positive or negative frequency components, and the complex conjugate of $\mathcal{E}^{(-)}$ is given by $\mathcal{E}^{(+)}$ [22]. In Eq. (32), $\mathcal{P} = \sum_\lambda \mu_0 \hat{e}_\lambda \int d\boldsymbol{r} \, \psi^h_{\bar{\lambda}}(\boldsymbol{r}) \psi^e_\lambda(\boldsymbol{r})$ is the usual interband polarization operator [12, 91].

Let us finally briefly discuss the optical dipole elements for excitonic and biexcitonic transitions. Within the framework of second quantization, the exciton and biexciton states $|x_\lambda\rangle$ and $|b\rangle$ can be expressed as

$$|x_\lambda\rangle = \int d\tau \, \Psi_x(\boldsymbol{r}_e, \boldsymbol{r}_h) \, \psi^{e\dagger}_\lambda(\boldsymbol{r}_e) \psi^{h\dagger}_{\bar{\lambda}}(\boldsymbol{r}_h) \, |0\rangle \tag{33}$$

$$|b\rangle = \int d\bar{\tau} \, \bar{\Psi}_b(\boldsymbol{r}_e, \boldsymbol{r}_h, \boldsymbol{r}'_e, \boldsymbol{r}'_h) \, \psi^{e\dagger}_\lambda(\boldsymbol{r}_e) \psi^{h\dagger}_{\bar{\lambda}}(\boldsymbol{r}_h) \psi^{e\dagger}_{\bar{\lambda}}(\boldsymbol{r}'_e) \psi^{h\dagger}_\lambda(\boldsymbol{r}'_h) \, |0\rangle \tag{34}$$

with $d\tau$ and $d\bar{\tau}$ denoting the phase space for excitons and biexcitons, respectively. In Eq. (33), the exciton state consists of one electron and hole with opposite spin orientations, and in Eq. (34) the biexciton of two electron–hole pairs with opposite spin orientations. From the light-matter coupling (31), we then find for the optical dipole elements

$$\langle 0|\mathcal{P}|x_\lambda\rangle = M_{0\lambda}\hat{\boldsymbol{e}}_\lambda = \mu_0\hat{\boldsymbol{e}}_\lambda \int d\boldsymbol{r}\, \Psi_\lambda(\boldsymbol{r},\boldsymbol{r}) \tag{35}$$

$$\langle x_{\bar{\lambda}}|\mathcal{P}|b\rangle = M_{\lambda b}\hat{\boldsymbol{e}}_\lambda = \mu_0\hat{\boldsymbol{e}}_\lambda \int d\boldsymbol{r}d\boldsymbol{r}_e d\boldsymbol{r}_h\, \Psi_\lambda^*(\boldsymbol{r}_e,\boldsymbol{r}_h)\hat{\Psi}_b(\boldsymbol{r},\boldsymbol{r},\boldsymbol{r}_e,\boldsymbol{r}_h) \tag{36}$$

In Eq. (35) the dipole moment is given by the spatial average of the exciton wavefunction $\Psi_\lambda(\boldsymbol{r},\boldsymbol{r})$ where the electron and hole are at the same position $\boldsymbol{r}$. Similarly, in Eq. (36) the dipole moment is given by the overlap of exciton and biexciton wavefunctions subject to the condition that the electron with spin $\bar{\lambda}$ and the hole with spin λ are at the same position $\boldsymbol{r}$, whereas the other electron and hole remain at the same position. In Appendix A.3, we show that in the weak confinement regime the oscillator strength for optical transitions scales with the confinement length L according to $|M_{0\lambda}|^2 \propto L^2$ (i.e., it is proportional to the confinement area). For that reason, excitons in the weak-confinement regime couple much stronger to the light than those in the strong confinement regime, which makes them ideal candidates for various kinds of optical coherence experiments [8, 26, 28, 90, 92–95].

4.2. Fluctuation-Dissipation Theorem

We next discuss how to compute optical spectra in linear response. As a preliminary task, we consider the general situation where a generic quantum system is coupled to an external perturbation $X(t)$ (e.g., an exciting laser light) via the system operator A through $A\,X(t)$ [96]. Let $\langle B\rangle$ be the expectation value of the operator B in the perturbed system and $\langle B\rangle_0$ that in the unperturbed one. In linear-response theory, the change $\langle \Delta B\rangle = \langle B\rangle - \langle B\rangle_0$ is assumed to be *linear* in the perturbation $X(t)$—an approximation valid under quite broad conditions provided that the external perturbation is sufficiently weak. We can then derive within lowest-order time-dependent perturbation theory the famous *fluctuation-dissipation theorem* [96]

$$\langle \Delta B(0)\rangle = i\int_{-\infty}^{0} dt' \langle [A(t'), B(0)]\rangle_0\, X(t') \tag{37}$$

where operators A and B are given in the interaction picture according to the unperturbed system Hamiltonian H_0. In Eq. (37), we have assumed that the external perturbation has been turned on at sufficiently early times such that the system has reached equilibrium. The important feature of Eq. (37) is that it relates the expectation value of B in the *perturbed* system to the correlation $[A, B]$—or equivalently to the fluctuation $[\Delta A, \Delta B]$ because commutators with c-numbers always vanish—of the *unperturbed system*. Usually, the expression on the right-hand side of Eq. (37) is much easier to compute than that on the left-hand side. We will next show how the fluctuation-dissipation theorem (37) can be used for the calculation of linear optical absorption and luminescence.

4.3. Optical Absorption

Absorption describes the process where energy is transferred from the light field to the quantum dot (i.e., light becomes absorbed). Absorption is proportional to the loss of energy of the light field, or equivalently to the gain of energy of the system

$$\alpha(\omega) \propto \frac{d}{dt}\langle H_0 + H_{\rm op}\rangle = \left\langle \frac{\partial H_{\rm op}}{\partial t}\right\rangle \tag{38}$$

Consider a monofrequent excitation $\mathscr{E}_0\hat{\boldsymbol{e}}_\lambda \cos\omega t$, where $\mathscr{E}_0$ is the amplitude of the light field. Inserting this expression into Eq. (38) gives after some straightforward calculation $\alpha(\omega) \propto -\omega\mathscr{E}_0\,\Im m(e^{i\omega t}\langle \hat{\boldsymbol{e}}_\lambda^*\mathcal{P}\rangle)$. From the fluctuation-dissipation theorem we then find

$$\alpha_\lambda(\omega) \propto \Im m\left(i\int_{-\infty}^{0} dt' \langle [\hat{\boldsymbol{e}}_\lambda^*\mathcal{P}(0), \hat{\boldsymbol{e}}_\lambda\mathcal{P}^\dagger(t')]\rangle_0 e^{-i\omega t'}\right) \tag{39}$$

(i.e., optical absorption is proportional to the spectrum of interband polarization fluctuations). We emphasize that this is a very general and important result that holds true for systems at finite temperatures, and is used for *ab initio* type calculations of optically excited semiconductors [97]. We next show how to evaluate Eq. (39). Suppose that the quantum dot is initially in its groundstate. Then only the term $\langle 0|\mathcal{P}(0)\,\mathcal{P}^\dagger(t')|0\rangle$ contributes in Eq. (39) because no electron–hole pair can be destroyed in the vacuum (i.e., $\mathcal{P}|0\rangle = 0$). Through $\mathcal{P}^\dagger(t')|0\rangle$, an electron–hole pair is created in the quantum dot, which propagates in presence of the quantum confinement and the Coulomb attraction between the electron and hole. Thus, the propagation of the interband polarization can be computed by use of the exciton eigenstates $|x\rangle$ through $\langle x|\mathcal{P}^\dagger(t')|0\rangle = e^{iE_x t}\,\langle x|\mathcal{P}^\dagger|0\rangle$. Inserting the complete set of exciton eigenstates in Eq. (39) gives

$$\alpha(\omega) \propto \Im m\left(i \sum_{x\lambda} \int_{-\infty}^{0} dt' |M_{0x_\lambda}|^2 \, e^{-i(\omega - E_{x_\lambda})t'} \right) \tag{40}$$

To evaluate the integral in Eq. (40), we have to assume that the exciton energy has a small imaginary part $E_x - i\gamma$ associated to the finite exciton lifetime because of environment couplings (Section 5). Then, $\int_{-\infty}^{0} dt' e^{-i(\omega - E_x + i\gamma)} = i/(\omega - E_x + i\gamma)$ and we obtain for the optical absorption the final result

$$\alpha(\omega) \propto \sum_{x\lambda} |M_{0x_\lambda}|^2 \, \delta_\gamma(\omega - E_{x_\lambda}) \tag{41}$$

Here, $\delta_\gamma(\omega) = \gamma/(\omega^2 + \gamma^2)$ is a Lorentzian that in the limit $\gamma \to 0$ gives Dirac's delta function. According to Eq. (41), the absorption spectrum of a single quantum dot is given by a comb of delta-like peaks at the energies of the exciton states, whose intensities—sometimes referred to as the *oscillator strengths*—are given by the square modulus of the dipole moments (35).

4.3.1. Weak Confinement

In the weak confinement regime the absorption spectrum is given by

$$\alpha(\omega) \propto \sum_x \left| \int d\boldsymbol{R}\, \Phi_x(\boldsymbol{R}) \right|^2 \delta_\gamma(\omega - \mathscr{E}_x) \tag{42}$$

where $\Phi_x(\boldsymbol{R})$ is the center-of-mass wavefunction introduced in Section 3.1. Let us consider the somewhat simplified example of a rectangular confinement with infinite barriers whose solutions are $\Phi(X, Y) = 2/(L_1 L_2)^{1/2} \sin(n_1 \pi X/L_1) \sin(n_2 \pi Y/L_2)$. Here, X and Y are the center-of-mass coordinates along x and y, L_1 and L_2 are the confinement lengths in x- and y-direction, and n_1 and n_2 the corresponding quantum numbers. The energy associated to this wavefunction is $\mathscr{E} = \pi^2/(2M)\,[(n_1/L_1)^2 + (n_2/L_2)^2]$. Inserting these expressions into Eq. (42) shows that the oscillator strength is zero when n_1 or n_2 is an even number and proportional to $L_1 L_2/(n_1 n_2)$ otherwise. Figure 12 shows absorption spectra computed within this framework for (a) an inhomogeneously broadened ensemble of quantum dots and (b) a single dot. The first situation corresponds to typical optical experiments performed on ensembles of quantum dots. Single dots can be measured by different types of local spectroscopy such as submicrometer apertures [4, 31, 48, 62, 98], solid immersion microscopy [50, 99], or scanning near-field microscopy [8, 94, 100]. Note that such single-dot spectroscopy is indispensable for the observation of the atomic-like optical density of states depicted in Fig. 12b, which is completely hidden in presence of the inhomogeneous broadening of Fig. 12a.

4.3.2. Strong Confinement

In the strong confinement regime, the optical response is governed by the single-particle properties. However, Coulomb interactions are responsible for renormalization effects that leave a clear fingerprint in the optical response. In the context of quantum-dot based quantum computation schemes (Section 7), it is precisely this fingerprint that allows the optical

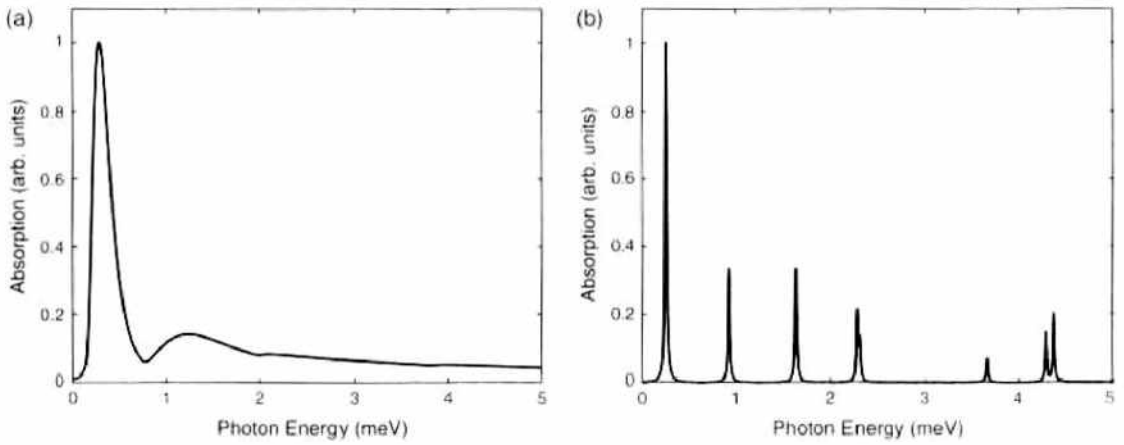


Figure 12. Absorption spectra for quantum dots in the weak-confinement regime and for (a) an inhomogenously broadened ensemble of quantum dots and (b) a single dot of dimension 100×70 nm^2. For the inhomogenous broadening, we assume a Gaussian distribution of the confinement lengths L_1 and L_2, which is centered at 70 nm and has a full-width of half-maximum of 60 nm, and for the homogeneous lifetime broadening $\gamma = 10$ μeV. We use material parameters representative for GaAs. Photon energy zero is given by the exciton energy of the two-dimensional quantum well.

manipulation of individual few-particle states. Similar to the absorption (42) in the weak-confinement regime, the linear optical absorption in the strong-confinement regime reads

$$\alpha(\omega) \propto \sum_x \left| \int dr\, \Psi_x(r, r) \right|^2 \delta(\omega - E_x) \tag{43}$$

Because the electron and hole are confined within a small space region, the oscillator strength is much smaller as compared to the weak confinement regime. The approximately product-type structure (17) of the exciton wavefunction and the similar shape of electron and hole wavefunctions gives rise to optical selection rules where only transitions between electron and hole states with corresponding quantum numbers (e.g. s–s or p–p) are allowed. Indeed, such behavior is observed in Fig. 13 showing absorption spectra representative for $In_xGa_{1-x}As$ dots: the three major peaks can be associated to transitions between the respective electron and hole ground states and the first and second excited states [3, 102, 103]. In our calculations we assume parabolic confinement potentials for electrons and holes, with a 2:1 ratio between the electron and hole single-particle splittings [101] and compute the spectra within a full configuration-interaction approach (Appendix B) for the respective six electron and hole single-particle states of lowest energy. The single-dot spectrum of Fig. 13b shows that Coulomb interactions result in a shift of oscillator strength to the transitions of lower energy (in a pure single-particle framework the ratio would be simply 1:2:3, reflecting the degeneracy of single-particle states), and the appearance of additional peaks [103, 104]. For the dot ensemble, Fig. 13a, we observe that the broadening of the groundstate transition is much narrower than that of the excited ones. This is because the excited states are less confined and are accordingly stronger affected by Coulomb interactions. Note that for the level broadening considered in the figure, the second and third exciton transitions even strongly overlap.

4.4. Luminescence

Luminescence is the process where in a carrier complex one electron-hole pair recombines by emitting a photon. To account for the creation of photons, we have to adopt the framework of second quantization of the light field [15, 22, 89] and use expression (30) for the electric field of photons. Then,

$$H_{cp} \cong i \sum_{k\lambda} \left(\frac{2\pi\omega_k}{\kappa_s} \right)^{1/2} (\hat{e}^*_{k\lambda}\, a^\dagger_{k\lambda}\, \mathcal{P} - \hat{e}_{k\lambda}\, a_{k\lambda}\, \mathcal{P}^\dagger) \tag{44}$$

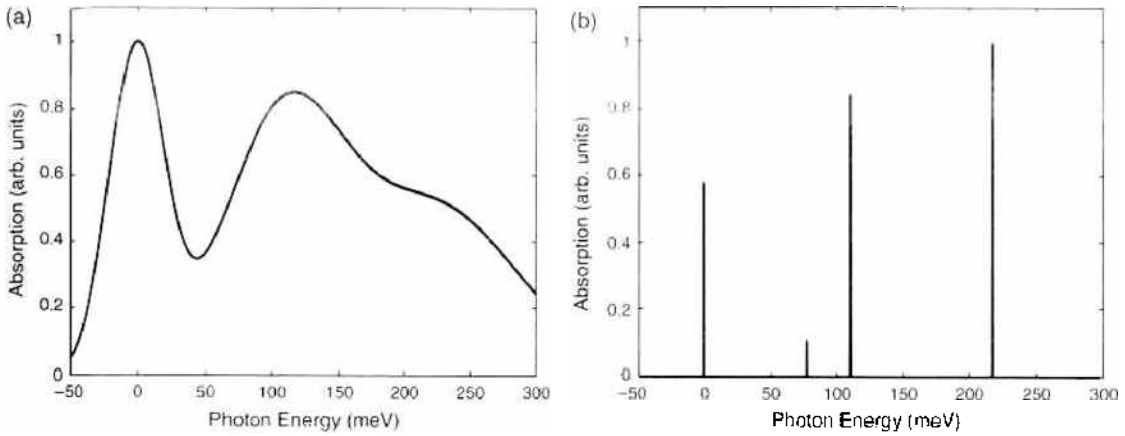


Figure 13. Absorption spectra for quantum dots in the strong confinement regime and for (a) an inhomogeneously broadened ensemble of quantum dots and (b) a single dot. We use prototypical material and dot parameters for $In_xGa_{1-x}As/GaAs$ dots [65, 101]. We assume a 2:1 ratio of single-particle splittings between electrons and holes and a 20% stronger hole confinement associated to the heavier hole mass and possible piezoelectric fields. The absorption spectra are computed within a full configuration-interaction approach for the respective six electron and hole single-particle states of lowest energy.

is the Hamilton operator that describes the photon-matter coupling within the envelope-function and rotating-wave approximations. The first term on the right-hand side describes the destruction of an electron-hole pair through photon emission, and the second one the reversed process. We shall now show how to compute from Eq. (44) the luminescence spectrum $L(\omega)$. It is proportional to the increase in the number of photons $\langle a^\dagger_{k\lambda} a_{k\lambda}\rangle$ emitted at a given energy ω. With this approximation we obtain

$$L(\omega) \propto \frac{d}{dt}\left(\sum_{k\lambda}\langle a^\dagger_{k\lambda} a_{k\lambda}\rangle\, \delta(\omega - \omega_{k\lambda})\right) \tag{45}$$

We next use Heisenberg's equation of motion $\dot{\mathcal{O}} = -i[\mathcal{O}, H]$, with $\mathcal{O}$ an arbitrary time-independent operator, and $\langle [a^\dagger_{k\lambda} a_{k\lambda}, H]\rangle = 2i\,\Im m\langle a^\dagger_{k\lambda}[a_{k\lambda}, H_{\rm up}]\rangle$. By computing the commutator with $H_{\rm up}$, Eq. (44), we obtain

$$L(\omega) \propto \sum_{k\lambda}\left(\frac{2\pi\omega_k}{\kappa_s}\right)^{1/2} \Im m\langle a^\dagger_{k\lambda}\, \hat{e}^*_{k\lambda}\mathcal{P}\rangle\, \delta(\omega - \omega_{k\lambda}) \tag{46}$$

The expression $\langle a^\dagger \mathcal{P}\rangle$ is known as the *photon-assisted density matrix* [27, 105]. It describes the correlations between the photon and the electron-hole excitations in the quantum dot. Again, we can use the fluctuation-dissipation theorem (37) to compute Eq. (46) and to provide a relation between the photon-assisted density matrix and the correlation function $\langle [a^\dagger_{k\lambda}(0)\mathcal{P}(0), a_{k'\lambda'}(t)\mathcal{P}^\dagger(t)]\rangle$. The calculation can be considerably simplified if we make the reasonable assumption that before photon emission, no other photons are present. Then, only $\langle a_{k'\lambda'} a^\dagger_{k\lambda}\rangle = \delta_{kk'}\delta_{\lambda\lambda'}$ does not vanish. It is diagonal in k and λ, because the only photon that can be destroyed by the annihilation operator a is the one created by $a^\dagger$. Thus,

$$L(\omega) \propto \pi\sum_{k\lambda}\frac{2\pi\omega_k\mu_0^2}{\kappa_s}\,\Im m\left(i\int_{-\infty}^{0} dt'\,\langle \hat{e}_{k\lambda}\mathcal{P}^\dagger(t')\hat{e}^*_{k\lambda}\mathcal{P}(0)\rangle e^{i\omega t}\right)\delta(\omega - \omega_{k\lambda}) \tag{47}$$

Suppose that the system is initially in the eigenstate $|i\rangle$ of the unperturbed system which has energy E_i. We next insert a complete set of eigenstates $\sum_f |f\rangle\langle f|$ in the above equation (note that because the interband polarization operator $\mathcal{P}$ can only remove one electron-hole pair the states f and i differ by one electron and hole), and assume that only photons propagating along z with polarization λ are detected. Then,

$$L(\omega) \propto \sum_f |\langle f|\hat{e}^*_\lambda \mathcal{P}|i\rangle|^2\, \delta_\gamma(E_f + \omega - E_i) \tag{48}$$

gives the expression for calculating luminescence spectra. Here, the photon energies at the peak positions equal the energy differences of initial and final states, and the oscillator strengths are given by the overlap between the two wavefunctions subject to the condition that one electron-hole pair is removed through the interband polarization operator $\mathcal{P}$.

4.5. Multiexciton and Multicharged Excitons

Let us first discuss the luminescence of multiexcitons (i.e., carrier complexes with an equal number of electron–hole pairs). To observe Coulomb renormalization effects in the optical spectra (such as, e.g., the biexciton shift Δ), it is compulsory to measure single dots. For dot ensembles, all line splittings would be completely hidden by the inhomogeneous broadening. The challenge to detect luminescence from single quantum dots (the density of typical self-assembled dots in the strong confinement regime is of the order of 5×10^{10} cm^{-2} [2]) is accomplished by means of various experimental techniques, such as shadow masks or mesas [106]. Such *single-dot spectroscopy* [57, 58, 60, 61, 107–110] has revealed a surprisingly rich fine-structure in the optical spectra, with the main characteristic that whenever additional carriers are added to the dot the optical spectra change because of the resulting additional Coulomb interactions. This has the consequence that each quantum-dot spectrum uniquely reflects its electron–hole configuration. In the following, we adopt the model of a quantum dot with cylinder symmetry (Section 3.1), and compute the luminescence spectra for an increasing number of electron–hole pairs within a full configuration-interaction approach. Results are shown in Fig. 14. For the single-exciton decay, the luminescence spectra exhibit a single peak at the exciton energy E_0 whose intensity is given by $|M_{0x}|^2$. In the biexciton

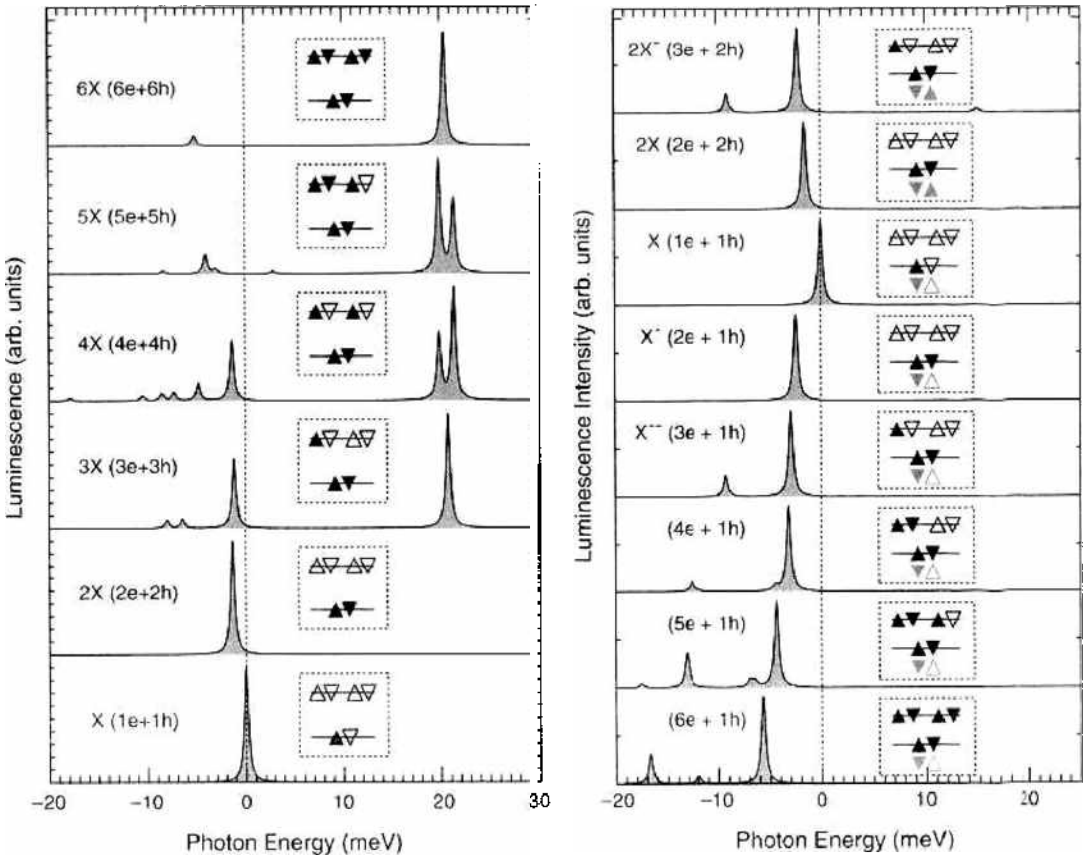


Figure 14. Luminescence spectra for multiexcitons (left panel) and multicharged excitons (right panel) as computed from a full configuration-interaction approach with a basis of approximately 10,000 states [102]. We use material parameters for GaAs and single-particle level-splittings of 20 meV for electrons and 3.5 meV for holes. The insets report the electron–hole configuration of the groundstate before photon emission (for multiexcitons, we use the same configuration for electrons and holes). For clarity we use a relatively large peak broadening.

decay, one electron–hole pair in the Coulomb-renormalized carrier complex recombines by emitting a photon, whose energy is reduced by Δ because of Coulomb correlation effects.

Things become more complicated when the number of electron–hole pairs is further increased. Here, one pair has to be placed in the excited p-shell, which opens up the possibility for different decay channels. Because of wavefunction symmetry, only electrons and holes in corresponding shells can recombine and emit a photon [3, 51]. For recombination in the p shell, the energy $\omega \sim E_0 + \Delta\epsilon_e + \Delta\epsilon_h$ of the emitted photon is blue-shifted by the energy splitting $\Delta\epsilon_e + \Delta\epsilon_h$ of single-particle states. On the other hand, recombination in the s-shell brings the system to an excited biexciton state with one electron–hole pair in the s and one in the p shell. Such excited states are subject to pronounced Coulomb renormalizations, which can be directly monitored in the luminescence spectra of Fig. 14. The two main features of luminescence from the different single-particle shells and the unambiguous spectroscopic fingerprint for each few-particle state because of Coulomb correlations prevail for the other multiexciton complexes. Similar conclusions also apply for multicharged excitons [63–67, 111–113]. Because of the strong single-particle character, the *aufbau* principle for negatively and positively charged excitons is dominated by successive filling of single-particle states, whereas Coulomb interactions only give rise to minor energy renormalizations. The only marked difference in comparison to multiexcitons is the additional Coulomb repulsion due to the imbalance of electrons and holes, which manifests itself in the carrier-capture characteristics [67, 114] and in the instability of highly charged carrier complexes [64, 65]. Typical multicharged exciton spectra are shown in the right panel of Fig. 14. For negatively charged dots, the main peaks red-shift with increasing doping because of exchange and correlation effects, and each few-particle state has its own specific fingerprint in the optical response. When the dot is positively charged, the emission-peaks preferentially shift to the blue [66]. This unique assignment of peaks or peak multiplets to given few-particle configurations allows in optical experiments to determine unambiguously the configuration of carrier complexes.

4.6. Near-Field Scanning Microscopy

Up to now, we have been concerned with optical excitation and detection in the far-field regime, where the spatial dependence of the electric field $\mathcal{E}(\boldsymbol{r})$ can be safely neglected. However, the diffraction limit $\lambda/2$ of light can be significantly overcome through near-field optical microscopy [115, 116]. This is a technique based on scanning tunneling microscopy, where an optical fiber is used as the tip and light is quenched through it. Most importantly, close to the tip the electric field contribution is completely different from that in the far-field [88, 115, 117].

For the quantum dots of our present concern, the carrier wavefunctions are always much stronger confined in the z-direction than in the lateral ones, which allows us to replace the generally quite complicated electromagnetic field distribution in the vicinity of the tip [115, 118, 119] by a more simple shape (e.g., a Gaussian with a given full width of half maximum σ_r). Up to now, most of the local-spectroscopy experiments were performed with spatial resolutions σ_r larger than the extension of the semiconductor nanostructures themselves [31]. This allowed to locate them but not not spatially resolve their electron–hole wavefunctions. Only very recently, Matsuda et al. [8, 120] succeeded in a beautiful experiment to map spatially the exciton and biexciton wavefunctions of a quantum dot in the weak confinement regime. In the following, we briefly discuss within the framework developed in Refs. [103, 121, 122] the main features of such local-spectroscopy experiments, and point to the difficulties inherent to their theoretical interpretation. Our starting point is given by the light-matter coupling (31). We assume, however, that the electric field $\mathcal{E}$ has an explicit space dependence through

$$\mathcal{E}^{(+)}(\boldsymbol{r}) = \mathcal{E}_0 e^{-i\omega t}\, \boldsymbol{\xi}(\boldsymbol{R} - \boldsymbol{r}) \tag{49}$$

Here, $\mathcal{E}_0$ is the amplitude of the exciting laser with frequency ω, and $\boldsymbol{\xi}(\boldsymbol{R} - \boldsymbol{r})$ is the profile of the electric field in the vicinity of the fiber tip. The tip is assumed to be located at position $\boldsymbol{R}$. Scanning the tip over the sample thus allows to measure the local absorption

(or luminescence [123, 124]) properties at different positions and to acquire information about the electron–hole wavefunction on the nanoscale.

The light-matter coupling for the electric field profile (49) is of the form

$$H_{\mathrm{op}} \cong \mathscr{E}_0 \mu_0 \sum_{\lambda} \int dr \, (\hat{e}_\lambda \psi_\lambda^h(r) \psi_\lambda^e(r) \, e^{i\omega t} \xi^*(R - r) + \mathrm{h.c.}) \tag{50}$$

The remaining calculation to obtain the optical near-field absorption spectra is completely analogous to that of far-field, Section 4.3, with the only difference that the light-matter coupling (50) has to be used instead of Eq. (31). We finally arrive at

$$\alpha(\omega) \propto \sum_{\chi\lambda} \left| \int dr \, \xi^*(R - r) \hat{e}_\lambda \Psi_{\chi_\lambda}(r, r) \right|^2 \delta_\gamma(\omega - E_{\chi_\lambda}) \tag{51}$$

In comparison to Eq. (41), the optical matrix element is given by the convolution of the electromagnetic profile $\xi^*(R - r)\hat{e}_\lambda$ with the exciton wavefunction, rather than the simple spatial average of $\Psi(r, r)$. Two limiting cases can be readily identified in Eq. (51). First, for a far-field excitation ξ_0 that does not depend on r, one recovers precisely the far-field absorption (41). In the opposite limit of infinite resolution, where ξ resembles a δ-function, the oscillator strength is given by the square modulus of the exciton wavefunction $\Psi(R, R)$ at the tip position. Finally, within the intermediate regime of a narrow but finite probe, $\Psi(r, r)$ is averaged over a region that is determined by the spatial extension of the light beam. Therefore, excitonic transitions that are optically forbidden in the far-field may become visible in the near-field. Figure 15 shows near-field spectra as computed from Eq. (51) for a quantum dot in the weak-confinement regime. The confinement for excitons and biexcitons is according to Fig. 5. In the second and third rows, we report our calculated optical near-field spectra for spatial resolutions of 25 and 50 nm. Note that the first (Fig. 15b) and second excited state (not shown) are dipole forbidden, but have large oscillator strengths for both resolutions. As a result of interference effects, the spatial maps at finite spatial resolutions differ somewhat from the wavefunction maps, particularly for the excited states: the apparent

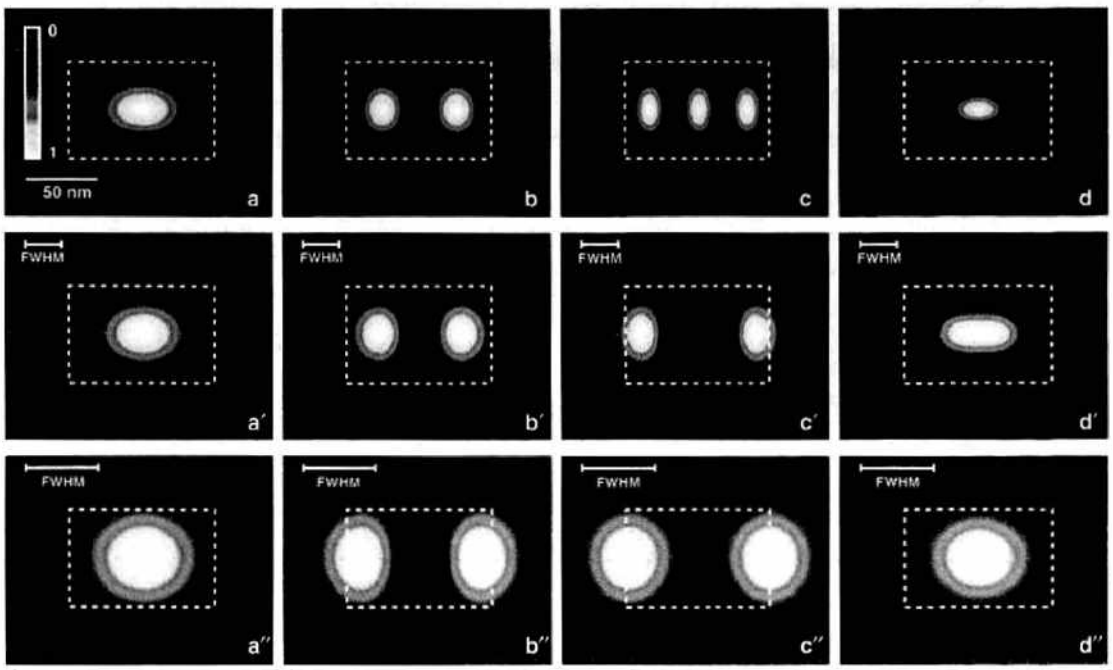


Figure 15. (a–d) Real-space map of the square modulus of the wavefunctions for the exciton (a) groundstate, (b) first and (c) third excited state, and (d) the biexciton groundstate. The dashed lines indicate the boundaries of the assumed interface fluctuation. (a'–d') Near-field spectra for a spatial resolution of 25 nm and (a''–d'') 51 nm, as computed according to Eqs. (51) and (52). The full-width at half maximum σ is indicated in the second and third row.

localization is weaker, and in (c) the central lobe is very weak for both resolutions [47, 103]. For the near-field mapping of the biexciton, we have to be more specific of how the system is excited. We assume that the dot is initially populated by the groundstate exciton and that the near-field tip probes the transition to the biexciton groundstate. This situation approximately corresponds to that of Refs. [8, 120] with nonresonant excitation in the nonlinear power regime. Similar to Eq. (51), the local spectra for biexcitons are given by

$$\alpha(\omega) \propto \left| \int dr dr_e dr_h \, \xi(\boldsymbol{R} - \boldsymbol{r}) \, \hat{e}_\lambda^* \, \Psi_0^*(\boldsymbol{r}_e, \boldsymbol{r}_h) \overline{\Psi}_0(\boldsymbol{r}, \boldsymbol{r}, \boldsymbol{r}_e, \boldsymbol{r}_h) \right|^2 \delta_\gamma(\omega + E_0 - \overline{E}_0) \qquad (52)$$

The corresponding spectra are shown in column (d) of Fig. 15. We observe that for the smaller spatial resolution, the biexciton groundstate depicts a stronger degree of localization than the exciton one, in nice agreement with the recent experiment of Matsuda et al. [8, 120].

4.7. Coherent Optical Spectroscopy

Absorption in a single quantum dot is the absorption of a single photon, which can usually not be measured. Other experimental techniques exist that allow to overcome this problem. In photo-luminescence-excitation spectroscopy, an exciton is created in an excited state; through phonon scattering it relaxes to its state of lowest energy (Section 5.5) and finally recombines by emitting a photon which is detected. Other techniques are more sensitive to the coherence properties and will be discussed below. For instance, in four-wave mixing spectroscopy [27, 125], the system is first excited by a sufficiently strong pump pulse, which creates a polarization. When at a later time a probe pulse arrives at the sample, a polarization grating is formed, and light is emitted into a direction determined by those of the pump and probe pulse [125]. This signal carries direct information about how much of the polarization introduced by the first pulse is left at a later time (i.e., it is a direct measure of the coherence properties). Another technique is coherent nonlinear optical spectroscopy [126], which is often used for quantum dots in the weak confinement regime. It offers a much better signal-to-noise ratio and gives detailed information about the coherence properties of excitons and biexcitons.

5. QUANTUM COHERENCE AND DECOHERENCE

Quantum coherence and decoherence are the two key players in the fields of quantum optics and semiconductor quantum optics. The light-matter coupling (32) is mediated by the interband polarization, which, in a microscopic description, corresponds to a coherent superposition of quantum states. For isolated systems, this provides a unique means for quantum control, where the system wavefunction can be brought to any desired state [127]. Decoherence is the process that spoils such ideal performance. It is due to the fact that any quantum system interacts with its environment (e.g., photons or phonons), and hereby acquires an uncontrollable phase. This introduces a kind of "random noise" and diminishes the quantum-coherence properties. Although from a pure quantum-control or quantum-computation perspective decoherence is often regarded as "the enemy" [128], from a more physics-oriented perspective it is the grain of salt: not only does it provide a means to monitor the state of the system, but it also allows for deep insights to the detailed interplay of quantum systems with their environment. This section is devoted to a more careful analysis of these two key players. We first briefly review the basic concepts of light-induced quantum coherence and its loss due to environment couplings. Based on this discussion, we then show how decoherence can be directly monitored in optical spectroscopy and how it can successfully be exploited for single-photon devices.

5.1. Quantum Coherence

In most cases, we do not have to consider the full spectrum of quantum-dot states. For instance, if the laser frequencies are tuned to the exciton groundstate, it completely suffices to know the energy of the exciton together with the optical matrix element connecting the states. It is physical intuition together with the proper choice of the excitation scenario that

allows to reduce a complicated few-particle problem to a relatively simple few-level scheme. This situation is quite different as compared to the description of carrier dynamics in higher-dimensional semiconductors, where such a clear-cut separation is not possible because of the scattering-type nature of carrier-carrier interactions [12, 27, 125]. It is, however, quite similar to quantum optics [15, 22, 89], which relies on phenomenological level schemes (e.g., Λ- or V-type schemes) with a few effective parameters—a highly successful approach despite the tremendously complicated nature of atomic states. Let us denote the generic few-level scheme with $|i\rangle$, where i labels the different states under consideration. If the artificial atom would be isolated from its environment, we could describe it in terms of the wavefunction

$$|\Psi\rangle = \sum_i C_i |i\rangle \tag{53}$$

with C_i the coefficients subject to the normalization condition $\sum_i |C_i|^2 = 1$. Such wavefunction description is no longer possible for a system in contact with its environment. Because the coefficients C_i acquire random phases through environment couplings and the system can suffer scatterings, we can only state with a certain probability that the system is in a given state. In statistical physics this lack of information is accounted for by the *density operator* [129]

$$\boldsymbol{\rho} = \sum_k p_k |\Psi_k\rangle\langle\Psi_k| = \overline{|\Psi\rangle\langle\Psi|} \tag{54}$$

where the sum of k extends over an ensemble of systems that are with probability p_k in the state $|\Psi_k\rangle$. The last term on the right-hand side provides the usual short-hand notation of this ensemble average. By construction, $\boldsymbol{\rho}$ is a Hermitian operator whose time evolution is given by the Liouville von-Neumann equation $i\dot{\boldsymbol{\rho}} = [H, \boldsymbol{\rho}]$. This can easily be proven by differentiating Eq. (54) with respect to time and using Schrödinger's equation for the Hamiltonian H [12, 15, 129]. If we insert the few-level wavefunctions (53) into (54), we obtain

$$\boldsymbol{\rho} = \sum_{ij} \overline{C_i C_j^*} |i\rangle\langle j| = \sum_{ij} \rho_{ij} |i\rangle\langle j| \tag{55}$$

Here, $\rho_{ij} = \overline{C_i C_j^*}$ is the *density matrix* of the few-level system, which contains the maximum information we possess about the system. The diagonal elements ρ_{ii} account for the probability of finding the system in state i, and the off-diagonal elements ρ_{ij} for the quantum coherence between states i and j. As consequence, $\boldsymbol{\rho}$ fulfills the trace relation $\mathrm{tr}\,\boldsymbol{\rho} = \sum_i \rho_{ii} = 1$, which states that the system has to be in one of its states.

5.1.1. Two-Level System

A particularly simple and illustrative example is given by a generic two-level system. This may correspond to an artificial atom that is either in its groundstate 0 or in the single-exciton state 1. In optical experiments, the population of excited exciton states can strongly be suppressed through appropriate frequency filtering (recall that the principal level splittings are of the order of several tens of meV) and that of biexcitons through appropriate light polarization. Thus, systems with dominant two-level character can indeed be identified in artificial atoms. The density matrix is of dimension two. It has four complex matrix elements corresponding to eight real numbers. Because $\boldsymbol{\rho}$ is a Hermitian matrix, only four of them are independent, which additionally have to fulfill the normalization condition $\mathrm{tr}\,\boldsymbol{\rho} = 1$. A convenient representation of $\boldsymbol{\rho}$ is through the Pauli matrices

$$\sigma_1 = |1\rangle\langle 0| + |0\rangle\langle 1|, \quad \sigma_2 = -i(|1\rangle\langle 0| - |0\rangle\langle 1|), \quad \sigma_3 = |1\rangle\langle 1| - |0\rangle\langle 0| \tag{56}$$

which together with the unit matrix $\mathbb{1} = |1\rangle\langle 1| + |0\rangle\langle 0|$ provides a complete basis within the two-level subspace. The Pauli matrices are Hermitian and have trace zero (Appendix C). Thus, the density matrix can be expressed as

$$\boldsymbol{\rho} = \frac{1}{2}\left(\mathbb{1} + \sum_i u_i \sigma_i\right) = \frac{1}{2}\mathbb{1} + \boldsymbol{u}\,\boldsymbol{\sigma} \tag{57}$$

with $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \sigma_3)$ and the first term guarantees the trace relation $\mathrm{tr}\,\boldsymbol{\rho} = 1$.

The system is thus fully characterized by the three-dimensional *Bloch vector* $\boldsymbol{u} = (u_1, u_2, u_3)$, which was already introduced in Section 2.3: its x- and y-components u_1 and u_2 account for the real and imaginary part of the quantum coherence—or interband polarization—respectively, and the z-component u_3 gives the population inversion between the excited and groundstate. Equation (57) demonstrates that the Bloch-vector picture advocated in Section 2 prevails for density matrices. What happens when the system is excited by an external laser? As discussed in Appendix C, within a rotating frame the system subject to an exciting laser can be described by the Hamiltonian [22]

$$H = \frac{1}{2}(\Delta\,\sigma_3 - \Omega^*\,|0\rangle\langle 1| - \Omega\,|1\rangle\langle 0|) \tag{58}$$

where Δ is the detuning between the laser and the two-level transition, and Ω is the *Rabi frequency*, which determines the strength of the light-matter coupling. Note that Ω describes the envelope part of the laser pulse which is constant for a constant laser and has an only small time dependence for typical pulses. From the Liouville von-Neumann equation $\dot{\boldsymbol{\rho}} = -i[H, \boldsymbol{\rho}]$, we then obtain the equation of motion for the Bloch vector

$$\dot{\boldsymbol{u}} = \boldsymbol{\Omega} \times \boldsymbol{u}, \quad \boldsymbol{\Omega} = (-\Re e(\Omega), \Im m(\Omega), \Delta) \tag{59}$$

We are now in the position to quantitatively describe the buildup of quantum coherence. Suppose that the system is initially in its groundstate 0 where the Bloch vector points into the negative z-direction. When the laser is turned on, the Bloch vector is rotated perpendicularly to $\boldsymbol{\Omega}$. Upon expansion of the solutions of the Bloch equations (59) in powers of the driving field Ω, we observe that to the lowest order the population (i.e., the z-component of the Bloch vector) remains unchanged and only a quantum coherence—described by u_1 and u_2—is created. This is due to the fact that the light couples indirectly (i.e., through the interband polarizations) to the quantum-state populations. When we consider in the solutions of Eq. (59) higher orders of Ω, we find that this induced polarization acts back on the system and modifies the populations. A particularly simple and striking example of such nonlinear light-matter interactions is given by a constant driving field where the solutions of (59) can be found analytically [22, 130]

$$\begin{aligned} u_1(t) &= (\Delta\,\Omega)/(\Omega_{\text{eff}}^2)\,(1 - \cos\Omega_{\text{eff}}\,t) \\ u_2(t) &= -\Omega/(\Omega_{\text{eff}})\,\sin\Omega_{\text{eff}}\,t \\ u_3(t) &= -(\Delta^2 + \Omega^2\cos\Omega_{\text{eff}}\,t)/(\Omega_{\text{eff}}^2) \end{aligned} \tag{60}$$

Here, $\Omega_{\text{eff}}^2 = \Omega^2 + \Delta$, and we have assumed that Ω is entirely real. From Eq. (60), we readily observe that after a time T given by $\Omega_{\text{eff}}\,T = 2\pi$ the system returns into the initial state. For that reason, pulses of duration T are called 2π-pulses. The phenomenon of a 2π-rotation of the Bloch vector has been given the name *Rabi rotation* [130]. Figure 16 shows the trajectories of the Bloch vector for a pulse with $\Omega\,T = 2\pi$ and for different detunings: only on resonance (i.e., for $\Delta = 0$), the Bloch vector returns at time T to its initial position, whereas off resonance $\boldsymbol{u}$ ends up in an excited state. We shall return to this point later in the discussion of self-induced transparency (Section 6.3). Rabi oscillations are a striking and impressive example of the nonlinear light-matter interaction. Indeed, the solutions (60) clearly show that terms up to infinite order in Ω are required to account for the return of $\boldsymbol{u}$ to its initial position. For semiconductor quantum dots, Rabi oscillations have been measured both in the time [95, 106, 131, 132] and frequency [133] domains. A particularly beautiful experimental setup is due to Zrenner et al. [106], where the authors used a quantum dot embedded in a field effect structure to convert the final exciton population to a photocurrent that could be directly measured. A somewhat different approach was pursued by Kamada et al. [133] where the appearance of additional peaks in the resonance-luminescence spectra at frequencies $\omega_0 \pm \Omega_{\text{eff}}$ centered around the laser frequency ω_0 were observed—a clear signature of Rabi-type oscillations [15, 22].

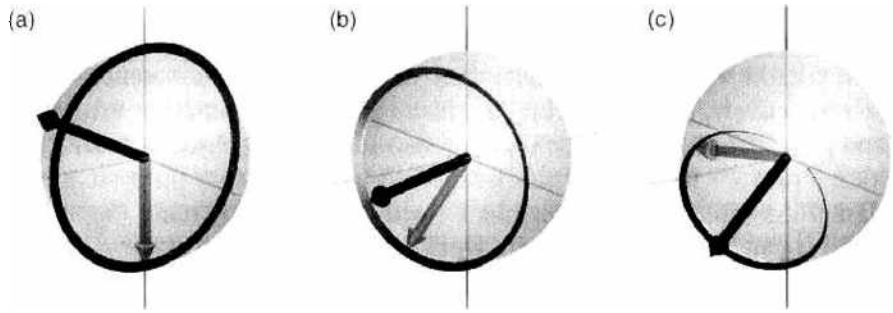


Figure 16. Trajectories of the Bloch vector $\boldsymbol{u}$ for a laser pulse with $\Omega t = 2\pi$ and for detunings Δ of (a) zero, (b) $-\Omega/2$, and (c) $-\Omega$. The light and dark arrows indicate the final position of the Bloch vector and the driving field Ω, respectively. Only on resonance, $\boldsymbol{u}$ returns to its initial positions, whereas off resonance, $\Omega_{\text{eff}} > \Omega$, and the Bloch vector is rotated further.

We conclude this section with a short comment on the coherence properties of the Bloch-vector propagation. In fact, the time evolution (60) of the isolated quantum system can be described in terms of the projector $\boldsymbol{\rho} = |\Psi\rangle\langle\Psi|$, where $|\Psi\rangle$ is the system wavefunction. Such projector-like density operators have the unique property $\boldsymbol{\rho}^2 = \boldsymbol{\rho}$. For the two-level system under consideration, this has the consequence that the length of the Bloch vector remains one throughout. In other words, the trajectory of $\boldsymbol{u}$ is located on the surface of the *Bloch sphere* (Figs. 3 and 16) (i.e., the unit sphere in the Bloch space).

5.2. Decoherence

Decoherence describes the process where a quantum system in contact with its environment loses its quantum-coherence properties. We shall assume that the environment—sometimes referred to as a *reservoir*—has an infinite number of degrees of freedom, and we are not able to specify precisely the corresponding state vector. To account for this lack of information, in the following we adopt the framework of statistical physics. Suppose that the problem under consideration is described by the Hamiltonian $H_S + H_R + V$, where H_S and H_R account for the system and reservoir, respectively, and V for their coupling. Let $w(t)$ be the density operator of the total system in the interaction representation according to $H_S + H_R$. The quantity of interest is the *reduced density operator* $\boldsymbol{\rho}$ of the system alone. It is obtained from the total density operator by tracing over the reservoir degrees of freedom through $\boldsymbol{\rho} = \mathrm{tr}_R w$. We shall now derive the equation of motion for $\boldsymbol{\rho}$ when it is in contact with the environment. As a starting point, we trace in the Lioville von-Neumann equation for w over the reservoir degrees of freedom and obtain

$$\dot{\boldsymbol{\rho}}(t) = \mathrm{tr}_R \dot{w}(t) = -i\,\mathrm{tr}_R[V(t), w(t)] \tag{61}$$

The important feature of this equation is that we are only able to trace out the reservoir on the left-hand side, but not on the right-hand side, which still depends on the full density operator w. The simple and physical deep reason for this is that through V, the system and environment become entangled and can no longer be described independently. Suppose that at an early time t_0, system and reservoir were uncorrelated such that $w(t_0) = \boldsymbol{\rho}(t_0) \otimes \boldsymbol{\rho}_R$, with $\boldsymbol{\rho}_R$ is the density operator of the reservoir. We can then use the usual time evolution operator $U(t, t_0)$ to establish a relation between $w(t)$ and $w(t_0)$. To the lowest order in the system-environment coupling V we find [89]

$$w(t) = w(t_0) - \int_{t_0}^{t} dt' [V(t'), \boldsymbol{\rho}(t_0) \otimes \boldsymbol{\rho}_R] + \mathcal{O}(V^2) \tag{62}$$

This expression no longer depends on the density operator $w(t)$ of the interacting system and environment. We can thus insert Eq. (62) into (61) to obtain an equation of motion for $\boldsymbol{\rho}$ that depends on the system degrees of freedom only. One additional approximation proves to be useful. To the same order in the system-environment interaction V one can replace

$\rho(t_0)$ by $\rho(t)$ [89]—a well-defined procedure that can be justified for any given order of V [134, 135]. This replacement is known as the *Markov approximation* and has the advantage that the time evolution of the density operator $\rho(t)$ only depends on its value at the same instant of time. Then,

$$\dot{\rho}(t) \cong -\int_{t_0}^{t} dt' \, \mathrm{tr}_R [V(t), [V(t'), \rho_R \otimes \rho(t)]] \tag{63}$$

and we have made the assumption $\mathrm{tr}_R [V, \rho(t_0) \otimes \rho_R] = 0$ which holds true in most cases of interest [15, 89]. Equation (63) is our final result. It has the evident structure that at time t' the system becomes entangled with the environment, the entangled system and reservoir propagate for a while—note that V is given in the interaction representation according to $H_S + H_R$—and finally a back-action on the system occurs at time t. Because of the finite interaction time, within the processes of decoherence and relaxation the system can acquire an uncontrollable phase or can exchange energy with the environment. The general structure of Eq. (63) prevails if higher orders of the interaction V are considered [136–139], as will be also discussed at the example of phonon-assisted dephasing in Section 5.5.

5.2.1. Caldeira–Leggett Type Model

To be more specific, in the following we consider the important case where a generic two-level system is coupled linearly to a bath of harmonic oscillators [140–142]

$$H = \frac{E_0}{2} \sigma_z + \sum_i \omega_i a_i^\dagger a_i + i \sum_i g_i (a_i^\dagger \sigma_- - a_i \sigma_+) \tag{64}$$

Here, E_0 is the energy splitting between ground and excited state, ω_i the energies of the harmonic oscillators that are described by the bosonic field operator a_i, and g_i the system-oscillator coupling constant, which is assumed to be real. The last term on the right-hand side defines the system-environment interaction V where we have introduced the lowering and raising operators $\sigma_- = |0\rangle\langle 1|$ and $\sigma_+ = |1\rangle\langle 0|$ for the two-level system. If we insert V in the interaction representation according to $H_S + H_R$ into Eq. (63), we obtain

$$\begin{aligned}
\dot{\rho}(t) \cong -\int_{t_0}^{t} dt' \sum_{ij} g_i g_j \{ &\langle a_i(t) a_j^\dagger(t') \rangle \sigma_+(t) \sigma_-(t') \rho(t) + \langle a_i(t') a_j^\dagger(t) \rangle \rho(t) \sigma_+(t') \sigma_-(t) \\
&- \langle a_i(t) a_j^\dagger(t') \rangle \sigma_-(t') \rho(t) \sigma_+(t) - \langle a_i(t') a_j^\dagger(t) \rangle \sigma_-(t) \rho(t) \sigma_+(t') \\
&+ \langle a_i^\dagger(t) a_j(t') \rangle \sigma_-(t) \sigma_+(t') \rho(t) + \langle a_i^\dagger(t') a_j(t) \rangle \rho(t) \sigma_-(t') \sigma_+(t) \\
&- \langle a_i^\dagger(t) a_j(t') \rangle \sigma_+(t') \rho(t) \sigma_-(t) - \langle a_i^\dagger(t') a_j(t) \rangle \sigma_+(t) \rho(t) \sigma_-(t') \}
\end{aligned} \tag{65}$$

Here, the terms in brackets have been derived by use of cyclic permutation under the trace and describe the propagation of excitations in the environment. The remaining terms with ρ and σ account for the effects of environment coupling on the system. The expression in the second and third line describe *emission* processes where an environment excitation is created prior to its destruction, and those in the fourth and fifth line *absorption* processes where the destruction is prior to the creation. The latter processes usually only occur at finite temperatures. In the following, we suppose that ρ_R describes the reservoir in thermal equilibrium such that $\langle a_i^\dagger a_j \rangle \cong \delta_{ij} n(\omega_i)$, with $n(\omega)$ the usual Bose–Einstein distribution function. Within this spirit, we have also neglected in Eq. (65) terms with $\langle aa \rangle$ and $\langle a^\dagger a^\dagger \rangle$, which would only play a role in specially prepared environments such as squeezed reservoirs [89]. Let us next consider one specific term in (65), which can be simplified according to

$$\int_{t_0}^{t-t_0} d\tau \sum_{ij} g_i g_j \langle a_i(\tau) a_j^\dagger(0) \rangle e^{i E_0 \tau} \cong \sum_i g_i^2 [1 + n(\omega_i)] \gamma(E_0 - \omega, t - t_0) \tag{66}$$

with $\gamma(\Omega, t) = \sin \Omega t / \Omega$ (and we have neglected terms that only contribute to energy renormalizations but not to decoherence and relaxation [89]). $\gamma(\Omega, t)$ has the important feature

that in the adiabatic limit $\lim_{t\to\infty} \gamma(\Omega, t) = \pi\delta(\Omega)$, it gives Dirac's delta function. Thus, for sufficiently long times, the various terms in Eq. (65) account for energy-conserving scattering processes (as discussed below the strict adiabatic $t \to \infty$ limit is usually not needed and it suffices to assume that the reservoir memory is sufficiently short-lived). Within the adiabatic limit, the environment couplings can then be described by the scattering rates $\Gamma_1 = 2\pi \sum_i g_i^2 [1 + n(\omega_i)]\, \delta(\omega_i - E_0)$ and $\Gamma_2 = 2\pi \sum_i g_i^2\, n(\omega_i)\, \delta(\omega_i - E_0)$ for emission and absorption, respectively, and the Lindblad operators $L_1 = \sqrt{\Gamma_1}\,\sigma_-$ and $L_2 = \sqrt{\Gamma_2}\,\sigma_+$ associated to emission and absorption. We can bring Eq. (65) to the compact form

$$\dot{\rho} = -i(H_{\text{eff}}\rho - \rho H_{\text{eff}}^{\dagger}) + \sum_i L_i \rho L_i^{\dagger} \tag{67}$$

where $H_{\text{eff}} = H_S - (i/2)\sum_i L_i^{\dagger} L_i$ is an effective, non-Hermitian Hamiltonian. Equation (67) is known as a *master equation of Lindblad form* [89, 143]. It has the intriguing feature that it is guaranteed that during the time evolution the trace over ρ remains one throughout.

5.2.2. Unraveling of the Master Equation

In many cases of interest, the master equation (67) of Lindblad form can be solved by a simple and particularly transparent scheme. It is known as the *unraveling of the master equation* [144–146]. Recall that the density operator is a statistical mixture of state vectors, $\rho = \sum_k p_k\, |\Psi_k\rangle\langle\Psi_k|$, where the summation over k results from the statistical average of the various pure states $|\Psi_k\rangle$. For simplicity, we restrict ourselves to a single state vector $|\Psi\rangle$. The general case (54) then follows from a straightforward generalization. On insertion of the projector $|\Psi\rangle\langle\Psi|$ into the master equation (67), we obtain

$$\frac{d}{dt}|\Psi\rangle\langle\Psi| = -i(H_{\text{eff}}|\Psi\rangle\langle\Psi| - |\Psi\rangle\langle\Psi|H_{\text{eff}}^{\dagger}) + \sum_i L_i|\Psi\rangle\langle\Psi|L_i^{\dagger} \tag{68}$$

The first term on the right-hand side can be interpreted as a non-Hermitian, Schrödinger-like evolution $i|\dot{\Psi}\rangle = H_{\text{eff}}|\Psi\rangle$ under the influence of H_{eff}. In contrast, the second term describes a time evolution where $|\Psi\rangle$ is projected—or jumps—to one of the possible states $L_i|\Psi\rangle$. For sufficiently small time intervals δt, the time evolution according to H_{eff} is given by $|\Psi(t+\delta t)\rangle = (1 - iH_{\text{eff}}\,\delta t)|\Psi(t)\rangle$. Note that H_{eff} is non-Hermitian and consequently the wavefunction at later time is not normalized. To the lowest order in δt the decrease of norm δp is given by

$$\delta p = i\delta t\langle\Psi(t)|H_{\text{eff}} - H_{\text{eff}}^{\dagger}|\Psi(t)\rangle = \delta t \sum_i \langle\Psi(t)|L_i^{\dagger}L_i|\Psi(t)\rangle = \sum_i \delta p_i \tag{69}$$

The full master equation evolution has to preserve the norm. This missing norm δp is brought in by the states $L_i|\Psi(t)\rangle$ to which the system is scattered with probability δp_i. The time evolution of the density operator can thus be decomposed into

$$\rho(t+\delta t) \cong (1 - \delta p)\rho_0(t+\delta t) + \delta p\, \rho_1(t+\delta t) \tag{70}$$

where $\rho_0(t+\delta t) = U_{\text{eff}}\rho(t)U_{\text{eff}}^{\dagger}$ with the nonunitary time evolution $U_{\text{eff}} \cong 1 - iH_{\text{eff}}\delta t$ accounts for the unscattered part of the density operator and $\rho_1(t+\delta t) = \sum_i \delta p_i\, L_i\, \rho(t)\, L_i^{\dagger}$ for the remainder where a scattering has occurred with probability δp. For that reason, the different parts of ρ are often refered to as *conditional density operators* [146]. This allows for a simple interpretation of the master equation (67): the first term on the right-hand describes the propagation of the system in presence of H_S and out-scatterings, which are responsible for decoherence, and the second one for in-scatterings which result in relaxation. This decomposition will prove particularly useful in the discussion of single-photon sources (Section 5.4).

5.3. Photon Scatterings

The light-photon coupling of Eq. (44) can be described within the framework of the Caldeira–Leggett type model. For a generic two-level system, it is of the form (64) where the summation index i includes the photon wavevector k and polarization λ. The coupling constant reads

$$g_{k\lambda} = \left(\frac{2\pi\omega_k}{\kappa_s}\right)^{1/2} (\hat{e}^*_{k\lambda}\hat{e}) M_{0x} \tag{71}$$

with $\hat{e}$ the exciton polarization defined in Section 3.1.3 and M_{0x} the optical dipole matrix element of Eq. (35). Inserting Eq. (71) into (66) gives for the memory kernel

$$\sum_{k\lambda} \frac{2\pi\omega_k}{\kappa_s} (\hat{e}^*_{k\lambda}\hat{e})^2 |M_{0x}|^2 \gamma(E_0 - \omega_k, t) = \frac{4n}{3\pi c^2} |M_{0x}|^2 \int_0^\infty \omega^3 d\omega \gamma(E_0 - \omega, t) \tag{72}$$

To arrive at the right-hand side, we have replaced the summation over k by $\sum_k \to (2\pi)^{-3} \int_0^\infty k^2 dk \int d\Omega$, and $\int d\Omega$ is the integration over all angles that has been performed analytically [15, 22, 147]. The integral over ω accounts for the temporal buildup of photon scatterings. It is shown in Fig. 17 as a function of time. Most remarkably, the asymptotic value is reached on a timescale of femtoseconds. Thus, in the description of scatterings, one does not have to invoke the strict adiabatic limit $t \to \infty$, but the asymptotic scattering behavior is rather due to the extremely short-lived memory kernel of the reservoir. This is because the system is coupled to an infinite number of photon modes that interfere destructively in the scattering process. If we replace the integral in Eq. (72) by its asymptotic value $\pi\omega_0^3$, we find for the scattering rate of spontaneous photon emission

$$\Gamma = \frac{4n_s\mu_0^2\omega_0^3}{3c^3} \int dr |\Psi_1(r,r)|^2 \sim \int dr |\Psi_1(r,r)|^2 \times 1 \text{ ns}^{-1} \tag{73}$$

This is the generalized Wigner–Weisskopf decay rate for a dipole radiator embedded in a medium with refractive index n_s. The nanosecond timescale given on the right-hand side of Eq. (73) represents a typical value for GaAs- or InGaAs-based quantum dots. The values for $\int dr |\Psi_1(r,r)|^2$ range from one in the strong-confinement regime to several tens in the weak-confinement regime (Appendix A.3), and the corresponding scattering times $1/\Gamma$ from nanoseconds to a few tens of picoseconds [24, 28, 34]. Such finite lifetime of excited exciton

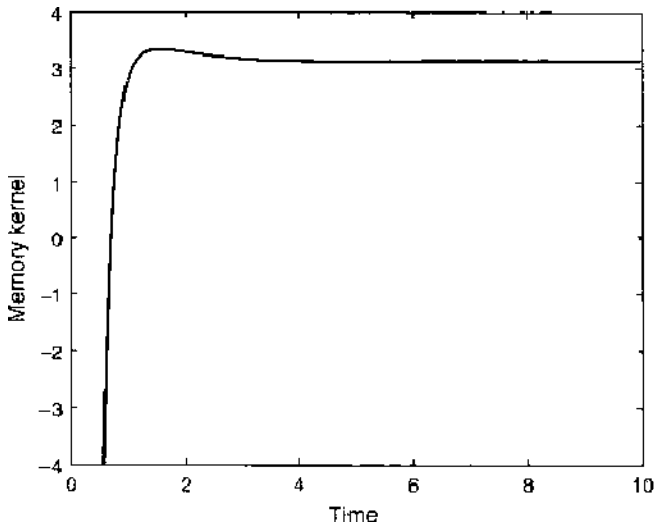


Figure 17. Memory function $\int_0^\infty \omega^3 d\omega \sin[(1-\omega)t]/(1-\omega)$ of Eq. (72), which describes the temporal buildup of photon scatterings. The asymptotic limit is reached on a timescale of $1/E_0$, where E_0 is the energy difference of ground and excited state. For a typical value of $E_0 = 1$ eV, the corresponding time is approximately 0.66 fs. The large negative values at early times are attributed to the somewhat unphysical assumption made in (62) that system and reservoir are initially completely decoupled.

states affects both the coherence properties and the lineshape of optical transitions. We first consider the case of Rabi-type oscillations in presence of a constant laser, which were already discussed at the beginning of this section. The equation of motion for the coherent time evolution follows from Eq. (59) and that for the incoherent part by Eq. (6), where the transverse and longitudinal scattering times $T_2/2 = T_1 = 1/\Gamma$ are given by the Wigner–Weisskopf decay time (for details, see Appendix C). For simplicity, we consider a resonant excitation $\Delta = 0$ and assume Ω to be real. The motion of the Bloch vector then only takes place in the (y, z)-plane and can be computed from

$$\dot{u}_2 = \Omega u_3 - \frac{\Gamma}{2} u_2, \quad \dot{u}_3 = -\Omega u_2 - \Gamma (u_3 + 1) \tag{74}$$

Typical results for the propagation are shown in Fig. 18. We observe that Rabi-flopping occurs but is dampened because of the finite exciton lifetime. In the limit $t \to \infty$, all oscillations become completely dampened and the Bloch vector approaches $\boldsymbol{u} \cong \mathbf{0}$. Such loss of coherence properties is a general property of decoherence and we will encounter similar results for the phonon-assisted dephasing (Section 5.5). We next discuss the influence of a finite exciton lifetime on the lineshape of optical transitions measured in absorption experiments (analogous conclusions hold for luminescence). Our starting point is given by Eq. (39), which, for the two-level system under consideration, states that the absorption spectrum is given by the spectrum of polarization fluctuations $\langle \sigma_-(0)\sigma_-(t)\rangle$. The objective to calculate from the equation of motion for the Bloch vector, which depends on only one time argument, the two-time correlation functions can be accomplished by different means. A popular one is based on the *quantum regression theorem*, which relates for a system initially decoupled from its environment the density-matrix to the two-time correlation functions [22, 89, 148, 149]. The primary idea of this approach is as follows. Let $\boldsymbol{\rho}(t_0)$ denote the density operator at time t_0. The density operator at later time can be obtained by use of the time evolution operator $U(t, t_0)$ through $\boldsymbol{\rho}(t) = U(t, t_0)\,\boldsymbol{\rho}(t_0)\,U^\dagger(t, t_0)$. Upon insertion of a complete set of eigenstates $|i\rangle$, this equation can be transformed to matrix form

$$\begin{aligned} \rho_{ij}(t) &= \sum_{kl} \langle i|U(t, t_0)|k\rangle \langle k|\boldsymbol{\rho}(t_0)|l\rangle \langle l|U^\dagger(t, t_0)|j\rangle \\ &= \sum_{kl} U_{ik}(t, t_0)\, U^*_{jl}(t, t_0)\, \rho_{kl}(t_0) \\ &= \sum_{kl} G_{ij,kl}(t, t_0)\, \rho_{kl}(t_0) \end{aligned} \tag{75}$$

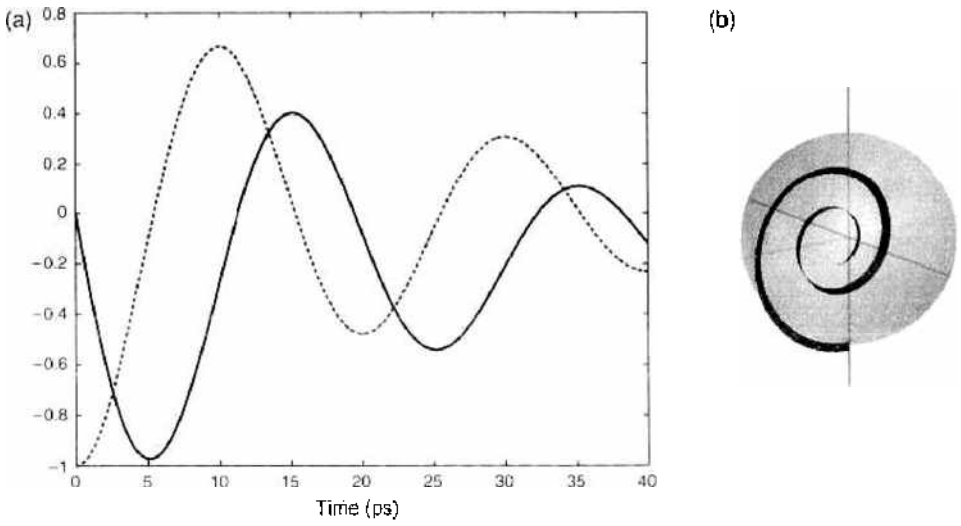


Figure 18. Time evolution of the Bloch vector as computed from Eq. (74) for a constant resonant laser with $\Omega = 0.2$ meV and for a finite upper-state lifetime $1/\Gamma = 40$ ps. The solid and dashed lines in (a) show $u_2(t)$ and $u_3(t)$, respectively, and (b) shows the corresponding trajectory of the Bloch vector $\boldsymbol{u}$.

where the last equality defines the Green function $G_{ij,kl}(t, t_0)$. Once we know the Green function, we can compute the expectation value for any operator A according to $\langle A \rangle = \sum_{ij} \rho_{ij}(t)\, A_{ji} = \sum_{ij,kl} G_{ij,kl}(t, t_0)\, \rho_{kl}(t_0)\, A_{ji}$. However, it also allows the calculation of multitime expectation values. Consider the correlation function for two operators A and B in the Heisenberg picture

$$\begin{aligned}\langle A(t) B(t_0) \rangle &= \operatorname{tr} \boldsymbol{\rho}(t_0)\, U^{\dagger}(t, t_0)\, A\, U(t, t_0)\, B \\ &= \langle i|\boldsymbol{\rho}(t_0)|j\rangle\langle j|U^{\dagger}(t, t_0)|l\rangle\langle l|A|m\rangle\langle m|U(t, t_0)|n\rangle\langle n|B|i\rangle \\ &= U_{mn}(t, t_0)\, U^{*}_{lj}(t, t_0)\, B_{ni}\, \rho_{ij}(t_0)\, A_{lm} \\ &= G_{ml,nj}(t, t_0)\, B_{ni}\, \rho_{ij}(t_0)\, A_{lm} \end{aligned} \tag{76}$$

where we have made use of the usual Einstein summation convention. Importantly, Eq. (76) shows that the two-time correlation function can be computed by replacing the density-matrix $\rho_{ij}(t_0)$ at time t_0 by the modified expression $\sum_k B_{ik}\, \rho_{kj}(t_0)$. The result, which is known as the quantum regression theorem, implies that the fluctuations regress in time like the macroscopic averages. Equation (76) holds exactly, but the factorization of the density operator at time t_0 plays an essential role in the derivation [22]. We shall now show how this result can be used to compute the polarization fluctuations $\langle \sigma_-(t)\sigma_+(t_0)\rangle$. According to the regression theorem, we have to use instead of the initial density operator $\boldsymbol{\rho}(t_0) = |0\rangle\langle 0|$ the modified $\sigma_+\, \boldsymbol{\rho}(t_0) = |1\rangle\langle 0| = \sigma_+$ one. Inserting σ_+ into the Lindblad equation (67) gives

$$\dot{\sigma}_+ = -i\left(E_0 - i\frac{\Gamma}{2}\right)\sigma_+ \tag{77}$$

which shows that the excitation σ_+ propagates with the transition energy E_0 but is dampened because of spontaneous photon emissions. For the correlation function, we obtain $\langle \sigma_-(t)\sigma_+(t_0)\rangle = \exp -i[E_0 - i(\Gamma/2)](t - t_0)$, which, upon insertion into Eq. (39), gives the final result

$$\alpha(\omega) \propto \frac{\Gamma/2}{(\omega - E_0)^2 + (\Gamma/2)^2} \tag{78}$$

The lineshape for optical transitions of a two-level system subject to spontaneous photon emissions is a Lorentzian centered at E_0 and with a full-width of half maximum of $\Gamma/2$. In a nonlinear coherent-spectroscopy experiment, Bonadeo et al. [26] made the important observation that for excitons in the weak-confinement regime energy relaxation and dephasing rates are comparable and predominantly due to photon emissions, thus reflecting the absence of significant pure dephasing. Such behavior is quite surprising for the solid state, as all interaction mechanisms can contribute to T_2 but only a few to T_1. Similar results were also found in the strong confinement regime where, however, things turn out to be more complicated (Section 5.5) [24, 34, 83, 150, 151].

5.4. Single-Photon Sources

Single-photon sources are one of the most promising quantum-dot based quantum devices. The creation of a single photon on demand—first a trigger is pushed and one single photon is emitted after a given time interval—plays an important role in quantum cryptography, for example, for secure key distributions [19, 21]. Gérard and Gayral [152] were the first to propose a turnstile single-photon source based on artificial atoms. Their proposal exploits two peculiarities of artificial atoms: first, because of Coulomb renormalizations of the few-particle states in the decay of a multiexciton state, each photon is emitted at a different frequency (see Fig. 14); second, because of environment couplings—for details see below—photons are always emitted from the few-particle state of lowest energy. Thus, in the cascade decay of a multiexciton complex, the last photon will always be that of the single-exciton decay, and this photon can be distinguished from the others through spectral filtering.

This is how the quantum-dot based single-photon source works: a short pump laser excites electron–hole pairs in the continuum states in the vicinity of the quantum dot, where some

become captured in the dot; the resulting multiexciton complex decays by emitting photons—because of Coulomb renormalizations each photon has a different frequency and because of environment couplings emission only takes place from the respective few-particle states of lowest energy; finally, the "single photon on demand" comes from the last single-exciton decay. Spectral filtering of the last photon is usually accomplished by placing the quantum dot in an optical resonator such as a microcavity [153, 154]. The theoretical description of single scatterings is a quite nontrivial task. The framework of environment couplings developed in this section is based on statistical physics and thus applies to ensembles of (identical) systems only. How do things have to be modified for the description of a single system? Surprisingly enough, not too much. The question of how to describe theoretically such problems first arose almost two decades ago when it became possible to store single ions in a Paul trap and to continuously monitor their resonance fluorescence and led to the development of the celebrated *quantum-jump approach* [144–146]. This approach combines the usual master-equation approach with the rules of demolition quantum measurements [146, 155] and provides a flexible tool for the description of single-system dynamics subject to continuous monitoring. Suppose that the artificial atom and the photon environment at time t_0 are described by the density operator $\boldsymbol{w}(t_0) = \boldsymbol{\rho}(t_0) \otimes \boldsymbol{\rho}_R$. We shall now let the system evolve for a short time δt in presence of the light-matter coupling. This time δt is supposed to be long enough to allow the photon to become separated from the dot—see Fig. 17 for the buildup time of scatterings—and short enough that only a single photon is emitted. What is the probability that a photon is detected within δt? Let $\mathbb{P}_0 = |0_R\rangle\langle 0_R|$ denote the projector on the photon vacuum 0_R. Then

$$P_0 = \mathrm{tr}\,\mathbb{P}_0\, U(t_0 + \delta t, t_0)\, \boldsymbol{w}(t_0)\, U^\dagger(t_0, t_0 + \delta t)\, \mathbb{P}_0 \tag{79}$$

gives the probability that within δt no photon is emitted. The term $U\boldsymbol{w}U^\dagger$ describes the propagation of the quantum dot coupled to photons and the projection operators $\mathbb{P}_0$ the photon detection. With probability P_0, no photon is detected. If we correspondingly project in Eq. (79) on the single-photon subspace $\mathbb{P}_1$, we get the probability P_1 that one photon is detected within δt. Obviously, $P_0 + P_1 = 1$ must be fulfilled. There is one important conclusion to be drawn from Eq. (79). If we compute according to Eq. (63) the time evolution of the density operator to the lowest order in V, but replace the trace over the reservoir by the projection operators $\mathbb{P}_0$ and $\mathbb{P}_1$—which are associated to the outcome of the measurement—we encounter expressions that are completely similar to those of the emission processes in the second and third line of Eq. (65). However, the terms in the second line only show up for projection on $\mathbb{P}_0$ and those in the third line only for projection on $\mathbb{P}_1$. This dependence can be understood as follows. In the quantum-mechanical time evolution (63) of the master equation, the density operator splits up into two terms associated to the situations where a photon is emitted or not. Through the measurement—described by the projection operators $\mathbb{P}_0$ and $\mathbb{P}_1$—we acquire additional information whether a photon has been emitted or not, and we correspondingly have to modify the density operator.

A particularly transparent description scheme for this propagation subject to quantum measurements is given by the master equation (67) of Lindblad form and its unraveling (70) [146, 155]. In the time evolution of $\boldsymbol{\rho}(t)$, we assume that after each time interval δt, a *gedanken* measurement is performed [146, 155], where either no photon or a single photon is detected. These two situations correspond to the two terms of Eq. (70) with probabilities $P_0 = 1 - \delta p$ and $P_1 = \delta p$. The decomposition of the time evolution of the density operator into no-scattering and scattering contributions provides an elegant means to calculate probabilities for finite time intervals $[t_0, t]$. Let us introduce the conditional density operator $\boldsymbol{\rho}_0$, associated to no-photon detection, whose time evolution is given by

$$\dot{\boldsymbol{\rho}}_0 = -i(H_{\mathrm{eff}}\,\boldsymbol{\rho}_0 - \boldsymbol{\rho}_0\, H_{\mathrm{eff}}^\dagger) \tag{80}$$

subject to the initial condition $\boldsymbol{\rho}_0(t_0) = \boldsymbol{\rho}(t_0)$. Because H_{eff} is a non-Hermitian operator, the trace $P_0(t) = \mathrm{tr}\,\boldsymbol{\rho}_0(t)$ decreases and gives the probability that the system has not emitted a photon within $[t_0, t]$. The probability that a photon is emitted at t is given by $-\delta t\, \dot{P}_0(t)$. Once a photon has been detected, we acquire additional information about the system and

accordingly have to change its density operator. This is the point where the second term of the unraveled master equation (70) comes into play. For the photon emission described by the Lindblad operator L_i (i may correspond to the photon polarization),

$$\rho(t+\delta t) \longrightarrow \frac{L_i \rho_0(t) L_i^\dagger}{\mathrm{tr}[L_i \rho_0(t) L_i^\dagger]} \tag{81}$$

gives the density operator right after the scattering (the denominator guarantees that the density operator after the scattering fulfills the trace relation). Table 4 lists some of the basic quantities of this scheme, which is known as the *quantum-jump approach* [146]. Let us consider as a first example the situation where a quantum dot is initially in the single-exciton state. With the Lindblad operator $L = \sqrt{\Gamma}\,\sigma$, associated to photon emission, we obtain for the effective Hamiltonian $H_{\mathrm{eff}} = -i(\Gamma/2)|1\rangle\langle 1|$. Thus, the decay of $P_1(t) = \exp -\Gamma t$ is mono-exponential and the probability to detect a photon at time t is given by $-\delta t\, \dot{P}_1(t) = \delta t\, \Gamma \exp -\Gamma t$.

5.4.1. Photon Antibunching

We next consider the situation where a single quantum dot is driven by a constant laser with Rabi frequency Ω, and the resonance luminescence—sometime refered to as resonance fluorescence—is measured. The quantity we are interested in is the probability that once a photon has been detected at time zero, the *next* photon is detected at time t. It is similar to the two-photon correlation function $g^{(2)}(t)$ [15, 22, 146], with the only difference that we ask for the next instead of any subsequent photon. With the framework of the quantum-jump approach, we are in the position to compute readily things.

This is how it goes. Suppose that the system is initially in state $|0\rangle$. The effective Hamiltonian in presence of the driving laser and of photon emissions is $H_{\mathrm{eff}} = (1/2)[\Delta\,\sigma_3 - \Omega\,\sigma_1 - i\Gamma(\mathbb{1} + \sigma_3)/2]$, with Δ the detuning between the laser and the two-level transition.

Table 4. The primary quantities and equations of interest of the quantum-jump approach [146] (for discussion, see text).

Description	Expression
Full density operator	$w(t)$
Time evolution (63) of $w(t)$	$\dot{w}(t) = -i[V(t), w(t)]$
Conditional density operator for no-photon emission	$\rho_0(t)$
Time evolution of $\rho_0(t)$	$\dot{\rho}_0 = -i(H_{\mathrm{eff}}\rho_0 - \rho_0 H_{\mathrm{eff}}^\dagger)$
Conditional time evolution operator for no-photon emission	$U_{\mathrm{eff}}(t, t_0) = \exp[-iH_{\mathrm{eff}}(t - t_0)]$ for time independent H_{eff}
Conditional time evolution for no-photon emission in $[t_0, t]$	$U_{\mathrm{eff}}(t, t_0)\,\rho(t_0)\,U_{\mathrm{eff}}^\dagger(t, t_0)$
Projection on photon vacuum; no photon detected	$\mathbb{P}_0$
Projection on single-photon subspace; photon detected	$\mathbb{P}_1$
Probability that photon is detected in $[t, t+\delta t]$[a]	$\mathrm{tr}\,\mathbb{P}_1 U(t, t+\delta t) w(t) U^\dagger(t, t+\delta t)\mathbb{P}_1 \rightarrow \delta t \sum_i \mathrm{tr}[\rho(t) L_i^\dagger L_i]$
Probability that no photon is detected in $[t_0, t]$[a]	$P_0(t) = \mathrm{tr}\,\mathbb{P}_0 U(t, t_0) w(t_0) U^\dagger(t, t_0) \rightarrow \mathrm{tr}[U_{\mathrm{eff}}(t, t_0)\rho(t_0)U_{\mathrm{eff}}^\dagger(t, t_0)]$
Full density operator after photon detection at time t	$w(t) = \mathrm{tr}_B[\mathbb{P}_1 w(t)\mathbb{P}_1]\,\mathbb{P}_1$
System density operator after detection of photon i	$\rho(t) = L_i \rho_0(t) L_i^\dagger / \mathrm{tr}[\,]$

[a] The two expressions on the right-hand side correspond, respectively, to the full density operator $w(t)$ and the reduced density operator $\rho(t)$ of the system. The latter is computed within the approximation of a master equation (67) in Lindblad form.

which we assume to be zero. As shown in Appendix C, the probability $P_0(t)$ for no-photon emission can then be computed analytically

$$P_0(t) = e^{-\sinh\theta\,\Omega_{\rm eff}t}\left(1 + \sinh\theta\,\sin\Omega_{\rm eff}t\right) \tag{82}$$

where $\Omega_{\rm eff}^2 = \Omega^2 - (\Gamma/2)^2$ for $\Omega > \Gamma/2$ and we have defined the angle θ through $\tanh\theta = (\Gamma/2\Omega)$. Equation (82) gives the probability that a system—which is initially in its groundstate and is subject to a constant laser field—has emitted no photon within $[0, t]$. Small angles θ refer to the case that photon scatterings occur seldomly on the timescale of $1/\Omega$ and large angles to the case where photon emissions and Rabi flopping take place on the same timescale. Figure 19a shows $P_0(t)$ as computed from Eq. (82) for different values of θ. We observe that in all cases $P_0(t)$ decays exponentially—the decay constant is given by $\sinh\theta\,\Omega_{\rm eff}$—and is modulated by the Rabi-type oscillations of $\sin\Omega_{\rm eff}t$. The latter oscillations reflect the fact that it requires the driving field Ω to bring the system from the ground to the excited state and eventually back to the groundstate and that photons can only be emitted from the excited state. This is also clearly shown in Fig. 19b, which shows the probability distribution $-\dot{P}_0(t)$ for the emission of the next photon: all three curves start at zero, and it requires a finite time to bring the system to the excited state where it can emit a photon. Finally, Fig. 19c shows the probability distribution that after a photon count at time zero, any other photon is detected at t. It is computed from the quantum regression theorem (76) for an initial density operator $|0\rangle\langle 0|$. Although at later times no correlation between the first and the subsequent photon count exist, at early times there is a strong anticorrelation because of the above-mentioned laser-mediated excitation of the upper state. This is a genuine single-system effect—for an ensemble of two-level systems the two-photon correlation would be a Poissonian distribution [22]—and is known as *photon antibunching*. Indeed, such behavior has clearly been observed in the two-photon correlations of single quantum dots. Using pulsed laser excitation, single-photon turnstile devices that generate trains of single-photon pulses were demonstrated [154, 156–160]. In a somewhat different scheme, electroluminescence from a single quantum dot within the intrinsic region of a *p-i-n* junction was shown to act as an electrically driven single-photon source [161, 162]. Also, the decay of multiexciton states has attracted great interest, as it provides a source for multicolor photons with tunable correlation properties [163]. In the quantum cascade decay of the biexciton, it was demonstrated that the first photon emitted from the biexciton-to-exciton decay is always followed by the photon of the single-exciton decay, that is, *photon bunching* [164]. When both photons are emitted along z—which could be achieved, for example, by an appropriate design of the microcavity—the two photons not only differ in energy but also in their polarizations, which could be used for the creation of entangled photons [147, 162, 165–167].

We conclude this section with a more conceptual problem. It is known that under quite broad conditions, quantum measurements lead to a wavefunction collapse. How does this

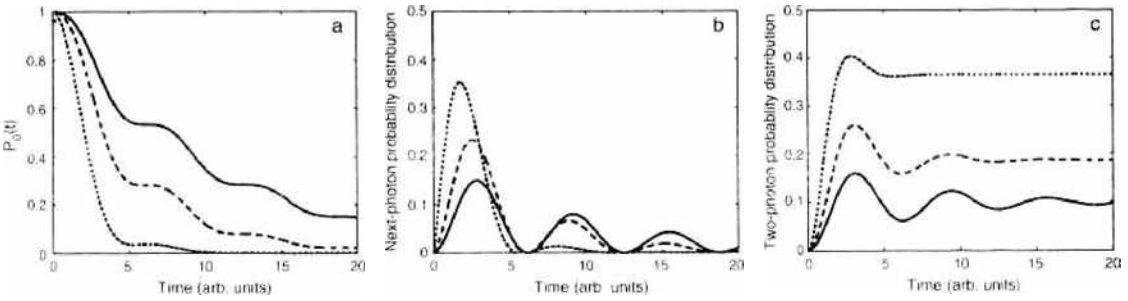


Figure 19. (a) Probability distribution $P_0(t)$ of Eq. (82) that a two-level system initially in the groundstate and subject to a driving laser field and spontaneous photon emissions has up to time t not emitted a photon. Times are measured in units of $1/\Omega_{\rm eff}$, and the angles are $\theta = 0.1$ (solid line), $\theta = 0.2$ (dashed line), and $\theta = 0.5$ (dotted line). (b) Probability distribution $-\dot{P}_0(t)$ that the second photon is emitted at time t. (c) Two-photon probability distribution that any other photon is detected at t, as computed from the quantum regression theorem 76.

collapse show up in the quantum-jump approach under consideration? Suppose that a photon is detected within the time interval $[t + \delta t]$. According to the von Neumann–Lüders rule, the total density operator w has to be changed to [146, 155]

$$w(t + \delta t) \longrightarrow \mathrm{tr}_R[\mathbb{P}_1 U(t + \delta t, t)\, w(t)\, U^\dagger(t + \delta t, t)\, \mathbb{P}_1]\mathbb{P}_0 \tag{83}$$

Here, the projector $\mathbb{P}_1$ accounts for the photon detection and the reservoir trace together with $\mathbb{P}_0$ for the demolition measurement where the photo detector absorbs the photon. The latter procedure leads to the collapse of the photon wavefunction. In Eq. (81), the influence of the photon measurement on $\boldsymbol{\rho}$ is less obvious. In comparison to the full master equation (67), the main effect of the measurement is that we acquire additional information about the photon environment, and we correspondingly modify the system density operator. Is this equivalent to a wavefunction collapse? The solution to this problem is quite subtle. In the derivation of the master equation (67) of Lindblad form, we made the adiabatic approximation $t \to \infty$, which, as shown at the example of photons in Fig. 17, is equivalent to the assumption of a sufficiently short-lived reservoir memory. In other words, in the process of photon emission described by the Lindblad operators L_i, the photon becomes fully decoupled from the system. Thus, when we measure the photon, no back-action on the system occurs. The environment is used "as a witness" [14] that provides information about the system (i.e., whether it has emitted a photon or not). For that reason, we are neither forced to introduce explicitly a wavefunction collapse in the purification process (81) nor does it matter whether the photon is detected directly after emission or travels some distance before detection (as the time evolution of the system is described identically in both cases). On the other hand, we promise that we will use the photon only to perform photon counting but will not try to measure accurately its frequency—which would require a sufficiently long interaction time between the quantum dot and the photon, within which the two objects would become entangled.

5.5. Phonon Scatterings

In addition to the photon coupling, carriers in artificial atoms experience interactions with genuine solid-state excitations such as, for example, phonons. For sufficiently small interlevel splittings, phonon scatterings can be described within the framework of the Caldeira–Leggett model of Section 5.2.1. Theoretical estimates for the corresponding relaxation times are of the order of several tens of picoseconds [168–171]. This finally justifies our assumption made in single-dot spectroscopy that photon emission always occurs from the few-particle states of lowest energy. However, things are considerably more difficult when the interlevel splitting is larger than the phonon energies. This is the case for most types of self-assembled dots where the level splitting is of the order of 50–100 meV, to be compared with the energy of longitudinal optical phonons of 36 meV in GaAs. According to Fermi's golden rule (7), scatterings should here become completely inhibited because of the lack of energy conservation. This led to the prediction of the so-called *phonon bottleneck* [168, 169]. Most experimental studies revealed, however, a fast intradot relaxation of optically excited carriers [101, 172–174]. Furthermore, Borri et al. [24, 83, 150] observed in optical coherence spectroscopy experiments that phonon-induced decoherence can even occur in complete absence of relaxation. Such decoherence is due to the lattice deformation induced by the optical excitation and the resulting formation of a *polaron* (i.e., a composite exciton-phonon excitation). In the following, we first briefly review the theoretical description of polarons and phonon-assisted dephasing. Based on this, we then reexamine relaxation processes beyond the framework of Fermi's golden rule.

5.5.1. Spin-Boson Model

Consider the model where a generic two-level system is coupled linearly to a reservoir of harmonic oscillators such that the interaction only occurs when the system is in the upper state [23]

$$H = E_0 |1\rangle\langle 1| + \sum_i \omega_i a_i^\dagger a_i + \sum_i g_i (a_i^\dagger + a_i) |1\rangle\langle 1| \tag{84}$$

Here, ω_i is the phonon energy, a_i the bosonic field operator for phonons, and g_i the coupling constant (details will be presented below). In comparison to the Caldeira–Leggett type model (64), the so-called *spin-boson model* of Eq. (84) does not induce transitions between the two levels. Yet, it leads to decoherence. This can easily be seen by writing Eq. (84) in the interaction picture according to the Hamiltonian for the uncoupled quantum dot and phonons

$$V(t) = \sum_i g_i (e^{i\omega_i t} a_i^\dagger + e^{-i\omega_i t} a_i) |1\rangle\langle 1| \tag{85}$$

Through the phonon coupling, the two-level system becomes entangled with the phonons, where each phonon mode evolves with a different frequency ω_i. If we trace out the phonon degrees of freedom—similarly to the procedure employed in the derivation of the Lindblad equation (67)—the different exponentials $e^{\pm i\omega t}$ interfere destructively, which leads to decoherence. Because this decoherence is not accompanied by relaxation, the process has been given the name *pure dephasing*. We shall now study things more thoroughly. We first note that the dot-phonon coupling term can be removed through the transformation [23, 175, 176]

$$e^s H e^{-s} = \bar{E}_0 |1\rangle\langle 1| + \sum_i \omega_i a_i^\dagger a_i \tag{86}$$

with $s = |1\rangle\langle 1| \sum_i \xi_i (a_i^\dagger - a_i)$ an anti-Hermitian operator, $\bar{E}_0 = E_0 - \sum_i (g_i)^2/\omega_i$ the renormalized two-level energy, and $\xi_i = g_i/\omega_i$. The simple physical reason is that for the Hamiltonian (84), the oscillator equilibrium positions are different for the ground and excited states of the two-level system, and e^s—which is closely related to the usual displacement operator $D(\xi) = e^{\xi a^\dagger - \xi^* a}$ of the harmonic oscillator [15, 22, 89]—accounts for this displacement of positions. Let us first study the lineshape of optical transitions resulting from the phonon coupling (84). As shown in Appendix D, within the spin-boson model, the polarization fluctuations governing the absorption spectra (39) can be computed analytically [23]

$$\langle \sigma_-(0)\sigma_+(t)\rangle = e^{i\bar{E}_0 t} \exp \sum_i \left(\frac{g_i}{\omega_i}\right)^2 \left[i \sin \omega_i t - (1 - \cos \omega_i t) \coth \frac{\beta\omega_i}{2} \right] \tag{87}$$

with β the inverse temperature. Because the final result (87) is exact (within the limits of our model Hamiltonian), it can be employed for arbitrarily strong phonon couplings g_i. In addition, it provides a prototypical model for decoherence that has found widespread applications in various fields of research [128, 135, 177–179]. For semiconductor quantum dots, Eq. (87) and related expressions have widely been used for the description of optical properties [25, 176, 179–182]. We now follow Krummheuer et al. [25] and derive explicit results for GaAs-based quantum dots. For simplicity, we assume a spherical dot model and acoustic deformation potential interactions [25, 29]

$$g_q = \left(\frac{q}{2\rho c_\ell}\right)^{1/2} (D_e - D_h)\, e^{-q^2 L^2/4} \tag{88}$$

as the only coupling mechanism. Here, q is the phonon wavevector, ρ the mass density, c_ℓ the longitudinal sound velocity, D_e and D_h the deformation potentials for electrons and holes, respectively, and the electron and hole wavefunctions have been approximated by Gaussians with the same carrier localization length L. Because the exponential in (88) introduces an effective cutoff for the wavevectors q, we do not have to account explicitly for the cutoff at the Debye frequency. It turns out to be convenient to measure length in units of L, wavevectors in units of $1/L$, energy in units of c_ℓ/L, and time in units of L/c_ℓ. With material parameters representative for GaAs [25, 29] and a carrier localization length $L = 5$ nm, we obtain, respectively, 1 ps, 0.7 meV, and 7.8 K for the time, energy, and temperature scale. The dot-phonon coupling strength can be expressed in dimensionless form

$$\alpha_p = \frac{(D_e - D_h)^2}{4\pi^2 \rho\, c_\ell^3\, L^2} \cong 0.033 \tag{89}$$

where the estimate on the right-hand side corresponds to the material and dot parameters listed above. With the natural units for time, energy, and temperature, the coupling strength α_p of Eq. (89) becomes the only parameter of the spin-boson model. We then get $\sum_i (g_i)^2/\omega_i = \alpha_p(\pi/2)^{1/2}$ for the renormalization of the two-level energy and $\exp \alpha_p \int_0^\infty x dx\, e^{-x^2/4} [i \sin xt - (1 - \cos xt) \coth(\beta x/2)]$ for the polarization correlation function (87). This function is shown in Fig. 20a for different temperatures. We observe a decay at early times associated to phonon dephasing—that is, part of the quantum coherence is transfered from the two-level system to the phonons—and the curves approach a constant value at later times. The asymptotic value decreases with increasing temperature.

In a sense, this finding is reminiscent of the Franck–Condon principle of optically excited molecules: because the equilibrium positions of the ions in the ground and excited state are different, after photoexcitation the molecule ends up in an excited vibrational state. However, in contrast to the molecule, which just couples to a few vibrational modes, optical excitations in artificial atoms couple to a continuum of phonon modes that all evolve with a different frequency. In the spirit of the random-phase approximation, this introduces decoherence. Figure 20b shows the imaginary part of the Fourier transform of $\langle \sigma_-(0)\sigma_+(t)\rangle$, which is proportional to absorption. In addition to the delta-peak at energy $\bar{E}_0$—which would acquire a Lorentzian shape (78) in presence of photon emissions—the spin-boson coupling (84) gives rise to a broad continuum in the optical spectra that increases with increasing temperature. Such behavior has been observed experimentally and attributed to phonon dephasing [24].

Rabi-Type Oscillations We next discuss the influence of the dot-phonon coupling (84) on the coherent optical response of artificial atoms. In contrast to the previous section, where all results could be obtained analytically, in presence of a laser pulse Ω with arbitrary strength, the solution of the equations of motion is more cumbersome [128, 177, 179], and one is forced to introduce an approximate description scheme [29, 30, 183]. To this end, in the following we adopt a density-matrix description. Our starting point is given by the Heisenberg equations of motion for $\boldsymbol{\sigma}$ and a_i, according to the Hamiltonian (58) and the spin-boson coupling (84),

$$\dot{\boldsymbol{\sigma}} = \boldsymbol{\Omega} \times \boldsymbol{\sigma} + \sum_i g_i (a_i^\dagger + a_i)\hat{\boldsymbol{e}}_3 \times \boldsymbol{\sigma}, \qquad \dot{a}_i = -i(\omega_i a_i + g_i |1\rangle\langle 1|) \tag{90}$$

The vector $\boldsymbol{\Omega}$ is defined in Eq. (59). Our objective now is to derive from (90) an approximate equation of motion for the Bloch vector $\boldsymbol{u} = \langle \boldsymbol{\sigma} \rangle$. Multiplying in Eq. (90) the total density operator w from the left-hand side and tracing over the system and phonon degrees

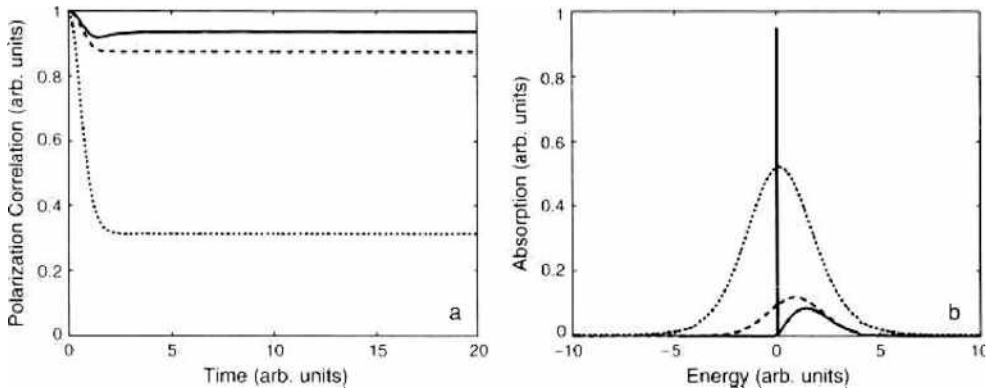


Figure 20. (a) Polarization fluctuations and (b) their Fourier transforms, which are proportional to absorption, as computed within the spin-boson model of Eq. (87) for temperatures of 0.1 (solid lines), 1 (dashed lines), and 10 (dotted lines). For material and dot parameters representative for GaAs (i.e., a mass density $\rho = 5.27$ g cm^{-3}, a longitudinal sound velocity $c_l = 5110$ m/s, deformation potentials $D_e = -14.6$ eV and $D_h = -4.8$ eV for electrons and holes, respectively, and a carrier localization length $L = 5$ nm [25, 29]) time, energy, and temperature are measured in units of 1 ps, 0.7 meV, and 7.8 K, respectively. We assume a dot-phonon coupling strength $\alpha_p = 0.033$ (for details, see text). In (b), energy zero is given by the renormalized energy $\bar{E}_0$ of the two-level system.

of freedom shows that the Bloch vector couples to $\langle a_i \boldsymbol{\sigma} \rangle$. If we would derive similarly to Eq. (90) the equation of motion for $\langle a_i \boldsymbol{\sigma} \rangle$, we would find that it couples to higher order density matrices such as $\langle a_i a_j \boldsymbol{\sigma} \rangle$. This is because through the spin-boson coupling (84) the two-level system becomes entangled with the phonons, and each density matrix couples to a matrix of higher order. The resulting infinite hierarchy of density matrices has been given the name *density matrix hierarchy* [27, 184]. We shall now introduce a suitable truncation scheme. Consider a generic expectation value $\langle AB \rangle$ with two arbitrary operators A and B. If there was no correlation between the two operators, the expectation value would simply be the product $\langle A \rangle \langle B \rangle$.

We shall now lump all correlations between A and B into the correlation function $\langle\langle AB \rangle\rangle$, and express the expectation value of the two operators through $\langle AB \rangle = \langle A \rangle \langle B \rangle + \langle\langle AB \rangle\rangle$. Corresponding factorization schemes—which are known as *cumulant expansions*—apply to expectation values with more operators such as, for example, $\langle ABC \rangle$ [27, 185, 129]. In the common truncation scheme of the density-matrix hierarchy, one selects a few cumulants, which are expected to be of importance, and neglects all remaining ones. To the lowest order of approximation, within the spin-boson model we keep the Bloch vector $\langle \boldsymbol{\sigma} \rangle$, the coherent phonon amplitude $s_i = \langle a_i \rangle$, and the phonon-assisted density matrix $\boldsymbol{u}_i = \langle\langle a_i \boldsymbol{\sigma} \rangle\rangle$ as dynamic variables. Their equations of motion can readily be obtained from Eq. (90), and we obtain

$$\begin{aligned} \dot{\boldsymbol{u}} &= \boldsymbol{\Omega}_{\text{eff}} \times \boldsymbol{u} + 2 \sum_i g_i \hat{\boldsymbol{e}}_3 \times \Re e(\boldsymbol{u}_i) \\ \dot{s}_i &= -i\omega_i s_i - \frac{i}{2} g_i (1 + u_3) \\ \dot{\boldsymbol{u}}_i &= \boldsymbol{\Omega}_{\text{eff}} \times \boldsymbol{u}_i - i\omega_i \boldsymbol{u}_\lambda + g_i \left(n_i + \frac{1}{2} \right) \hat{\boldsymbol{e}}_3 \times \boldsymbol{u} + \frac{i}{2} g_i (u_3 \boldsymbol{u} - \hat{\boldsymbol{e}}_3) \end{aligned} \tag{91}$$

with $\boldsymbol{\Omega}_{\text{eff}} = \boldsymbol{\Omega} + 2\Re e \sum_i g_i s_i \hat{\boldsymbol{e}}_3$, and $n_i = \langle\langle a_i^\dagger a_i \rangle\rangle$ the phonon distribution function that we approximate by the thermal distribution $n(\omega_i)$. We can correspondingly keep higher order cumulants such as $\langle\langle a_i^\dagger a_j \boldsymbol{\sigma} \rangle\rangle$ whose equations of motion are considerably more complicated [29, 30]. On general grounds, we expect that the neglect of higher order cumulants is appropriate for sufficiently low temperatures and weak dot-phonon couplings g_i—or equivalently α_p defined in Eq. (89)—which is a valid assumption for GaAs-based quantum dots ($\alpha_p \cong 0.033$) at low temperatures. Figure 21 shows results of calculations based on Eq. (91) for a 2π-pulse and for temperatures of $T = 1$ and $T = 10$. We observe that Rabi flopping occurs but is dampened because of the phonon-assisted dephasing. In particular at higher temperatures, phonon dephasing is of strong importance and gives rise to decoherence on a picosecond timescale. We shall return to this point in Section 6.

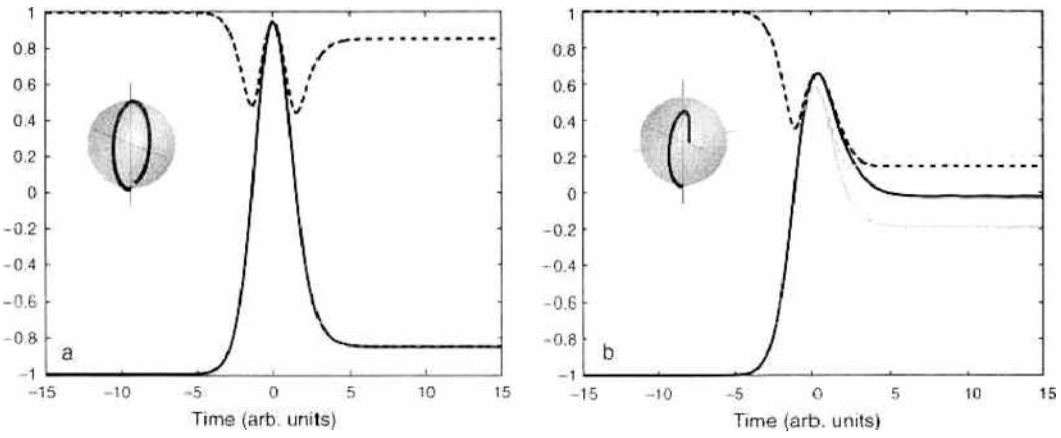


Figure 21. Rabi flopping in presence of phonon-assisted dephasing at temperatures of (a) $T = 1$ and (b) $T = 10$ and for a Gaussian laser pulse with a full-width at half maximum of 5 (for units, see caption to Fig. 20). The solid and dashed lines show u_3 and $\|\boldsymbol{u}\|$, respectively. The dark curves show results of calculations including $\boldsymbol{u}$, s_i, and $\boldsymbol{u}_i$ as dynamic variables, and the gray ones [which are indistinguishable in panel (a)] those of calculations that additionally include $\langle\langle a_i a_i \boldsymbol{\sigma} \rangle\rangle$, $\langle\langle a_i a_i^\dagger \boldsymbol{\sigma} \rangle\rangle$, and $\langle\langle a_i^\dagger a_i \boldsymbol{\sigma} \rangle\rangle$. The insets report the trajectories of the Bloch vector.

Beyond the Spin-Boson Model Although in many cases of interest the spin-boson model (84) provides a sufficiently sophisticated description of optical excitations in artificial atoms, there are situations where it is expected to break down. For instance, at higher temperatures, anharmonic decay of phonons—a phonon decays in an energy-conserving scattering into two phonons of lower energy—or higher order phonon processes [186] could play a decisive role. If the artificial atom can no longer be described as a genuine two-level system, one has to consider additionally phonon-mediated scattering channels. Quite generally, the strong polar-optical coupling to longitudinal optical phonons introduces a marked deformation of the surrounding lattice and the formation of a *polaron* [187–189]. Because optical phonons have a very small dispersion, this interaction channel has no significant impact on decoherence. However, when the system is in an excited exciton or multiexciton state, the anharmonic decay of phonons contributing to the polaron allows for relaxation processes even in absence of energy matching between the unrenormalized dot transition and the phonons [180, 190–194]. This demonstrates that phonon relaxation and decoherence in artificial atoms is more efficient than one would expect in a simple-minded Fermi's golden-rule picture. Future work will show to what extent carrier-phonon interactions can be tailored in artificial atoms [195, 196] and whether phonon-assisted dephasing can eventually be strongly suppressed or fully overcome [183, 197, 198].

5.6. Spin Scatterings

So far, we have seen that photon and phonon scatterings occur on a timescale ranging from several tens of picoseconds to nanoseconds. Such decoherence times are remarkably long for the solid state but are rather short when it comes to more sophisticated quantum control applications (Sections 6 and 7). Optical excitations in quantum dots possess another degree of freedom that has recently attracted enormous interest: *spin* [53]. Spin couples weakly to the solid-state environment and is therefore expected to be long lived. Optics provides a simple means to modify spin degrees of freedom through coupling to the charge degrees. This dual nature of optical excitations is exploited in quantum-computation proposals, to be discussed in Section 7. What are the typical spin relaxation and decoherence times in artificial atoms? Things are quite unclear. Experimentally, it was found that at low temperature spin relaxation is almost completely quenched [34, 35, 199, 200]. Theoretical estimates indicate relaxation times of the order of microseconds or above [201], whereas almost no conclusive results exist for the pertinent decoherence times. Thus, spin keeps its secret in the game and holds a lot of promise and hope.

6. QUANTUM CONTROL

Recent years have witnessed enormous interest in controlling quantum phenomena in a variety of nanoscale systems [202]. Quite generally, such control allows to modify the system's wavefunction at will through appropriate tailoring of external fields (e.g., laser pulses): whereas in *quantum optics*, the primary interest of this wavefunction engineering lies on the exploitation of quantum coherence among a few atomic levels [15, 22, 89], in *quantum chemistry*, optical control of molecular states has even led to the demonstration of optically driven chemical reactions of complex molecules [203]; furthermore, starting with the seminal work of [204], coherent-carrier control in semiconductors and semiconductor nanostructures has been established as a mature field of research on its own. These research arenas have recently received further impetus from the emerging fields of quantum computation and quantum communication [19], aiming at quantum devices where the wavefunction can be manipulated with highest possible precision. It is worth emphasizing that hitherto there exists no clear consensus of how to tailor optimally the system's control, and it appears that each field of research has come up with its own strategies: for instance, quantum-optical implementations in atoms benefit from the long atomic coherence times of metastable states, and it usually suffices to rely on the solutions of effective models (e.g., adiabatic population transfer in an effective three-level system [205]): in contrast, in quantum chemistry, the complexity of molecular states usually does not permit schemes that are solely backed from

the underlying level schemes, and learning algorithms, which receive direct feedback from experiment, appear to be the method of choice. Finally, coherent control in semiconductor nanostructures has hitherto been primarily inspired by quantum-optical techniques; however, it is clear that control in future quantum devices will require more sophisticated techniques to account for the enhanced decoherence in the solid state.

This section is devoted to an introduction into the field. Throughout, we shall use the laser-induced quantum coherence as the workhorse, which allows to bring the system to almost any desired state [127]. On the other hand, such ideal performance is spoiled by the various decoherence channels at play (e.g., the ones discussed in the previous section). From the field of quantum optics, a number of control strategies are known that allow to suppress or even overcome decoherence losses. In Section 6.1, we shall discuss one of the most prominent ones: stimulated Raman adiabatic passage [205], a technique that exploits the renormalized states in presence of strong laser fields for a robust and high-fidelity population transfer. Because of its simplicity, there has recently been strong interest in possible solid-state implementations [206–213]. There exist other control strategies, which are either based on schemes developed in the fields of nuclear magnetic resonance [214–218] or rely on general optimization approaches such as optimal control [203, 219, 220] or genetic algorithms [203, 221]. Below, we shall review the latter two approaches. Similar to the last section, it will prove useful to rely on effective level schemes, which grasp the main features of the excitonic and multiexciton quantum dot states (Fig. 22). Finally, for a more extensive discussion of coherent optical spectroscopy and coherent carrier control in quantum dots, the reader is referred to the literature [28, 34, 93, 95, 100, 106, 222].

6.1. Stimulated Raman Adiabatic Passage

Let us first consider the Λ-type level scheme depicted in Fig. 22b. It consists of two long-lived states $|0\rangle$ and $|1\rangle$, which are optically connected through a third short-lived state $|2\rangle$. Such a level scheme may correspond to a coupled dot charged with one surplus carrier, where states $|0\rangle$ and $|1\rangle$ are associated to the carrier localization in one of the dots and $|2\rangle$ to the charged exciton, which allows optical coupling between states $|0\rangle$ and $|1\rangle$ [206]; alternatively, we may associate the two lower states $|0\rangle$ and $|1\rangle$ to the spin orientation of one surplus electron in a single quantum dot, where in presence of a magnetic field along x (i.e., Voigt geometry, Section 3.1.3), the two states can be optically coupled through the charged exciton $|2\rangle$ (see Refs. [213, 223, 224] and Section 7). Quite generally, for the level scheme of Fig. 22b and assuming that the system is initially prepared in state $|0\rangle$, in the following we shall ask the question: What is the most efficient way to bring the system from $|0\rangle$ to $|1\rangle$? Suppose that the frequencies of two laser pulses are tuned to the 0–2 and 1–2 transitions, respectively. For reasons to become clear in a moment, we shall refer to the pulses as *pump* and *Stokes*.

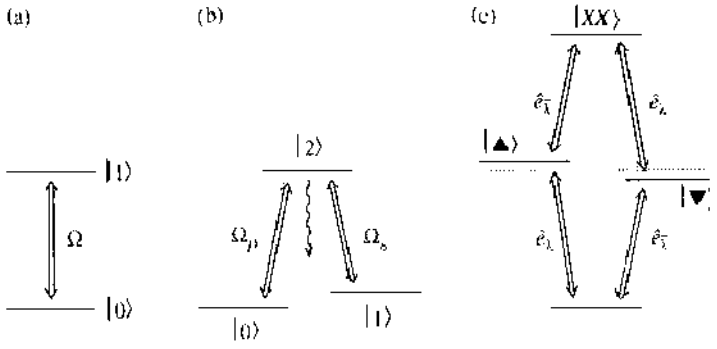


Figure 22. Prototypical dot-level schemes. (a) Two-level system with $|0\rangle$ and $|1\rangle$ the ground and excited states; Ω denotes the Rabi frequency in presence of a light field. (b) Λ-type scheme, for example, carrier states in coupled dots [206]; $|0\rangle$ and $|1\rangle$ are long-lived states, whereas $|2\rangle$ is a short-lived state that is optically coupled to both $|1\rangle$ and $|2\rangle$ (for details, see Section 6.1); the wiggled line indicates spontaneous photon emission. (c) Exciton states in a single dot: $|0\rangle$ is the vacuum state; $|\blacktriangle\rangle$ and $|\blacktriangledown\rangle$ are the spin-degenerate single-exciton groundstates, and $|XX\rangle$ is the biexciton groundstate; optical selection rules for light polarizations $\hat{e}_\alpha$ or $\hat{e}_\beta$ (either circular or linear) apply as indicated in the figure.

Because direct optical transitions between $|0\rangle$ and $|1\rangle$ are forbidden, we have to use $|2\rangle$ as an auxiliary state; however, intermediate population of $|2\rangle$ introduces losses through environment coupling (e.g., spontaneous photon emissions or phonon-assisted dephasing). We can use the master equation (67) of Lindblad form to describe the problem. Within a rotating frame, we obtain the effective hamiltonian [205, 225]

$$H_{\text{eff}} = -\frac{1}{2}\begin{pmatrix} 0 & 0 & \Omega_p^* \\ 0 & 0 & \Omega_s^* \\ \Omega_p & \Omega_s & \epsilon \end{pmatrix} \tag{92}$$

with $\epsilon = 2\Delta + i\Gamma$, Δ the detuning of the lasers with respect to the 0–2 and 1–2 transitions, and Γ the inverse lifetime of the upper state. Ω_p and Ω_s are the Rabi frequencies for the pump and Stokes pulse, respectively. Throughout we assume that Ω_p only affects the 0–2 transition and Ω_s only the 1–2 one, which can, for example, be achieved through appropriate polarization filtering, or for sufficiently large 0–1 splittings through appropriate choice of laser frequencies. Figure 23b shows results of simulations for different time delays between the Stokes and pump pulse, and for different pulse areas $A_s = \int_{-\infty}^{\infty} dt\ \Omega_s(t)$ (we assume $A_s = A_p$): black corresponds to successful and white to no population transfer. In the case of the "intuitive" ordering of laser pulses where the pump pulse excites the system *before* the Stokes pulse (i.e., negative time delays in Fig. 23b), one observes enhanced population transfer for odd multiples of π. This is associated to processes where the pump pulse first excites the system from $|0\rangle$ to $|2\rangle$, and the subsequent Stokes pulse brings the system from $|2\rangle$ to $|1\rangle$. However, the large black area at positive time delays in Fig. 23b suggests that there is a more efficient way for a population transfer. Here, the two pulses are applied in the "counterintuitive" order (i.e., the pump pulse is turned on *after* the Stokes pulse). Because of the resemblance of this scheme with a Raman-type process, it has become convenient to introduce the expression of a *Stokes* pulse, and the whole process has been given the name *stimulated Raman adiabatic passage* [205]. This process fully exploits the quantum coherence introduced by the intense laser fields.

In presence of the Stokes pulse, the dot-states become renormalized, and these renormalized states are used by the pump pulse for a robust and high-fidelity population transfer. Although Fig. 23 presents solutions of the full master equation (67), in the following we shall only consider the time evolution of ρ due to the effective Hamiltonian (92). If Ω_p and Ω_s

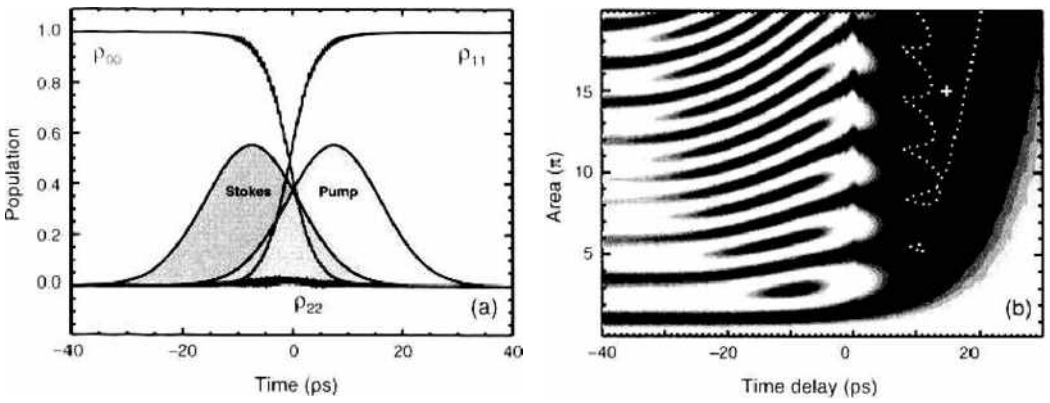


Figure 23. Simulations of coherent population transfer in coupled dots: (a) transients of the populations ρ_{00}, ρ_{11}, and ρ_{22} [for level scheme, see Fig. 22b]; (b) contour plot of final population ρ_{11} as a function of time delay between Stokes and pump pulse and of pulse area $A_s = A_p$; white corresponds to values below 0.1, black to values above 0.9; the dashed line gives the contour of $\rho_{11} \geq 0.999$, and the cross indicates the values used in panel (a). In our simulations, we use the same Gaussian envelopes for the Stokes and pump pulses (with the time delay given in the figure) and a full-width at half maximum of 20 ps. Parameters are chosen according to Ref. [206].

have a sufficiently slow time dependence—as will be specified in more detail further below—at each instant of time the system is characterized by the eigenvalues and eigenvectors of $H_{\rm eff}$. Straightforward algebra yields for $\Gamma \ll \Omega_s$, $\Gamma \ll \Omega_p$ the eigenvalues

$$\varpi_0 = 0, \quad \varpi_\pm = \frac{1}{2}(\Delta \pm \Omega_{\rm eff}) + i(1 - \Delta/\Omega_{\rm eff})\Gamma \tag{93}$$

with $\Omega_{\rm eff}^2 = \Delta^2 + \Omega_p^2 + \Omega_s^2$. Most importantly, eigenvalue ϖ_0 has no imaginary part and consequently does not suffer radiative losses (this holds even true for large values of Γ). Indeed, introducing the time-dependent angle θ through $\tan\theta = \Omega_p/\Omega_s$ we observe that the corresponding eigenvector

$$|a_0\rangle = \cos\theta|1\rangle \quad \sin\theta|2\rangle \tag{94}$$

has *no* component of the "leaky" state $|2\rangle$—in contrast to the eigenvectors $|a_\pm\rangle$ which are composed of all three states $|0\rangle$, $|1\rangle$, and $|2\rangle$. In $|a_0\rangle$, the amplitudes of the 0–2 and 1–2 transitions interfere destructively, such that the state is completely stable against absorption and emission from the radiation fields. For that reason, state $|a_0\rangle$ has been given the name *trapped state*. The population transfer process exploits this trapped state as a vehicle in order to transfer population between states $|0\rangle$ and $|1\rangle$. It is achieved by using overlap in time between the two laser pulses (Fig. 23a, Table 5). Initially, the system is prepared in state $|0\rangle$. When the first (Stokes) laser is smoothly turned on, the system is excited at the Stokes frequency. At this frequency no transitions can be induced; what the pulse does, however, is to align the time-dependent state vector $|\Psi(t)\rangle$ with $|a_0(t)\rangle = |0\rangle$ (as $\theta = 0$ in the sector $\Omega_s \neq 0$, $\Omega_p = 0$), and to split the degeneracy of the eigenvalues ϖ_0 and $\varpi_\perp$. Thus, if the pump laser is smoothly turned on—such that throughout $\Omega_{\rm eff}(t)$ remains large enough to avoid non-adiabatic transitions between $|a_0(t)\rangle$ and $|a_+(t)\rangle$ [205]—all population is transferred between states $|0\rangle$ and $|1\rangle$ within an adiabatic process where $|\Psi(t)\rangle$ directly follows the time-dependent trapped state $|a_0(t)\rangle$. Stimulated Raman adiabatic passage is a process important for a number of reasons. First, it is a prototypical example of how intense laser fields can cause drastic renormalizations of carrier states; quite generally, these "dressed" states exhibit novel features in case of quantum interference (i.e., if three or more states are optically coupled). Second, quantum control in quantum dots has recently attracted increasing interest in view of possible quantum computation applications [213, 216, 226, 227] aiming at an all-optical control of carrier states; in this respect, the adiabatic transfer scheme might be of some importance because of its robustness and its high fidelity (Section 7). More specifically, from Fig. 23b it becomes apparent that the transfer works successfully within a relatively large parameter regime—in contrast to the "intuitive" order of pulses, where a detailed knowledge of the dipole matrix elements and a precise control over the laser parameters is required. Thus, it provides a robust scheme that only relies on sufficiently smooth and strong laser pulses.

6.2. Optimal Control

In many cases of interest it is more difficult—or even impossible—to guess control strategies solely based on physical intuition, and one is forced to rely on more general control schemes. Suppose that the system under investigation is described by the n-component vector $x = (x_1, x_2, \ldots x_n)$ of dynamic variables, whose equations of motion are given by the differential

Table 5. Time evolution of the quantities characterizing the stimulated Raman adiabatic passage [205] (for discussion, see text).

Quantity	Ω_s on	Ω_s and Ω_p on	Ω_p on
Angle θ	0	$0 \to \pi/2$	$\pi/2$
$\sim \varpi_\pm - \Delta/2$	$\pm\frac{1}{2}(\Delta^2 + \Omega_s^2)^{1/2}$	$\pm\frac{1}{2}(\Delta^2 + \Omega_s^2 + \Omega_p^2)^{1/2}$	$\pm\frac{1}{2}(\Delta^2 + \Omega_p^2)^{1/2}$
ϖ_0	0	0	0
Trapped state $\lvert a_0\rangle$	$\lvert 0\rangle$	$\lvert 0\rangle \to \lvert 1\rangle$	$\lvert 1\rangle$

equations $\dot{x}_i = F_i(x, \Omega)$ with F a functional that depends on all state variables x and the control fields Ω. Here, x may refer to the different components of a wavefunction ψ that obeys Schrödinger's equation, or to the different components of the density operator ρ, or to the cumulants of a density-matrix approach. The main assumption we shall make is that of a Markovian time evolution, that is, we assume that for the time evolution of $x_i(t)$, the functional F only depends on the variables $x(t)$ at the same instant of time. The components x_i are supposed to be real, which can always be achieved by separating ψ_i or ρ_{ij} into their respective real and imaginary parts. Initially the system is in the state x_0. Quite generally, in the field of quantum control we are seeking for control fields Ω that bring the system from x_0 at time zero to the desired final state x_d at time T, or promote the system within the time period $[0, T]$ through a sequence of desired states. To evaluate a given control Ω, we have to quantify its success through the *cost functional*

$$J(x, \Omega) = J_T(x) + \tilde{J}(x, t) + \frac{\gamma}{2} \int_0^T dt \sum_i \|\Omega_i(t)\|^2 \tag{95}$$

Here, $J_T(x)$, which only depends on $x(T)$, accounts for the terminal conditions and rates how close $x(T)$ is to the desired state x_d (e.g., through $\frac{1}{2}\|\psi(T) - \psi_d\|^2$), $\tilde{J}(x, t)$ is a functional that accounts for other control objectives within $[0, T]$ (e.g., the wish to suppress the population of certain states), and the last term accounts for the limited laser resources. Our task now is to determine a control that minimizes $J(x, \Omega)$ subject to the constraint that x fulfills the dynamic equations $\dot{x}_i = F_i(x, \Omega)$ with the initial condition $x(0) = x_0$. Within the framework of *optimal control* [203, 219, 220], this is accomplished by introducing Lagrange multipliers $\tilde{x}$ for the constraints, and turning the constrained minimization of (95) into an unconstrained one. For this purpose, we define the Lagrangian function

$$L(x, \tilde{x}, \Omega) = \int_0^T dt \sum_i \tilde{x}_i(\dot{x}_i - F_i(x, \Omega)) + J(x, \Omega) \tag{96}$$

We next use that the Lagrange function admits a stationary point at the solution.

Taking the functional derivatives of L with respect to x_i, $\tilde{x}_i$, and Ω_k (k labels the different components of the control fields) and performing integration by parts for the term $\tilde{x}_i \dot{x}_i$, we arrive at

$$\dot{x}_i = F_i(x, \Omega), \quad \dot{\tilde{x}}_i = -\frac{\partial F_i(x, \Omega)}{\partial x_i} - \frac{\partial \tilde{J}(x, t)}{\partial x_i}, \quad \Omega_k = \frac{1}{\gamma} \sum_i \tilde{x}_i \frac{\partial F_i(x, \Omega)}{\partial \Omega_k} \tag{97}$$

together with the intial $x_i(0) = x_{i0}$ and terminal $\tilde{x}_i(T) = -\partial J_T(x, \Omega)/\partial x_i$ conditions. For the "optimal control," this set of equations has to be fulfilled simultaneously for x, $\tilde{x}$, and Ω. In general, analytic solutions can be only found for highly simplified systems [228], whereas numerical calculation schemes have to be adopted for more realistic systems. A numerical algorithm for the solution of the optimality system (97) was formulated in Borzì et al. [220]. Suppose that we have an intial guess for the control fields Ω. We can then solve the dynamic equations for x subject to the initial conditions $x(0) = x_0$ *forward* in time, and the dynamic equations for $\tilde{x}$ subject to the terminal conditions $\tilde{x}_i(T) = -\partial J_T(x, \Omega)/\partial x_i$ *backward* in time. The last equation in (97) then provides the search directions $d_{\Omega_k} = -(1/\gamma)\sum_i \tilde{x}_i \partial F_i(x, \Omega)/\partial \Omega_k$ for the new control fields $\Omega_k \to \Omega_k + \lambda d_{\Omega_k}$, where for sufficiently small λ it is guaranteed that the new cost functional $J(x, \Omega)$ decreases [220]. Figure 24 sketches the basic ingredients of the resulting algorithm.

6.2.1. Adiabatic Passage

As a first example, we shall revise the adiabatic passage scheme of Section 6.1 within the framework of optimal control. This can be done in absence of decoherence analytically [228, 229] or, as we shall do in the following, in presence of decoherence numerically [220]. Our starting point is given by the master equation (67) of Lindblad form, where the effective Hamiltonian H_{eff} is given by (92). Because we are aiming at solutions that minimize environment losses, we neglect in Eq. (67) the in-scattering contributions and are left with

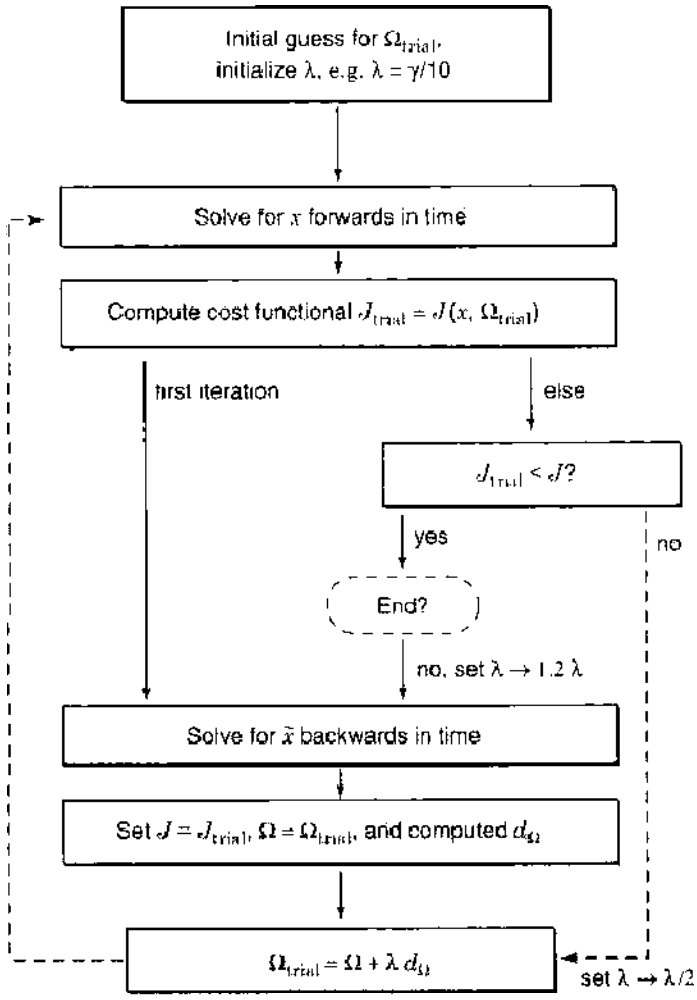


Figure 24. Schematic sketch of the numerical algorithm for optimal control [220]. One starts with a guess Ω_{trial} for the control fields and sets the initial stepsize λ to, for example, $\gamma/10$. Then the equations of motion for x are solved forward in time and the cost functional J_{trial} is computed. In the first iteration, Ω_{trial} is accepted. Upon acceptance, the equations of motion for the dual variables $\tilde{x}$ are solved backward in time, one sets $J = J_{trial}$ and $\Omega = \Omega_{trial}$, and finally the new search directions d_Ω are computed. Finally, a new Ω_{trial} is computed, and the loop is started again. In the ensuing iterations, Ω_{trial} is only accepted if $J_{trial} < J$, and otherwise the step-size λ is decreased and the loop restarted (i.e., linesearch with Armijo back-tracking). Through the procedure of increasing and decreasing λ, it is guaranteed that the algorithm always finds an appropriate step-size. Finally, the algorithm comes to an end after a certain number of iterations or when x has come close enough to the desired state x_f.

the Schrödinger-like equation of motion $i\dot{\psi} = H_{\text{eff}}\psi$ for the three-component wavefunction ψ. The effective Hamiltonian reads

$$H_{\text{eff}} = \left(\Delta - i\frac{\Gamma}{2}\right)|2\rangle\langle 2| - \frac{1}{2}\sum_{l=p,s}\left(M_l^{\dagger}\Omega_l + M_l\Omega_l^{*}\right) \tag{98}$$

with the optical transition matrix elements $M_p = |0\rangle\langle 2|$ and $M_s = |1\rangle\langle 2|$ for the pump and Stokes pulse, respectively. We assume that at time zero the system is in state ψ_0. The objective of the control is expressed through

$$J(\psi, \Omega) = \frac{1}{2}(1 - |\psi_1(T)|^2) + \frac{\gamma}{2}\int_0^T dt \sum_{l=p,s}|\Omega_l(t)|^2 \tag{99}$$

The first term has its minimum zero for $|\psi_1(T)| = 1$, that is, when the system is finally in state $|1\rangle$, and the second term accounts for the limited laser resources and is needed to make the optimal-control problem well posed (we set $\gamma = 10^{-8}$). For this system (as well as

for its Lindblad generalization [230]), we can easily obtain the optimality system in complex form [220]

$$i\dot{\psi} = H_{\text{eff}}\,\psi\,, \quad i\dot{\tilde{\psi}} = H_{\text{eff}}^{+}\,\tilde{\psi}\,, \quad \Omega_i = -\frac{1}{2\gamma}(\langle\psi|M_i^{\dagger}|\tilde{\psi}\rangle + \langle\tilde{\psi}|M_i^{\dagger}|\psi\rangle) \tag{100}$$

subject to the initial $\psi(0) = \psi_0$ and terminal $\tilde{\psi}_i(T) = -i\partial J_f(\psi)/\partial\psi_i = -i\delta_{i1}\psi_1(T)$ conditions. The last term in Eq. (100) is written in the usual bra and ket shorthand notation. Figure 25 shows results of prototypical optimal-control calculations and for different initial control fields Ω_{trial}. We start with pump and Stokes pulses of Gaussian form centered at time 30 and with a full-width of half maximum of 15, and use different pulse areas of 10, 25, and 50. For the two fields of lowest area, the optimal control loop (Fig. 24) comes up with control fields where in addition to the initial Gaussians components at the detuning frequency Δ are present. Reminiscent of the stimulated Raman adiabatic scheme of Section 6.1, at this frequency the Stokes pulse is turned on prior to the pump one. In contrast to that, for the highest initial pulse areas, the pump and Stokes fields keep their Gaussian shape (and no additional frequency components show up). The resulting control strategy is similar to that of the black regions in Fig. 23b at zero time delay. Thus, different initial control fields Ω_{trial} in the optimal control algorithm lead to different control strategies.

This is not particularly surprising, as even for the relatively simple situation shown in Fig. 23, where the pump and Stokes pulse are characterized by the two parameters of area and pulse delay, a huge number of successful control strategies (indicated by the black regions) is found. In the optimal control case shown in Fig. 25, where the control fields Ω are discretized at about 10,000 points in time, the control space is tremendously increased, and correspondingly an even much larger number of possible solutions can be expected. In absence of decoherence (i.e., for an isolated few-level system), it can be shown that the only allowed extrema of the Lagrange function (96) correspond to perfect control or no control [127]. In presence of decoherence, things are modified. This is shown in Fig. 25 and even more clearly in Fig. 26, where one observes that the deviation of the final state $\psi(T)$ from the desired state ψ_1 differs for different initial conditions. In other words, starting at one point of the high-dimensional control space, the search algorithm of Fig. 24 proceeds along the direction of the steepest descent and becomes trapped in a suboptimal local minimum. The way the search algorithm evolves can be studied in Fig. 26. For the small pulse areas (dotted and dashed lines), one observes that initially no population is transfered (and J has its largest possible value of 0.5), whereas for the largest pulse area (solid line), the initial guess for

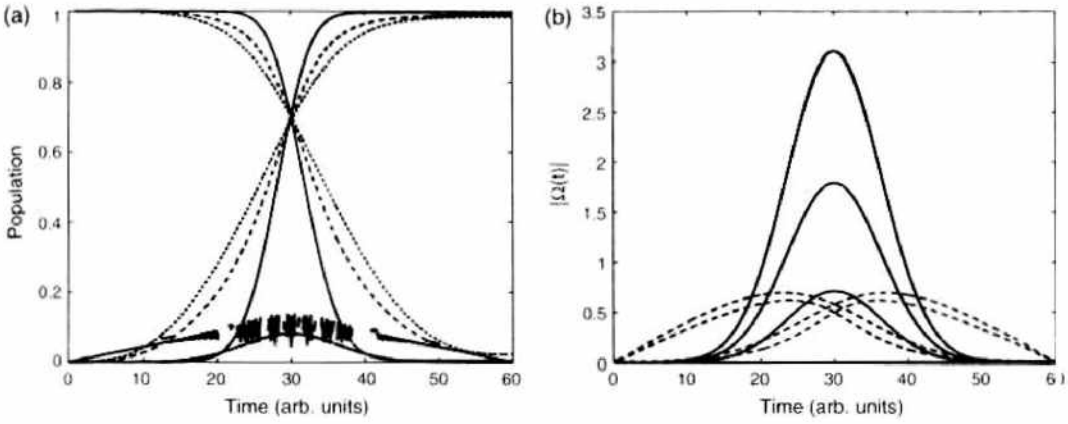


Figure 25. Results of optimal control calculations for the adiabatic population transfer. We assume that the system is initially in state $|1\rangle$. The objective of the optimal control is to maximize the population of $|1\rangle$ at time $T = 60$. As an initial guess for the pump and Stokes laser pulses, we assume Gaussians centered at time 30 with a full-width at half maximum of 15 and areas of 10, 25, and 50. We use $\Delta = 20$ and $\Gamma = 0.1$. Panel (a) shows the population transients [see Fig. 23a] for the initial areas of 10 (dotted lines), 25 (dashed lines), and 50 (solid lines). Panel (b) shows the corresponding optimal control fields; the solid and dashed lines, respectively, correspond to frequency components centered around zero and Δ. The zero-frequency components are attributed in order of increasing magnitude to the initial areas of 10, 25, and 50.

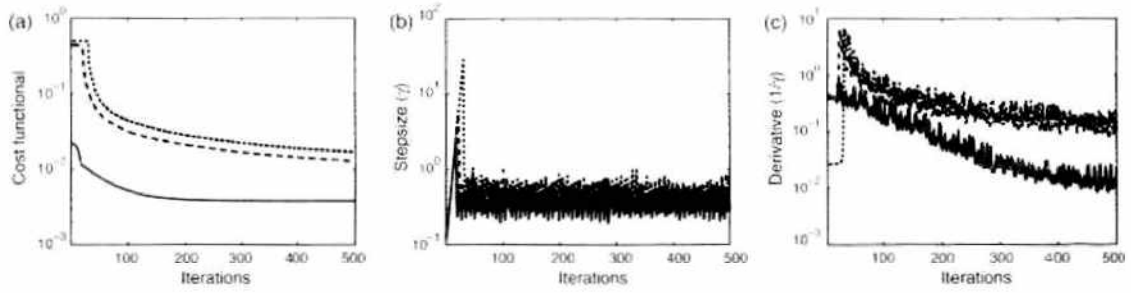


Figure 26. Details of the optimal control calculations shown in Fig. 25 for initial pulse areas of 10 (dotted lines), 25 (dashed lines), and 50 (solid lines). Panel (a) shows the decrease of the cost functional $J(\psi, \Omega)$ of Eq. (99) as a function of the number of iterations, panel (b) the step-size λ chosen in the optimal control algorithm, and panel (c) the derivative $\|d_u\|$ (which should vanish at the extremum).

Ω_{trial} already leads to a quite efficient transfer. Within the first few iterations of the optimal control loop, J does not decrease significantly. This is because the initial stepsize λ is too small [panel (b)] to cause a significant decrease of J. After approximately 10 iterations, an adequate step-size is obtained (through the increase of $\lambda \rightarrow 1.2\,\lambda$ after successful steps, see Fig. 24), and J decreases rapidly over approximately 100 iterations. This regime is followed by one with a much slower decrease of J, where the control approaches slowly that associated to the local minimum. This discussion allows us to pinpoint the respective advantages and disadvantages of the optimal control algorithm. If we start with a reasonable guess for Ω_{trial} (whose solutions can be even completely away from the desired ones, such as for the smaller pulse areas), we obtain a strongly improved Ω whose solutions fulfill the objective of the control much better than that of Ω_{trial}. On the other hand, the solutions Ω of the control algorithm (Fig. 24) are most probably associated to local minima of the control space rather than to the *global* minimum.

6.2.2. Phonon-Assisted Dephasing

The adiabatic passage scheme that we have just discussed is an extreme example in the sense that there exist numerous solutions—as indicated in Fig. 23b—and one expects that the search algorithm becomes quickly trapped in a suboptimal, local minimum. As another example, we shall now discuss the coherent control of the two-level system in presence of phonon couplings. In Section 5.5, we showed for the spin-boson model that in presence of an exciting laser pulse Rabi flopping occurs but is dampened because of phonon decoherence. However, contrary to other decoherence channels in solids where the system's wavefunction acquires an uncontrollable phase through environment coupling, in the independent Boson model the loss of phase coherence is due to the coupling of the electron–hole state to an ensemble of harmonic oscillators that all evolve with a coherent time evolution but different phase. This results in destructive interference and dephasing and thus spoils the direct applicability of coherent carrier control. On the other hand, the coherent nature of the state-vector evolution suggests that more refined control strategies might allow to suppress dephasing losses. To address the question whether such losses are inherent to the system under investigation, in the following we examine phonon-assisted dephasing within the optimal-control framework aiming at a most efficient control strategy to channel the system's wavefunction through a sequence of given states. We quantify the objective of the control through the cost function

$$J(\boldsymbol{u}, \Omega) = \frac{1}{2}\left(\int_{-T}^{T} dt\, \beta(t) |\boldsymbol{u}(t) - \hat{\boldsymbol{e}}_3|^2 + |\boldsymbol{u}(T) + \hat{\boldsymbol{e}}_3|^2 + \gamma \int_{t}^{T} dt\, |\Omega(t)|^2 \right) \tag{101}$$

with β a Gaussian centered at time zero with a narrow full-width of half maximum of $0.1\,\omega_t^{-1}$ and $\gamma = 10^{-5}$ a small constant. In other words, we are seeking for solutions where $\boldsymbol{u}$ passes through the excited state $\hat{\boldsymbol{e}}_3$ at time zero and goes back to the ground state $-\hat{\boldsymbol{e}}_3$ at T. For the system dynamics, we assume the equations of motion (91) for the cumulants $\boldsymbol{u}$, s_t, and $\boldsymbol{u}_t$, subject to the initial conditions $\boldsymbol{u}(-T) = -\hat{\boldsymbol{e}}_3$, $s_t(-T) = 0$, and $\boldsymbol{u}_t(-T) = 0$. Again, the

method of Lagrange multipliers is used to minimize $J(\boldsymbol{u}, \Omega)$ subject to Eq. (91), and to obtain the adjoint equations [183]

$$\begin{aligned}
\dot{\tilde{\boldsymbol{u}}} &= \boldsymbol{\Omega} \times \tilde{\boldsymbol{u}} + \sum_i g_i \left[\left(n_i + \frac{1}{2} \right) \hat{\boldsymbol{e}}_3 \times \Re e(\tilde{\boldsymbol{u}}_i) + \frac{1}{2} \Im m\big((\tilde{s}_i - \boldsymbol{u}\tilde{u}_i) \hat{\boldsymbol{e}}_3 - u_3 \tilde{\boldsymbol{u}}_i \big) \right] + \beta(t)(\boldsymbol{u} - \hat{\boldsymbol{e}}_3) \\
\dot{\tilde{s}}_i &= -i\omega_i \tilde{s}_i + 2g_i \left[(\tilde{\boldsymbol{u}} \times \boldsymbol{u}) \hat{\boldsymbol{e}}_3 + \Re e \sum_j (\tilde{\boldsymbol{u}}_j^* \times \boldsymbol{u}_j) \hat{\boldsymbol{e}}_3 \right] \\
\dot{\tilde{\boldsymbol{u}}}_i &= \boldsymbol{\Omega} \times \tilde{\boldsymbol{u}}_i - i\omega_i \tilde{\boldsymbol{u}}_i + 2g_i \hat{\boldsymbol{e}}_3 \times \tilde{\boldsymbol{u}}
\end{aligned} \tag{102}$$

with terminal conditions $\tilde{\boldsymbol{u}}(T) = -\hat{\boldsymbol{e}}_3 - \boldsymbol{u}(T)$, $\tilde{s}_i(T) = 0$, and $\tilde{\boldsymbol{u}}_i(T) = 0$. Equations (91) and (102) together with

$$\Omega = \frac{1}{\gamma} \left[(\tilde{\boldsymbol{u}} \times \boldsymbol{u}) + \Re e \sum_i (\tilde{\boldsymbol{u}}_i^* \times \boldsymbol{u}_i) \right] (\hat{\boldsymbol{e}}_1 - i\hat{\boldsymbol{e}}_2) \tag{103}$$

form the optimality system. It is solved iteratively through the scheme depicted in Fig. 24, with integration of Eq. (91) forward and Eq. (102) backward in time, and computing an improved control by use of Eq. (103). Results of such optimal-control calculations are shown in Fig. 27. Most remarkably, one can indeed obtain a control field for which $\boldsymbol{u}(t)$ passes through the desired states of $\hat{\boldsymbol{e}}_3$ at time zero and $-\hat{\boldsymbol{e}}_3$ at T. Thus, appropriate pulse shaping allows to control fully the two-level system even in presence of phonon couplings. We emphasize that, with the exception of the somewhat pathological quantum "bang-bang" control [128], where the system is constantly flipped to suppress decoherence, no such simple control strategy for suppression of environment losses is known in the literature. This result, which also prevails in presence of finite but low temperatures [183], clearly highlights the strength and flexibility of optimal control.

6.2.3. Genetic Algorithms

In several cases of interest, one is often neither able to solve the equations of motion nor know the full Hamiltonian characterizing the system. This holds in particular true for laser-induced reactions of molecules [203], where the configuration landscape is highly complicated. Judson and Rabitz [231] were the first to propose an evolutionary algorithm that allows the search for control fields even without any knowledge of the Hamiltonian. For the sake of completeness, in the following we shall briefly outline the main ideas of this approach. Suppose that the laser pulse can be encoded in terms of a "gene," that is, a

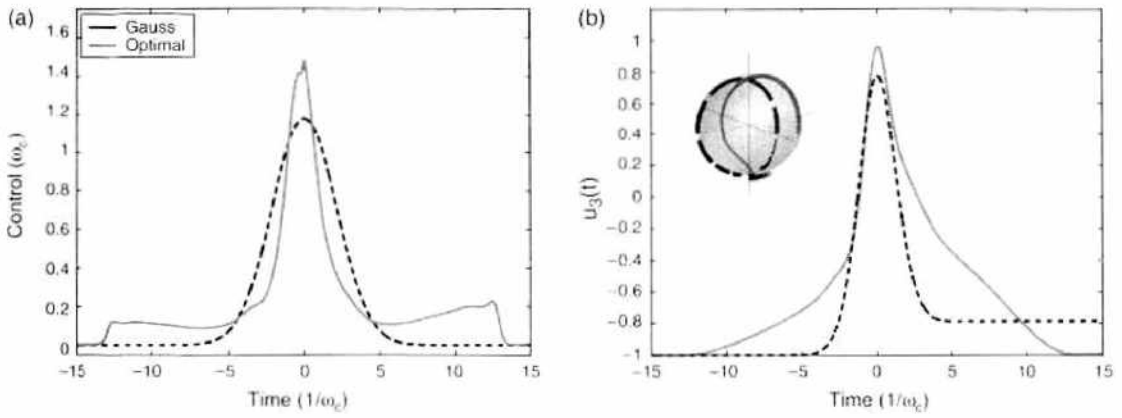


Figure 27. Results of our calculations with a Gaussian 2π (dashed lines) and optimal-control (solid lines) laser pulse and for zero temperature and an electron–phonon coupling of $\alpha_p = 0.1$. Panel (a) shows $|\Omega(t)|$ and panel (b) the time evolution of $u_3(t)$ and the insets the trajectories of the Bloch vector $\boldsymbol{u}(t)$. For the Gaussian 2π-pulse, Rabi flopping occurs but is dampened due to electron–phonon interactions. For the optimal control, decoherence losses are completely suppressed, and the the system passes through the desired states of $\hat{\boldsymbol{e}}_3$ at time zero and $-\hat{\boldsymbol{e}}_3$ at T.

vector $\Omega = (\Omega_1, \Omega_2, \ldots \Omega_n)$ with typically $n \sim 10$ to 100 components. Within the evolutionary approach, a population $N_{\text{pop}} \cong 48$ of different genes Ω^μ is considered. At the beginning, the different components of each gene are chosen randomly. Next, we compute for each gene the objective function J^μ. Within the evolutionary approach, the next population of genes is determined in biological terms according to the fitness J^μ of each individual [221, 231]. Following Zeidler et al. [221], this can be accomplished as follows: first, the individual with the best J^μ is included without change in the next generation (*elitism*); next, the $N_{\text{parent}} \sim N_{\text{pop}}/7$ individuals with the best J^μ are chosen as parents for the next generation; about $2\,N_{\text{parent}}$ of the individuals of the next generation are determined by randomly choosing two parents, cutting their genes at one given point (*single-point crossover*) or two points (*two-point crossover*), and pasting the different pieces of the genes together (*recombination*); finally, the remaining individuals of the next generation are obtained by randomly taking one parent and randomly modifying its genetic information Ω_i^μ (*mutation*).

There exist numerous other implementations that differ in one or several points. However, the grand idea of all these approaches is to provide a sufficiently large gene pool and to let the individuals benefit from their respective advantages and peculiarities. This has the consequence that once an individual acquires a successful control strategy (either through recombination or mutation), it will distribute it to the next generation where it possibly becomes optimized through further mixing or mutation. For that, it is compulsory to keep not only the fittest individual but to provide a larger gene pool. In this respect, mutation plays an important role as it determines the degree of modification from one generation to the next. For normalized control-field components $\Omega_i^\mu \in [0, 1]$ mutation can be computed according to [221, 232]

$$\Omega_i^{\mu'} = \Omega_i^\mu - \sigma \log\left[\tan\left(\frac{\pi}{2}\, r\right)\right] \tag{104}$$

where $r \in (0, 1)$ is a uniformly distributed random number (and $\Omega_i^{\mu'} \in [0, 1]$ has to be asserted). σ is the quantity that determines how much $\Omega^{\mu'}$ can deviate from Ω^μ. It has a similar role as the step-size λ in the optimal control scheme, and its value should be adapted during optimization. Here, one can proceed as follows [221]: let N_{mut} be the number of mutated individuals and N_{succ} the number of successful mutations with $J^{\mu'} < J^\mu$; for $N_{\text{succ}} < 0.2\,N_{\text{perm}}$, we conclude that the mutation rate is too high and set $\sigma \to 0.9\,\sigma$; otherwise we increase the rate according to $\sigma \to \sigma/0.9$. Quite generally, the genetic algorithm works formidably well for laser fields that can be by characterized by a few parameters (e.g., laser-pulse shaping experiments [203, 221]). For instance, parameterizing the pump and Stokes pulses of the adiabatic passage scheme by Gaussians (i.e., in terms of areas, detunings, time delay, and full-width at half maxima), a highly successful transfer scheme is found after a few generations. In comparison to the optimal control approach, the evolutionary one examines a larger portion of the control space (through its different individuals), and therefore chooses out of several local minima the lowest one. On the other hand, evolutionary approaches are usually much slower [no information about the steepest descent $\nabla_\Omega J(x, \Omega)$ is used] and have huge problems to find nontrivial pulse shapes, such as the one depicted in Fig. 27a.

6.3. Self-Induced Transparency

We conclude this section with an at-first-sight somewhat different topic, namely laser pulse propagation in a macroscopic sample of inhomogenously broadened quantum dots. Let the central frequency of the laser pulse be tuned to the exciton groundstate transition (see Fig. 13). We then describe the dot states in terms of generic two-level systems with different detunings Δ. When the light pulse enters the dot region, it excites excitons and hereby suffers attenuation. It should, however, be emphasized that inhomogeneous line broadening leads to losses which are substantially different from those induced by homogeneous broadening [233]. Through the pulse propagation in the medium of inhomogenously broadened dots, all of them are excited in *in phase*, where—at variance with homogeneous broadening—each dot has a coherent time evolution. However, the phase varies from dot to dot, thus

leading to interference effects that in most cases prevent the observation of the coherent radiation-matter interaction. A striking exception is the phenomenon of *self-induced transparency* [22, 234–238], a highly nonlinear optical coherence phenomenon that directly exploits inhomogeneous level broadening. Light-matter coupling plays a crucial role in its theoretical analysis. Not only one has to consider the material response in the presence of the driving light pulse, but also the back-action of the macroscopic material polarization on the light propagation (through Maxwell's equations). For the inhomogenously broadened two-level systems, we assume a time evolution according to the master equation (67) of Lindblad form, where the coherent part is given by the usual Bloch equations (59) for different detunings Δ. For the light pulse, we assume a geometry (Fig. 28) where the laser enters from the left-hand side into the sample of inhomogenously broadened dots. Denoting the pulse propagation direction with z and assuming an electric-field profile with envelope $\mathcal{E}_0(z, t)$ and a central frequency of ω_0 [239], we describe the light propagation in the slowly varying envelope approximation [22]

$$\left(\partial_z + \frac{n}{c}\partial_t\right)\mathcal{E}_0(z, t) \cong -\frac{2\pi\omega_0}{nc}\,\Im m\mathcal{P}(z, t) \tag{105}$$

Here, $\mathcal{P}(z) = M_{0x}\hat{e}_\lambda N \int g(\Delta)d\Delta\, 1/2[u_1(z, \Delta) - i\,u_2(z, \Delta)]$ is the material polarization, with M_{0x} the excitonic dipole moment (assumed to not depend on Δ), $\hat{e}_\lambda$ the exciton polarization, N the dot density, $g(\Delta)$ the inhomogenous broadening, and $u_1(z, \Delta)$ and $u_2(z, \Delta)$ the real and imaginary part of the Bloch vector, respectively, at position z and for a detuning Δ. For a coherent time evolution, there exists a remarkable theorem that asserts that the pulse area, defined through

$$A(z) = M_{0x}\lim_{t\to\infty}\int_{-\infty}^{t} dt'\,\hat{e}_\lambda^*\,\mathcal{E}(z, t') \tag{106}$$

satisfies the equation [22, 235]

$$\frac{dA(z)}{dz} = -\frac{\alpha}{2}\sin A(z), \quad \alpha = \frac{2\pi^2 N\omega_0 M_{0x}^2}{nc}\,g(0) \tag{107}$$

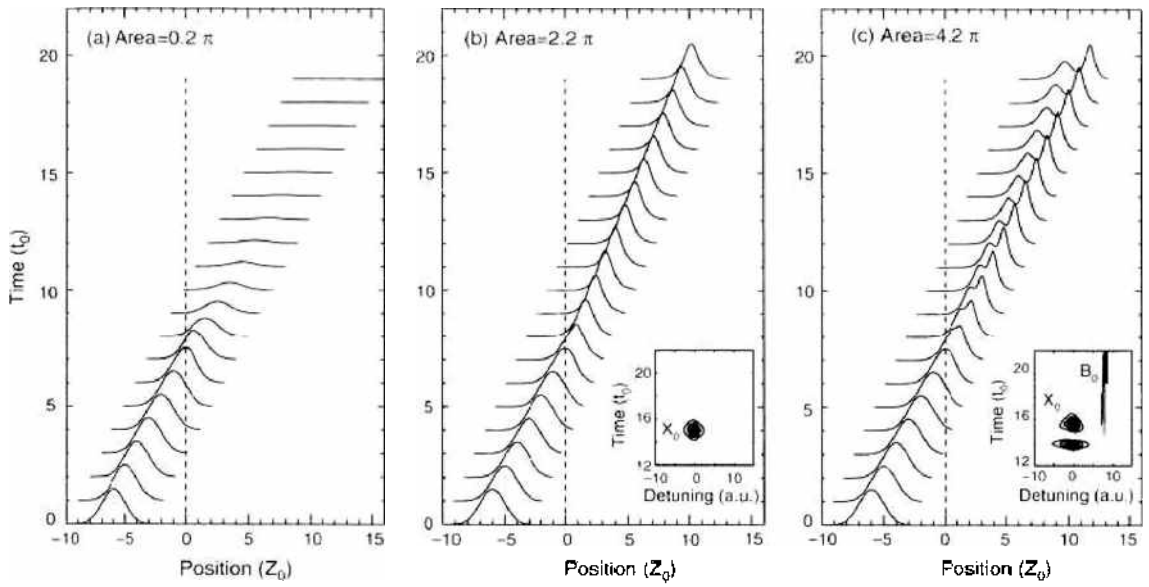


Figure 28. Results of our simulations for $\mathcal{E}_0(z, t)$ of pulse propagation in a sample of inhomogenously broadened quantum dots and for different pulse areas; we assume a setup where the pulse enters from a dot-free region (negative z-values) into the dot region. Length is measured in units of $z_0 = 1/\alpha$, time in units of $t_0 = nz_0/c$, and energy in units of $E_0 = 1/t_0$, with $z_0 \sim 250$ μm, $t_0 \sim 3$ ps, and $E_0 \sim 0.2$ meV for typical InGaAs dot samples. The insets report contour plots of the time evolution of the exciton and biexciton population at position $z = 5$ [225, 239, 240].

Here, α provides a characteristic length scale. For a weak incident pulse, one immediately observes from the linearized form of Eq. (107) that A decays according to $\exp -\alpha z/2$, as expected from Beer's law of linear absorption. Within the Bloch vector picture, this decay is due to the small rotations of Bloch vectors out of their equilibrium positions and the resulting intensity loss of the light pulse. However, completely new features appear when $A \geq \pi$. Most importantly, if A is an integer of π, the pulse area suffers no attenuation in propagating along z. Indeed, such behavior is observed in Fig. 28, which shows results of simulations for the more complete level scheme depicted in Fig. 22c: for small field strengths, Fig. 28a, the pulse becomes attenuated quickly. However, if the pulse area exceeds a certain value, Fig. 28b, self-modulation occurs and the pulse propagates without suffering significant losses. In a sense, this situation resembles a material control of the laser pulse. The latter acquires a 2π hyperbolic secant shape [22, 235], which—contrary to the situation depicted in Fig. 16 for a constant laser—rotates all Bloch vectors from their initial state through a sequence of excited states back to the initial ones, irrespective of their detuning Δ (inset of Fig. 28b). Here, the leading edge of the pulse coherently drives the system in a predominantly inverted state; before decoherence takes place, the trailing edge brings then the population back to the groundstate by means of stimulated emission, and an equilibrium condition is reached in which the pulse receives through induced emission of the system the same amount of energy transferred to the sample through induced absorption. Finally, at the highest pulse area, Fig. 28c, we observe pulse breakup [22, 234, 235]. The inset shows a 4π-rotation of the exciton states and an additional population of the biexciton ones. As apparent from the figure, this additional biexciton channel does not spoil the general pulse propagation properties (for details, see Ref. [240]). Self-induced transparency in semiconductor quantum dots has been demonstrated recently [132, 241]. No pulse breakup was observed, a finding attributed to a possible dependence of the dipole moments $M_{0i}(L)$ on the quantum dot size L.

7. QUANTUM COMPUTATION

Quantum computation is a quantum control with unprecedented precision [19, 18]. Its key elements are the quantum bit, or *qubit*, which is a generic two-level system, and a register of such qubits (with a typical size ranging from a few tens to several hundreds). This register allows to store the quantum information, which is processed by means of unitary transformations (*quantum gates*) through an external control. Besides the single-qubit rotations (*unconditional gates*), one also requires two-qubit rotations (*conditional gates*) where the "target qubit" is only rotated when the "control qubit" is in an inverted state. Through the latter transformations, it becomes possible to create entanglement, which is at the heart of quantum computation. In his seminal work, Shor [242] showed that such quantum computation could—if implemented successfully—eventually outperform classical computation. Yet, the hardware requirements and the degree of controllability are tremendous (error-correction schemes permit only one error in approximately 10^4 operations [243]), and it is completely unclear whether a quantum computer will ever be built. Despite this very unclear situation, recent years have seen huge efforts in identifying possible candidates for quantum computers and performing proof-of-principle experiments. Quite generally, any few-level system with sufficiently long-lived states, that allows for efficient readout and scalability, can serve as a possible candidate [244]. Among the vast amount of work devoted to the implementation of quantum-information processing in physical systems, several have been concerned with optical and spin excitations in quantum dots. In the following, we shall briefly review some of the key proposals and experiments.

There are a number of crucial elements to be met in any implementation of a quantum computer, among which the most important ones are the identification of qubits with long decoherence times, of a coupling mechanism between different qubits (for performing conditional gates), and of a readout capability for the quantum information. In the field of quantum dots, optical and spin excitations have been considered as qubits. With the limits discussed in the previous sections, long decoherence times, efficient control, and reliable readout schemes are available. The strategies for coupling qubits are motivated by related

schemes in different fields of research, which either exploit some local nearest-neighbor interactions [245], (e.g., hyperfine interactions in nuclear magnetic resonance [246]), or rely on a common "bus" that connects all qubits (e.g., phonon excitations of a linear chain of ions [247]). For quantum dots, Barenco et al. [248] were the first to propose the quantum-confined Stark effect as a means to couple optical excitations in different dots. This proposal was elaborated by Troiani et al. [226] and Biolatti et al. [249, 250], who proposed to use the Coulomb renormalizations of few-particle states for an efficient inter-qubit coupling. Let us briefly address the first proposal [226] at the example of the level scheme depicted in Fig. 22c.

We denote the groundstate with $|0\rangle|0\rangle$, where the first and second expression account, respectively, for the (missing) exciton with spin-up and spin-down orientation; within this qubit language, the single-exciton states correspond to $|1\rangle|0\rangle$ and $|0\rangle|1\rangle$ and the biexciton state to $|1\rangle|1\rangle$. Because of the polarization selection rules (Section 3.1.3) and the Coulomb renormalization Δ of the biexciton, all these states can be addressed individually by means of coherence spectroscopy. Indeed, optical control of these two exciton-based qubits was demonstrated [93, 95, 251]. To allow within this framework for scalability, optical excitations in an array of quantum dots were proposed; enhancement of the Coulomb couplings between different dots could be either achieved by relying on the quantum-confined Stark effect [249, 250]—in an electric field electron and hole wavefunctions become spatially separated, and in turn the dipole–dipole interaction between excitons in different dots is strongly enhanced—or on intrinsic exciton–exciton couplings [252, 253]. Other work has proposed Förster-type processes where optical excitations are near-field coupled [254–256].

Qubits based on optical excitations have the glaring shortcoming of a fast decoherence on the sub-nanosecond timescale. Much longer decoherence times are expected for spin excitations. Loss and DiVincenzo [257] proposed a quantum computation scheme based on spin states of coupled single-electron doped quantum dots, with electrical gating as a means for the unconditional and conditional operations. A mixed approach was put forward by Imamoglu et al. [223], where the quantum information is encoded in the spin degrees of freedom, and coupling to the optical degrees is used for efficient and fast quantum gates. Within this proposal, the quantum computer is realized through single-electron charged quantum dots. The unconditional gates are performed through optical coupling of the different electron-spin states to the charged-exciton state (Voigt geometry, Section 3.1.3), and the conditional ones by means of cavity quantum electrodynamics where all quantum dots are located in a microcavity and coupled to a common cavity mode [223, 258, 259]. Differently, Piermarocchi et al. [260] proposed a coupling via virtual excitations of delocalized excitons as a genuine solid-state coupling mechanism between electron-spins in different quantum dots. There are a number of further proposals for quantum computation with spin memory and optical gating, where either the quantum-confined Stark effect [224, 261] or the enhanced flexibility of molecular states in artificial molecules [213, 262] is used for switchable qubit–qubit interactions. In addition, some work has been concerned with strategies for sophisticated optical gating (e.g. based on pulse shaping [216, 217] or spin-flip Raman transitions [218]). As regarding stimulated Raman adiabatic passage (Section 6.1), a slightly modified level scheme and a somewhat different control strategy is required for qubit rotations or entanglement creation by means of adiabatic population transfers [263–265]. Corresponding quantum-dot implementations were proposed for unconditional and conditional gates [213], and for storage qubits [227]. Finally, in the context of molecular systems, the applicability of optimal control for quantum gates was shown to be feasible [266].

ACKNOWLEDGMENTS

During the past years, I had the opportunity to collaborate with many people who have influenced my way of thinking and have substantially contributed to the results presented in this overview. I am indebted to all of them. My deepest gratitude goes to Giovanna Panzarini (1968—2001), who has introduced me to the field of quantum optics and has shared with me her physical intuition and her joy for the beauty of physics. Her memory will remain alive. Elisa Molinari is gratefully acknowledged for generous support, most helpful discussions, and for providing a lively and stimulating atmosphere in the Modena group. I sincerely thank

Filippo Troiani for his pioneering contributions in our collaboration on quantum-dot–based quantum computation. He, as well as Guido Goldoni, Costas Simserides, Claudia Sifel, Pekka Koskinen, Alfio Borzí, Georg Stadler, and Jaro Fabian I wish to thank for many fruitful discussions and the pleasure of collaborating during the past years.

APPENDIX A: RIGID EXCITON AND BIEXCITON APPROXIMATION

Excitons

Consider the trial exciton wavefunction

$$\Psi^{\mathrm{I}}(\boldsymbol{r}_e, \boldsymbol{r}_h) = \Phi(\boldsymbol{R})\,\phi_0(\boldsymbol{\rho}) \tag{A.1}$$

which consists of the groundstate exciton wavefunction $\phi_0(\boldsymbol{\rho})$ of an ideal quantum well and an envelope function $\Phi(\boldsymbol{R})$. In other words, we assume that in presence of a quantum confinement, the electron and hole are Coulomb bound in the same way as they would be in an ideal quantum well, and only the center-of-mass motion is affected by the quantum confinement. In Eq. (A.1) the center-of-mass and relative coordinates are given by the usual expressions $\boldsymbol{R} = (m_e\boldsymbol{r}_e + m_h\boldsymbol{r}_h)/M$ and $\boldsymbol{\rho} = \boldsymbol{r}_e - \boldsymbol{r}_h$, respectively. We next insert the trial wavefunction (A.1) into the Schrödinger equation (15) and obtain

$$\left[\mathscr{H} + h + \sum_{i=e,h} U_i(\boldsymbol{r}_i)\right]\Phi(\boldsymbol{R})\,\Phi_0(\boldsymbol{\rho}) = E\,\Phi(\boldsymbol{R})\,\phi_0(\boldsymbol{\rho}) \tag{A.2}$$

with $\mathscr{H}$ and h defined in Eq. (12). The left-hand side can be simplified by using $h\phi_0(\boldsymbol{\rho}) = \epsilon_0\phi_0(\boldsymbol{\rho})$. Multiplying Eq. (A.2) with $\delta(\boldsymbol{R} - \boldsymbol{R}_x)\,\phi_0(\boldsymbol{\rho})$ and integrating over the entire phase space $\boldsymbol{\tau}$ finally gives

$$\left[-\frac{\nabla^2_{R_x}}{2M} + \sum_{i=e,h}\int d\boldsymbol{\tau}\,\delta(\boldsymbol{R} - \boldsymbol{R}_x)\,U_i(\boldsymbol{r}_i)\,|\phi_0(\boldsymbol{\rho})|^2\right]\Phi(\boldsymbol{R}_x) = \mathscr{E}\,\Phi(\boldsymbol{R}_x) \tag{A.3}$$

Comparing this expression with Eq. (16) shows that the term on the left-hand side is identical to the averaged potential $\overline{U}(\boldsymbol{R}_x)$.

Biexcitons

A similar procedure can be applied for biexcitons. In analogy to Eq. (A.1), we make the *ansatz*

$$\overline{\Psi}(\bar{\boldsymbol{\tau}}) = \overline{\Phi}(\boldsymbol{R})\,\bar{\phi}_0(\bar{\boldsymbol{\tau}}) \tag{A.4}$$

with $\bar{\phi}_0$ the variational function (23) and $\overline{\Phi}(\boldsymbol{R})$ the corresponding envelope function which depends on the center-of-mass coordinate $\boldsymbol{R} = m_e(\boldsymbol{r}_e + \boldsymbol{r}_{e'})/M + m_h(\boldsymbol{r}_h + \boldsymbol{r}_{h'})/M$ with $M = 2(m_e + m_h)$; finally $\bar{\boldsymbol{\tau}}$ denotes the set of variables $\boldsymbol{r}_e$, $\boldsymbol{r}_h$, $\boldsymbol{r}_{e'}$, and $\boldsymbol{r}_{h'}$. Suppose that the Hamiltonian can be decomposed into the parts

$$H = -\frac{\nabla^2_R}{2M} + h + \sum_i U_i(\boldsymbol{r}_i) \tag{A.5}$$

where in analogy to excitons, $h\bar{\phi}_0 = \bar{\epsilon}_0\bar{\phi}_0$ gives the energy of the quantum-well biexciton [44] and i runs over all electrons and holes. From the Schrödinger equation defined by Eq. (A.5) and (A.4) we then obtain after multiplication with $\delta(\boldsymbol{R} - \boldsymbol{R}_b)\,\bar{\phi}_0$ and integration over the entire phase space $\bar{\boldsymbol{\tau}}$ the final result

$$\left[-\frac{\nabla^2_{R_b}}{2M} + \sum_i\int d\bar{\boldsymbol{\tau}}\,\delta(\boldsymbol{R} - \boldsymbol{R}_b)\,U_i(\boldsymbol{r}_i)\,|\bar{\phi}_0(\bar{\boldsymbol{\tau}})|^2\right]\overline{\Phi}(\boldsymbol{R}_b) = \overline{\mathscr{E}}\,\overline{\Phi}(\boldsymbol{R}_b) \tag{A.6}$$

where the term on the left-hand side defines the effective confinement potential for biexcitons (see Fig. 5).

Optical Dipole Elements

Let us investigate the dependence of the dipole matrix elements (35) on the confinement length L for the exciton states under consideration. Because of the product-type exciton wavefunction (A.1), the optical dipole moment (35) is given by the spatial average of the envelope part $\Phi(\boldsymbol{R})$. We shall now show how this average depends on the confinement length L. Our starting point is given by the normalization condition $\int d\boldsymbol{R}\,|\Phi(\boldsymbol{R})|^2 = 1$. We next introduce the dimensionless space variable $\boldsymbol{\xi} = \boldsymbol{R}/L$, which is of the order one. Through $L^d \int d\boldsymbol{\xi}\,|\Phi(L\boldsymbol{\xi})| = \int d\boldsymbol{\xi}\,|\tilde{\Phi}(\boldsymbol{\xi})|^2 = 1$, we define the wavefunction $\tilde{\Phi}(\boldsymbol{\xi}) = L^{d/2}\Phi(L\boldsymbol{\xi})$, where $d = 2$ denotes the two-dimensional nature of the electron-hole states. Then,

$$M_{0\mathrm{X}} = \mu_{cv}\,\phi_{cv}(0)\int d(L\boldsymbol{\xi})\,L^{-d/2}\tilde{\Phi}(\boldsymbol{\xi}) = \mu_{cv}\,\phi_{cv}(0)L^{d/2}\int d\boldsymbol{\xi}\,\tilde{\Phi}(\boldsymbol{\xi}) \tag{A.7}$$

is the dipole moment for excitonic transitions in the weak confinement regime. Equation (A.7) is the result we were seeking. The integral on the right-hand side of the last expression is of the order unity. Thus, the oscillator strength for optical transitions scales with $|M_{0\mathrm{X}}|^2 \propto L^d$ (i.e., it is proportional to the area L^2 of the confinement potential).

APPENDIX B: CONFIGURATION INTERACTIONS

Second Quantization

Second quantization is a convenient tool for the description of few- and many-particle problems [23, 267–269]. The central objects are the field-operators $\psi^\dagger(\boldsymbol{r})$ and $\psi(\boldsymbol{r})$, which, respectively, describe the creation and destruction of an electron at position $\boldsymbol{r}$. The field operators obey the usual anticommutation relations $\{\psi(\boldsymbol{r}), \psi^\dagger(\boldsymbol{r}')\} = \delta(\boldsymbol{r}-\boldsymbol{r}')$ and zero otherwise. Within the framework of second quantization, one replaces all one- and two-particle operators $\mathcal{O}_1(\boldsymbol{r})$ and $\mathcal{O}_2(\boldsymbol{r}_1,\boldsymbol{r}_2)$ by [23, 268]

$$\mathcal{O}_1(\boldsymbol{r}) \longrightarrow \int d\boldsymbol{r}\psi^\dagger(\boldsymbol{r})\mathcal{O}_1(\boldsymbol{r})\psi(\boldsymbol{r}) \tag{A.8}$$

$$\mathcal{O}_2(\boldsymbol{r},\boldsymbol{r}') \longrightarrow \int d\boldsymbol{r}d\boldsymbol{r}'\psi^\dagger(\boldsymbol{r})\psi^\dagger(\boldsymbol{r}')\mathcal{O}_2(\boldsymbol{r},\boldsymbol{r}')\psi(\boldsymbol{r}')\psi(\boldsymbol{r}) \tag{A.9}$$

When a semiconductor is described in the envelope-function approximation, electrons and holes have to be treated as independent particles. This can be accomplished by introducing the field operators $\psi^e_\lambda(\boldsymbol{r})$ and $\psi^h_\lambda(\boldsymbol{r})$ accounting for the electron and hole degrees of freedom, where λ is the spin of the electron or hole (Section 3.1.3); below we shall denote the spin orientation orthogonal to λ with $\bar{\lambda}$. We find it convenient to expand $\psi^{e,h}(\boldsymbol{r})$ in the single-particle bases of Eq. (10),

$$\psi^e_\lambda(\boldsymbol{r}) = \sum_\mu \phi^e_\mu(\boldsymbol{r})c_{\mu\lambda}, \qquad \psi^h_\lambda(\boldsymbol{r}) = \sum_\nu \phi^h_\nu(\boldsymbol{r})d_{\nu\lambda} \tag{A.10}$$

where $c^\dagger_{\mu\lambda}$ creates an electron with spin orientation λ in the single-particle state μ, and $d^\dagger_{\nu\lambda}$ a hole with spin λ in state ν. With these field operators, we can express the few-particle Hamiltonian accounting for the propagation of electrons and holes in presence of the quantum confinement and mutual Coulomb interactions as [23, 269]

$$\begin{aligned} H = &\sum_{\mu\lambda}\epsilon^e_{\mu\lambda}\,c^\dagger_{\mu\lambda}c_{\mu\lambda} + \sum_{\nu\lambda}\epsilon^h_{\nu\lambda}\,d^\dagger_{\nu\lambda}c_{\nu\lambda} + \frac{1}{2}\sum_{\substack{\mu\mu'\bar{\mu}\bar{\mu}'\\ \lambda\lambda'}} V^{ee}_{\mu'\bar{\mu}\cdot\bar{\mu}'\mu}\,c^\dagger_{\mu'\lambda}c^\dagger_{\bar{\mu}'\lambda'}c_{\bar{\mu}\lambda'}c_{\mu\lambda} \\ &+\frac{1}{2}\sum_{\substack{\nu'\bar{\nu}\bar{\nu}'\nu\\ \lambda\lambda'}} V^{hh}_{\nu'\bar{\nu}\cdot\bar{\nu}'\nu}\,d^\dagger_{\nu'\lambda}d^\dagger_{\bar{\nu}'\lambda'}d_{\bar{\nu}\lambda'}d_{\nu\lambda} - \sum_{\substack{\mu'\bar{\mu}\bar{\nu}'\nu\\ \lambda\lambda'}} V^{eh}_{\mu'\bar{\mu}\cdot\bar{\nu}'\nu}\,c^\dagger_{\mu'\lambda}d^\dagger_{\bar{\nu}'\lambda'}d_{\nu\lambda'}c_{\mu\lambda} \end{aligned} \tag{A.11}$$

where the terms in the first line account for the single-particle properties of electrons and holes, those in the second line for the mutual electron and hole Coulomb interactions, and those in the third line for the Coulomb attractions between electrons and holes. All Coulomb

couplings in Eq. (A.11) preserve the spin orientations of the particles. For simplicity, we have neglected the electron-hole exchange interaction discussed in Section 3.1.3 as well as Auger-type Coulomb processes [270, 271]. The Coulomb matrix elements are given by

$$V^{ij}_{\mu'\mu,\nu'\nu} = \int dr dr' \, \frac{\phi^{i*}_{\mu'\lambda}(r)\phi^{i}_{\mu\lambda}(r)\,\phi^{j*}_{\nu'\lambda'}(r')\phi^{j}_{\nu\lambda'}(r')}{\kappa_s|r-r'|} \tag{A.12}$$

A word of caution is in order. It might be tempting to assume that Eq. (A.11) can be obtained in a first-principles manner from Eq. (9). This is not the case. Whereas Eq. (9) has a rather precise meaning in the first-principles framework of density functional theory [36], no comparably simple interpretation exists for the Coulomb terms of Eqs. (A.11) and (A.12). It turns out that dielectric screening in semiconductors is a highly complicated many-particle process [97, 271], which, surprisingly enough, approximately results in the dielectric screening constant κ_s. Thus, Eqs. (A.11) and (A.12) should be understood as an effective rather than first-principles description.

Direct Diagonalization

We will now show how the framework of second quantization can be used for the calculation of few-particle states in the strong confinement regime. Throughout, we shall assume that the single-particle description of Eq. (10) provides a good starting point and that Coulomb interactions only give rise to moderate renormalization effects. More precisely, for $\Delta\epsilon$ a typical single-particle level splitting and V a typical Coulomb matrix element, we assume that $V \ll \Delta\epsilon$, which allows to describe approximately the interacting few-particle system in terms of a limited basis of single-particle states—typically around 10 states for electrons and holes [58, 65, 67]. We stress that exciton or biexciton states in the weak confinement regime (i.e., electron-hole complexes that are bound because of Coulomb correlations) could not be described within such an approach.

Excitons

Consider first the Coulomb correlated states for one electron–hole pair–that is, the exciton states in the strong-confinement regime. Although they could be easily calculated without invoking the framework of second quantization, this analysis will allow us to grasp the essential features of configuration interaction calculations. We first define the Hilbert space under consideration. In view of the above discussion and keeping in mind that we are aiming at a computational scheme, we restrict our basis to a limited number of single-particle states (e.g., the 10 states of lowest energy for electrons and holes). Then, $|\mu, \nu\rangle = c^{\dagger}_{\mu\lambda} d^{\dagger}_{\nu\bar{\lambda}} |0\rangle$ provides a basis of approximately hundred states suited for the description of one electron and hole with opposite spin orientations (Fig. 7). We next expand the exciton in this basis,

$$|x\rangle = \sum_{\mu\nu} \Psi^{x}_{\mu\nu} |\mu, \nu\rangle \tag{A.13}$$

The exciton eigenstates $\Psi^{x}_{\mu\nu}$ and energies E_x are then obtained from the Schrödinger equation $H|x\rangle = E_x|x\rangle$, where H is the many-body hamiltonian defined in Eq. (A.11). To this end, we multiply the Schrödinger equation from the left-hand side with $\langle\mu, \nu|$ and obtain after some straightforward calculation the eigenvalue equation

$$\sum_{\mu'\nu'} \left((\epsilon^{e}_{\mu} + \epsilon^{h}_{\nu})\delta_{\mu\mu'}\delta_{\nu\nu'} - V^{eh}_{\mu\mu',\nu\nu'} \right) \Psi^{x}_{\mu'\nu'} = E_x \Psi^{x}_{\mu\nu} \tag{A.14}$$

Here, the term in parentheses on the left-hand side is the Hamiltonian matrix in the single-particle basis, and E_x and $\Psi^{x}_{\mu\nu}$ can be obtained by its direct diagonalization.

Biexcitons

Things can easily be extended to biexcitons. Before presenting the details of the underlying analysis, two points are worth mentioning. First, the proper anti-symmetrization of the electron–hole wavefunction is automatically guaranteed within the framework of second quantization. Second, the size of the Hilbert space for an n-body problem scales according to $\sim N^n$, where N is the number of single-particle states under consideration. This number becomes exceedingly fast prohibitively large for computational approaches. One thus introduces a further cutoff adapted from the single-particle energies of the few-particle basis states. Consider the basis $|\mu, \nu; \mu', \nu'\rangle = c^\dagger_{\mu\lambda} d^\dagger_{\nu\bar{\lambda}} c^\dagger_{\mu'\bar{\lambda}} d^\dagger_{\nu'\lambda} |0\rangle$ for the description of a biexciton where the two electron–hole pairs have opposite spin orientations. Then,

$$|b\rangle = \sum_{\mu\nu,\mu'\nu'} \bar{\Psi}^b_{\mu\nu,\mu'\nu'} |\mu, \nu; \mu', \nu'\rangle \tag{A.15}$$

defines the biexciton state. The biexciton wavefunctions $\bar{\Psi}^b_{\mu\nu,\mu'\nu'}$ and energies $\bar{E}_b$ are obtained from Schrödinger's equation with the many-body Hamiltonian (A.11),

$$\begin{aligned}(\epsilon^e_\mu + \epsilon^h_\nu + \epsilon^e_{\mu'} + \epsilon^h_{\nu'})\bar{\Psi}^b_{\mu\nu,\mu'\nu'} &+ \sum_{\bar{\mu}\bar{\mu}'} V^{ee}_{\mu\bar{\mu},\mu'\bar{\mu}'} \bar{\Psi}^b_{\bar{\mu}\nu,\mu'\nu'} + \sum_{\bar{\nu}\bar{\nu}'} V^{hh}_{\nu\bar{\nu},\nu'\bar{\nu}'} \bar{\Psi}^b_{\mu\bar{\nu},\mu'\bar{\nu}'} \\ &- \sum_{\bar{\mu}\bar{\nu}} V^{eh}_{\mu\bar{\mu},\nu\bar{\nu}} \bar{\Psi}^b_{\bar{\mu}\bar{\nu},\mu'\nu'} - \sum_{\mu'\bar{\nu}'} V^{eh}_{\mu'\bar{\mu}',\nu'\bar{\nu}'} \bar{\Psi}^b_{\mu\nu,\bar{\mu}'\bar{\nu}'} \\ &- \sum_{\mu\bar{\nu}'} V^{eh}_{\mu\bar{\mu},\nu'\bar{\nu}'} \bar{\Psi}^b_{\bar{\mu}\nu,\mu'\bar{\nu}'} - \sum_{\bar{\mu}'\bar{\nu}} V^{eh}_{\mu'\bar{\mu}',\nu\bar{\nu}} \bar{\Psi}^b_{\mu\bar{\nu},\bar{\mu}'\nu'} \\ &= \bar{E}_b \bar{\Psi}^b_{\mu\nu,\mu'\nu'}\end{aligned} \tag{A.16}$$

Here, the terms in the first line account for the single-particle energies and the repulsive electron–electron and hole–hole interactions and those in the second and third line for the various attractive Coulomb interactions between electrons and holes. Again, the biexciton eigenstates $\bar{\Psi}^b_{\mu\nu,\mu'\nu'}$ and energies $\bar{E}_b$ are obtained through direct diagonalization of the Hamiltonian matrix. The same scheme can further be extended to other few-particle complexes, such as, for example, triexcitons or multicharged excitons. It turns out to be advantageous to derive general rules for the construction of the Hamiltonian matrix. The interested reader is refered to the literature [272–274].

APPENDIX C: TWO-LEVEL SYSTEM

Two-level systems are conveniently described in terms of the Pauli matrices

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{A.17}$$

They are Hermitian $\sigma_i^\dagger = \sigma_i$, have trace zero $\mathrm{tr}\,\sigma_i = 0$, and fulfill the important relation

$$\sigma_i \sigma_j = \delta_{ij} \mathbb{1} + i\epsilon_{ijk} \sigma_k \tag{A.18}$$

Here, ϵ_{ijk} is the total antisymmetric tensor, and we have used the Einstein summation convention. It immediately follows that $\sigma_i^2 = \mathbb{1}$. For the commutation and anticommutation relations, we obtain

$$[\sigma_i, \sigma_j] = 2i\epsilon_{ijk} \sigma_k, \quad \{\sigma_i, \sigma_j\} = 2\delta_{ij} \mathbb{1} \tag{A.19}$$

We have now all important relations at hand. Let us first compute the expression $\exp(\lambda \boldsymbol{a}\boldsymbol{\sigma})$, with $\boldsymbol{a} = a\hat{\boldsymbol{e}}$ an arbitrary real vector that has the norm $a = \|\boldsymbol{a}\|$ and the direction described

by the unit vector $\hat{\boldsymbol{e}} = \boldsymbol{a}/a$. To this end, we expand the exponential into a power series and obtain

$$\begin{aligned} e^{\lambda a\sigma} &= \mathbb{1} + \lambda(\boldsymbol{a}\boldsymbol{\sigma}) + \frac{\lambda^2}{2!}(\boldsymbol{a}\boldsymbol{\sigma})^2 + \frac{\lambda^3}{3!}(\boldsymbol{a}\boldsymbol{\sigma})^3 + \frac{\lambda^4}{4!}(\boldsymbol{a}\boldsymbol{\sigma})^4 + \ldots \\ &= \mathbb{1} + (\lambda a)\hat{\boldsymbol{e}}\boldsymbol{\sigma} + \frac{(\lambda a)^2}{2!}\mathbb{1} + \frac{(\lambda a)^3}{3!}\hat{\boldsymbol{e}}\boldsymbol{\sigma} + \frac{(\lambda a)^4}{4!}\mathbb{1} + \ldots \\ &= \cos a\lambda\mathbb{1} + \sin a\lambda\hat{\boldsymbol{e}}\boldsymbol{\sigma} \end{aligned} \tag{A.20}$$

To arrive at the second line, we have used $(\boldsymbol{a}\boldsymbol{\sigma})^2 = a^2\,\mathbb{1}$, which immediatly follows from Eq. (A.18). This expression can be used, for example, for computing the time evolution operator of a two-level system. Things have to be slightly modified for the conditional time evolution in the unraveling (80) of the master equation. The conditional time evolution (80) of a two-level system driven by the resonant laser Ω and subject to spontaneous photon emissions is described by the effective Hamiltonian

$$H_{\rm eff} = -\frac{1}{2}\left[\Omega\,\sigma_1 + i\Gamma\frac{1}{2}(\mathbb{1} + \sigma_3)\right] = -i\frac{\Gamma}{4}\mathbb{1} - \Omega_{\rm eff}\,\hat{\boldsymbol{e}}\boldsymbol{\sigma} \tag{A.21}$$

Here, $\Omega_{\rm eff}^2 = \Omega^2 - (\Gamma/2)^2$ and $\tilde{\boldsymbol{e}} = (\cosh\theta, 0, i\sinh\theta)$, where the angle θ is defined through $\tanh\theta = \Gamma/(2\Omega)$. By use of $\hat{\boldsymbol{e}}^2 = 1$, we obtain for the conditional time evolution operator $e^{-iH_{\rm eff}t}$ [in a similar manner to Eq. (A.20)] the result

$$U_{\rm eff}(t) = e^{-iH_{\rm eff}t} = e^{-\Gamma t/4}\left(\cos\frac{\Omega_{\rm eff}t}{2}\,\mathbb{1} + i\sin\frac{\Omega_{\rm eff}t}{2}\,\hat{\boldsymbol{e}}\boldsymbol{\sigma}\right) \tag{A.22}$$

For the initial density operator $|0\rangle\langle 0| = (\mathbb{1} - \sigma_3)/2$, the probability that within $[0, t]$ the system has not emitted a photon is

$$P_0(t) = \frac{1}{2}{\rm tr}[U_{\rm eff}(t)\,(\mathbb{1} - \sigma_3)\,U_{\rm eff}^\dagger(t)] \tag{A.23}$$

In the evaluation of the above expression we only have to consider the products of those terms which give $\mathbb{1}$ (because those with σ vanish when performing the trace). Then,

$$P_0(t) = e^{-\Gamma/2t}\left[\cos^2\frac{\Omega_{\rm eff}t}{2} + \sin^2\frac{\Omega_{\rm eff}t}{2} - i\cos\frac{\Omega_{\rm eff}t}{2}\,\sin\frac{\Omega_{\rm eff}t}{2}\,(\hat{\boldsymbol{e}} - \hat{\boldsymbol{e}}^*)\hat{\boldsymbol{e}}_3\right] \tag{A.24}$$

which after a few minor manipulations finally gives Eq. (82).

Bloch Equations

Consider the Hamiltonian (58) of a two-level system subject to the driving laser $e^{-i\omega_l t}\,\Omega$. In the interaction representation according to $\omega_0\,|1\rangle\langle 1|$, we can remove the fast time dependence of $e^{-i\omega_0 t}\,\Omega$ (*rotating frame* [15, 225]), and obtain

$$\begin{aligned} H &= \frac{1}{2}(\Delta\,\sigma_3 - \Omega^*\,|0\rangle\langle 1| - \Omega\,|1\rangle\langle 0|) \\ &= \frac{1}{2}\,\Delta\,\sigma_3 - \frac{1}{4}[(\Re e\,\Omega - i\Im m\,\Omega)(\sigma_1 - i\sigma_2) + (\Re e\,\Omega + i\Im m\,\Omega)(\sigma_1 + i\sigma_2)] \\ &= \frac{1}{2}(\Delta\,\sigma_3 - \Re e\,\Omega\,\sigma_1 + \Im m\,\Omega\,\sigma_2) \end{aligned} \tag{A.25}$$

Here, we have used $|0\rangle\langle 1| = (\sigma_1 - i\sigma_2)/2$ and $|1\rangle\langle 0| = (\sigma_1 + i\sigma_2)/2$ and have decomposed Ω into its real and imaginary part. $\Delta = E_{01} - \omega_{01}$ is the detuning of the two-level system with respect to the laser frequency ω_{01}. Inserting this Hamiltonian together with the density operator (57) into the Liouville von-Neumann equation gives

$$\dot{\boldsymbol{u}}\boldsymbol{\sigma} = -i\left[\frac{1}{2}\,\boldsymbol{\Omega}\boldsymbol{\sigma}\right] = (\boldsymbol{\Omega}\times\boldsymbol{u})\boldsymbol{\sigma} \tag{A.26}$$

where we have used Eq. (A.19) to arrive at the last term. We next multiply this equation with $\boldsymbol{\sigma}$ and take the trace to arrive finally at the coherent part (59) of the optical Bloch equations. For the incoherent part, we express the Lindblad operators according to $L = a_0 \mathbb{1} + \boldsymbol{a}\boldsymbol{\sigma}$, with the complex coefficients $a_0 = a_0' + ia_0''$ and $\boldsymbol{a} = \boldsymbol{a}' + i\boldsymbol{a}''$. When this operator is inserted into the master equation (67) of Lindblad form, we obtain after some lengthy but straightforward calculation the incoherent part of the Bloch equations [147]

$$\dot{\boldsymbol{u}} \cong 2(2(\boldsymbol{a}' \times \boldsymbol{a}'') - (a_0'\boldsymbol{a}'' - a_0''\boldsymbol{a}') \times \boldsymbol{u} - |\boldsymbol{a}|^2 \boldsymbol{u} + (\boldsymbol{a}'\boldsymbol{u})\,\boldsymbol{a}' + (\boldsymbol{a}''\boldsymbol{u})\,\boldsymbol{a}'') \tag{A.27}$$

For the Lindblad operator $L = \sqrt{\Gamma}|0\rangle\langle 1| = \sqrt{\Gamma}(\sigma_1 - i\sigma_2)/2$, corresponding to $a_0 = 0$ and $\boldsymbol{a} = \sqrt{\Gamma}(\hat{\boldsymbol{e}}_1 - i\hat{\boldsymbol{e}}_2)/2$, we finally arrive at Eq. (6), with the longitudinal and transverse scattering times $T_1 = 1/\Gamma$ and $T_2 = 2/\Gamma$, respectively.

APPENDIX D: INDEPENDENT BOSON MODEL

In this appendix, we show how to evaluate the polarization fluctuations $G(t) = \langle \sigma_-(0)\sigma_+(t)\rangle$ for the spin-boson Hamiltonian $H = E_0|1\rangle\langle 1| + H_0 + V$ of Eq. (84), with $H_0 = \sum_i \omega_i a_i^\dagger a_i$ and V the dot-phonon coupling. $E_0|1\rangle\langle 1|$ commutes with both H_0 and V, and correspondingly $e^{iHt} = e^{iE_0 t|1\rangle\langle 1|}\, e^{i(H_0+V)t}$. Inserting this expression into $G(t)$ allows to evaluate all expressions involving the system operators $|1\rangle\langle 1|$ and $\sigma_\pm$ explicitly, and we obtain

$$G(t) = e^{iE_0 t}\langle e^{i(H_0+V)t}\, e^{-iH_0 t}\rangle \tag{A.28}$$

Here, we have assumed that the expectation value $\langle . \rangle$ is for the system in the groundstate and for a thermal distribution of phonons, and we have used that $e^{-i(H_0+V)t}|0\rangle = e^{-iH_0 t}|0\rangle$, which follows upon expanding the exponential in its power series and using that $V|0\rangle = 0$. We next introduce the displacement operator

$$D(\xi) = e^{\xi a^\dagger - \xi^* a}, \quad D^\dagger(\xi) = D^{-1}(\xi) = D(-\xi) \tag{A.29}$$

of the harmonic oscillator [89, 275]. It has the important properties

$$\begin{aligned} D^\dagger(\xi)\, a\, D(\xi) &= a + \xi \\ D^\dagger(\xi)\, a^\dagger D(\xi) &= a^\dagger + \xi^* \\ D^\dagger(\xi) f(a, a^\dagger) D(\xi) &= f(a + \xi, a^\dagger + \xi^*) \end{aligned} \tag{A.30}$$

with $f(a, a^\dagger)$ an arbitrary function of the field operators a and $a^\dagger$. The last expression can easily be proven by inserting $D(\xi)D^\dagger(\xi) = \mathbb{1}$ in the power series of $f(a, a^\dagger)$. Because in Eq. (A.28) the different oscillators propagate independently of each other, in the following it suffices to consider only one phonon mode. Then, Eq. (A.30) can be used to simplify the time evolution operator according to

$$e^{i(H_0+V)t} = e^{-i\xi^2\omega t}\, D^\dagger(\xi)\, e^{iH_0 t}\, D(\xi) \tag{A.31}$$

with $\xi = g/\omega$. Accordingly, we can express $G(t)$ for a single phonon mode through

$$G(t) = e^{iE_0 t}\langle D^\dagger(\xi)\, e^{iH_0 t}\, D(\xi)\, e^{-iH_0 t}\rangle = e^{i\bar{E}_0 t}\langle D^\dagger(\xi)\, D(\xi\, e^{i\omega t})\rangle \tag{A.32}$$

where $\bar{E}_0 = E_0 - \xi^2\omega$, and the last expression has been derived by evaluating $D(\xi)$ in the interaction representation according to H_0. We can now use the relation $D(\xi)\, D(\xi') = e^{-i\,\mathrm{Im}(\xi\xi')}\, D(\xi + \xi')$ for the displacement operators [89, 275] to simplify expression (A.32) to

$$G(t) = e^{i(\bar{E}_0 t + \xi^2 \sin\omega t)}\langle D(\xi[e^{i\omega t} - 1])\rangle \tag{A.33}$$

In the remainder, we discuss how this expression can be evaluated for a thermal phonon distribution. To this end, we use the factorization $D(\xi) = e^{-|\xi|^2} e^{\xi a^\dagger} e^{-\xi^* a}$ of the displacement operator. For a Bose–Einstein distribution $n(\omega)$ of the phonons, it can be shown that [23, 135, 275]

$$\langle (a^\dagger)^l a^l \rangle = l! [n(\omega)]^l \tag{A.34}$$

To compute $\langle D(\xi) \rangle$, we expand the exponentials $e^{\xi a^\dagger}$ and $e^{-\xi^* a}$ in power series and use that only terms with an equal number of creation and annihilation operators give a nonvanishing contribution. Then,

$$\langle D(\xi) \rangle = \exp\left\{ -|\xi|^2 \left[n(\omega) + \frac{1}{2} \right] \right\} \tag{A.35}$$

The final result (87) is obtained by using $|e^{i\omega t} - 1|^2 = 2(1 - \cos \omega t)$, $n(\omega) + 1/2 = 1/2 \coth(\beta\omega/2)$ and introducing an appropriate summation over all phonon modes.

REFERENCES

1. U. Woggon, "Optical Properties of Semiconductor Quantum Dots." Springer, Berlin, 1997.
2. D. Bimberg, M. Grundmann, and N. Ledentsov, "Quantum Dot Heterostructures." John Wiley, New York, 1998.
3. L. Jacak, P. Hawrylak, and A. Wojs, "Quantum Dots." Springer, Berlin, 1998.
4. A. Zrenner, L. V. Butov, M. Hagn, G. Abstreiter, G. Böhm, and G. Weimann, *Phys. Rev. Lett.* 72, 3382 (1994).
5. H. F. Hess, E. Betzig, T. D. Harris, L. N. Pfeiffer, and K. W. West, *Science* 264, 1740 (1994).
6. D. Gammon, E. S. Snow, B. V. Shanabrook, D. S. Katzer, and D. Park, *Phys. Rev. Lett.* 76, 3005 (1996).
7. D. Gammon, E. S. Snow, B. V. Shanabrook, D. S. Katzer, and D. Park, *Science* 273, 87 (1996).
8. K. Matsuda, T. Saiki, S. Nomura, M. Mihara, Y. Aoyagi, S. Nair, and T. Takagahara, *Phys. Rev. Lett.* 91, 177401 (2003).
9. J. Y. Marzin, J. M. Gérard, A. Izraël, D. Barrier, and G. Bastard, *Phys. Rev. Lett.* 73, 716 (1994).
10. M. Grundmann, J. Christen, N. N. Ledentsov, J. Böhrer, D. Bimberg, S. S. Ruvimov, P. Werner, U. Richter, U. Gösele, J. Heydenreich, V. M. Ustinov, A. Yu. Egorov, A. E. Zhukov, P. S. Kopev, and Zh. I. Alferov, *Phys. Rev. Lett.* 74, 4043 (1995).
11. R. Leon, P. M. Petroff, D. Leonard, and S. Fafard, *Science* 267, 1966 (1995).
12. H. Haug and S. W. Koch, "Quantum Theory of the Optical and Electronic Properties of Semiconductors." World Scientific, Singapore, 1993.
13. P. Y. Yu and M. Cardona, "Fundamentals of Semiconductors." Springer, Berlin, 1996.
14. W. H. Zurek, *Rev. Mod. Phys.* 75, 715 (2003).
15. M. O. Scully and M. S. Zubairy, "Quantum Optics." Cambridge University Press, Cambridge, UK, 1997.
16. H. Fritzsch, *CERN Courier* 43, 13 (2003).
17. S. Chu, *Nature* 416, 206 (2002).
18. C. H. Bennett and D. P. DiVincenzo, *Nature* 404, 247 (2000).
19. D. Bouwmeester, A. Ekert, and A. Zeilinger (eds.), "The Physics of Quantum Information." Springer, Berlin, 2000.
20. M. A. Nielsen and I. L. Chuang, "Quantum Computation and Quantum Information." Cambridge University Press, Cabmridge, UK, 2000.
21. N. Gisin, G. Ribordy, W. Tittel, and H. Zbinden, *Rev. Mod. Phys.* 74, 145 (2002).
22. L. Mandel and E. Wolf, "Optical Coherence and Quantum Optics." Cambridge University Press, Cambridge, UK, 1995.
23. G. D. Mahan, "Many-Particle Physics." Plenum, New York, 1981.
24. P. Borri, W. Langbein, S. Schneider, U. Woggon, R. L. Sellin, D. Ouyang, and D. Bimberg, *Phys. Rev. Lett.* 87, 157401 (2001).
25. B. Krummheuer, V. M. Axt, and T. Kuhn, *Phys. Rev. B* 65, 195313 (2002).
26. N. H. Bonadeo, G. Chen, D. Gammon, D. S. Katzer, D. Park, and D. G. Steel, *Phys. Rev. Lett.* 81, 2759 (1998).
27. F. Rossi and T. Kuhn, *Rev. Mod. Phys.* 74, 895 (2002).
28. N. H. Bonadeo, J. Erland, D. Gammon, D. Park, D. S. Katzer, and D. G. Steel, *Science* 282, 1473 (1998).
29. J. Förstner, C. Weber, J. Dankwerts, and A. Knorr, *Phys. Rev. Lett.* 91, 127401 (2003).
30. J. Förstner, C. Weber, J. Dankwerts, and A. Knorr, *Phys. Status Solidi.* (b) 238, 419 (2003).
31. J. R. Guest, T. H. Stievater, X. Li, J. Cheng, D. G. Steel, D. Gammon, D. S. Katzer, D. Park, C. Ell, A. Thränhardt, G. Khitrova, and H. M. Gibbs, *Phys. Rev. B* 65, 241310 (2002).
32. J. G. Tischler, A. S. Bracker, D. Gammon, and D. Park, *Phys. Rev. B* 66, 081310 (2002).
33. M. Bayer, G. Ortner, O. Stern, A. Kuther, A. A. Gorbunov, A. Forchel, P. Hawrylak, S. Fafard, K. Hinzer, T. L. Reinecke, S. N. Walck, J. P. Reithmaier, F. Klopf, and F. Schäfer, *Phys. Rev. B* 65, 195315 (2002).
34. A. S. Lenihan, M. V. Gurudev Dutt, D. G. Steel, S. Ghosh, and P. K. Bhattacharya, *Phys. Rev. Lett.* 88, 223601 (2002).

35. M. Paillard, X. Marie, P. Renucci, T. Amand, A. Jbeli, and J. M. Gérard, *Phys. Rev. Lett.* 86, 1634 (2001).
36. R. M. Dreizler and E. U. Gross, "Density Functional Theory," Springer, Berlin, 1990.
37. H. Jiang and J. Singh, *Phys. Rev. B* 56, 4696 (1997).
38. C. Pryor, *Phys. Rev. B* 57, 7190 (1998).
39. C. Pryor, *Phys. Rev. B* 60, 2869 (1999).
40. O. Stier, M. Grundmann, and D. Bimberg, *Phys. Rev. B* 59, 5688 (1999).
41. A. J. Williamson, L. W. Wang, and A. Zunger, *Phys. Rev. B* 62, 12 963 (2000).
42. J. Shumway, A. Franceschetti, and A. Zunger, *Phys. Rev. B* 63, 155316 (2001).
43. G. Bester, S. Nair, and Alex Zunger, *Phys. Rev. B* 67, 161306 (2003).
44. D. A. Kleinman, *Phys. Rev. B* 28, 871 (1983).
45. G. Bastard, "Wave Mechanics Applied to Semiconductor Heterostructures," Les Editions de Physique, Les Ulis, 1989.
46. R. Zimmermann, F. Große, and E. Runge, *Pure Appl. Chem.* 69, 1179 (1997).
47. U. Hohenester, G. Goldoni, and E. Molinari, *Appl. Phys. Lett.* 84, 3963 (2004).
48. K. Brunner, G. Abstreiter, G. Böhm, G. Tränkle, and G. Weimann, *Phys. Rev. Lett.* 73, 1138 (1994).
49. D. Gammon, E. S. Snow, and D. S. Katzer, *Appl. Phys. Lett.* 67, 2391 (1995).
50. Q. Wu, R. D. Grober, D. Gammon, and D. S. Katzer, Excitons, *Phys. Rev. B* 62, 13022 (2000).
51. P. Hawrylak, *Phys. Rev. B* 60, 5597 (1999).
52. E. O. Kane, in "Semiconductors and semimetals," (R. K. Willardson and A. C. Beer, Eds.), Vol. 39, p. 75, Academic, New York, 1966.
53. I. Žutić, J. Fabian, and S. Das Sarma, *Rev. Mod. Phys* 76, 323 (2004).
54. M. Bayer, A. Kuther, A. Forchel, A. Gorbunov, V. B. Timofeev, F. Schäfer, J. P. Reithmaier, T. L. Reinecke, and S. N. Walck, *Phys. Rev. Lett.* 82, 1748 (1999).
55. A. V. Filinov, M. Bonitz, and Y. E. Lozovik, *Phys. Status Solidi* (c) 0, 1441 (2003).
56. S. Rodt, R. Heitz, A. Schliwa, R. L. Sellin, F. Guffarth, and D. Bimberg, *Phys. Rev. B* 68, 035331 (2003).
57. M. Bayer, O. Stern, P. Hawrylak, S. Fafard, and A. Forchel, *Nature* 405, 923 (2000).
58. R. Rinaldi, S. Antonaci, M. D. Vittorio, R. Cingolani, U. Hohenester, E. Molinari, H. Lipsanen, and J. Tulkki, *Phys. Rev. B* 62, 1592 (2000).
59. S. Raymond, S. Studenikin, A. Sachrajda, Z. Wasilewski, S. J. Cheng, W. Sheng, P. Hawrylak, A. Babinski, M. Potemski, G. Ortner, M. Bayer, *Phys. Rev. Lett.* 92, 187402 (2004).
60. L. Landin, M. S. Miller, M. E. Pistol, C. E. Pryor, and L. Samuelson, *Science* 280, 262 (1998).
61. E. Dekel, D. Gershoni, E. Ehrenfreund, D. Spektor, J. M. Garcia, and M. Petroff, *Phys. Rev. Lett.* 80, 4991 (1998).
62. A. Zrenner, *J. Chem. Phys.* 112, 7790 (2000).
63. R. J. Warburton, C. S. Dürr, K. Karrai, J. P. Kotthaus, G. Medeiros-Ribeiro, and P. M. Petroff, *Phys. Rev. Lett.* 79, 5282 (1997).
64. R. J. Warburton, C. Schäflein, D. Haft, F. Bickel, A. Lorke, K. Karrai, J. M. Garcia, W. Schoenfeld, and P. M. Petroff, *Nature* 405, 926 (2000).
65. F. Findeis, M. Baier, A. Zrenner, M. Bichler, G. Abstreiter, U. Hohenester, and E. Molinari, *Phys. Rev. B* 63, 121309(R) (2001).
66. D. V. Regelman, E. Dekel, D. Gershoni, E. Ehrenfreund, A. J. Williamson, J. Shumway, A. Zunger, W. V. Schoenfeld, and P. M. Petroff, *Phys. Rev. B* 64, 165301 (2001).
67. A. Hartmann, Y. Ducommun, E. Kapon, U. Hohenester, and E. Molinari, *Phys. Rev. Lett.* 84, 5648 (2000).
68. G. L. Snider, A. O. Orlov, I. Amlani, X. Zuo, G. H. Bernstein, C. S. Lent, J. L. Merz, and W. Porod, *J. Appl. Phys.* 85, 4283 (1999).
69. M. Rontani, F. Troiani, U. Hohenester, and E. Molinari, *Solid State Commun.* 119, 309 (2001).
70. B. Partoens and F. M. Peeters, *Phys. Rev. Lett.* 84, 4433 (2000).
71. L. Martín-Moreno, L. Brey, and C. Tejedor, *Phys. Rev. B* 62, 10 633 (2000).
72. T. H. Oosterkamp, S. F. Godijn, M. J. Uilenreef, Y. V. Nazarov, N. C. van der Vaart, and L. P. Kouwenhoven, *Phys. Rev. Lett.* 80, 4951 (1998).
73. T. H. Oosterkamp, T. Fujisawa, W. G. van der Wiel, K. Ishibashi, R. Hijman, S. Tarucha, and L. P. Kouwenhoven, *Nature* 395, 873 (1998).
74. T. Fujisawa, T. H. Oosterkamp, W. G. van der Wiel, B. W. Broer, R. Aguado, S. Tarucha, and L. P. Kouwenhoven, *Science* 282, 923 (1998).
75. T. Schmidt, R. J. Haug, K. V. Klitzing, A. Förster, and H. Lüth, *Phys. Rev. Lett.* 78, 1544 (1997).
76. R. H. Blick, D. Pfannkuche, R. J. Haug, K. V. Klitzing, and K. Eberl, *Phys. Rev. Lett.* 80, 4032 (1998).
77. R. H. Blick, D. W. van der Weide, R. J. Haug, and K. Eberl, *Phys. Rev. Lett.* 81, 689 (1998).
78. M. Brodsky, N. B. Zhitenev, R. C. Ashoori, L. N. Pfeiffer, and K. W. West, *Phys. Rev. Lett.* 85, 2356 (2000).
79. S. Amaha, D. G. Austing, Y. Tokura, K. Muraki, K. Ono, and S. Tarucha, *Solid State Commun.* 119, 183 (2001).
80. S. Fafard, M. Spanner, J. P. McCaffrey, and Z. R. Wasilewski, *Appl. Phys. Lett.* 76, 2268 (2000).
81. G. Schedelbeck, W. Wegscheider, M. Bichler, and G. Abstreiter, *Science* 278, 1792 (1997).
82. M. Bayer, P. Hawrylak, K. Hinzer, S. Fafard, M. Korkusinski, R. Wasilewski, O. Stern, and A. Forchel, *Science* 291, 451 (2001).
83. P. Borri, W. Langbein, U. Woggon, M. Schwab, M. Bayer, S. Fafard, Z. Wasilewski, and P. Hawrylak, *Phys. Rev. Lett.* 91, 267401 (2003).
84. F. Troiani, U. Hohenester, and E. Molinari, *Phys. Rev. B* 65, 161301(R) (2002).

85. K. L. Janssens, B. Partoens, and F. M. Peeters, *Phys. Rev. B* 65, 233301 (2002).
86. K. L. Janssens, B. Partoens, and F. M. Peeters, *Phys. Rev. B* 66, 075314 (2002).
87. P. Koskinen and U. Hohenester, *Solid State Commun.* 125, 529 (2003).
88. J. D. Jackson, "Classical Electrodynamics." Wiley, New York, 1962.
89. D. F. Walls and G. J. Millburn, "Quantum Optics." Springer, Berlin, 1995.
90. L. C. Andreani, G. Panzarini, and J.-M. Gérard, *Phys. Rev. B* 60, 13 276 (1999).
91. F. Rossi, *Semicond. Sci. Technol.* 13, 147 (1998).
92. N. H. Bonadeo, A. S. Lenihan, G. Chen, J. R. Guest, D. G. Steel, D. Gammon, D. S. Katzer, and D. Park, *Appl. Phys. Lett.* 75, 2933 (1999).
93. G. Chen, N. H. Bonadeo, D. G. Steel, D. Gammon, D. S. Katzer, D. Park, and L. J. Sham, *Science* 289, 1906 (2000).
94. J. R. Guest, T. H. Stievater, Gang Chen, E. A. Tabak, B. G. Orr, D. G. Steel, D. Gammon, and D. S. Katzer, *Science* 293, 2224 (2001).
95. X. Li, Y. Wu, D. Steel, D. Gammon, T. H. Stievater, D. S. Katzer, D. Park, C. Piermarocchi, and L. J. Sham, *Science* 301, 809 (2003).
96. R. Kubo, M. Toda, and M. Hashitsume, "Statistical Physics II." Springer, Berlin, 1985.
97. G. Onida, L. Reining, and A. Rubio, *Rev. Mod. Phys.* 74, 601 (2002).
98. E. T. Batteh, Jun Cheng, Gang Chen, D. G. Steel, D. Gammon, D. S. Katzer, and D. Park, *Appl. Phys. Lett.* 84, 1928 (2004).
99. Q. Wu, R. D. Grober, D. Gammon, and D. S. Katzer, *Phys. Rev. Lett.* 83, 2652 (1999).
100. T. Guenther, C. Lienau, T. Elsaesser, M. Glanemann, V. M. Axt, T. Kuhn, S. Eshlaghi, and A. D. Wieck, *Phys. Rev. Lett.* 89, 057401 (2002).
101. T. S. Sosnowskii, T. B. Norris, H. Jiang, J. Singh, K. Kamath, and P. Bhattacharya, *Phys. Rev. B* 57, R9423 (1998).
102. U. Hohenester and E. Molinari, *Phys. Status Solidi* (b) 221, 19 (2000).
103. C. Simserides, U. Hohenester, G. Goldoni, and E. Molinari, *Phys. Rev. B* 62, 13657 (2000).
104. P. Hawrylak, G. A. Narvaez, M. Bayer, and A. Forchel, *Phys. Rev. Lett.* 85, 389 (2000).
105. M. Kira, F. Jahnke, and S. W. Koch, *Phys. Rev. Lett.* 81, 3263 (1998).
106. A. Zrenner, E. Beham, S. Stufler, F. Findeis, M. Bichler, and B. Abstreiter, *Nature* 418, 612 (2002).
107. J. Motohisa, J. J. Baumberg, A. P. Heberle, and J. Allam, *Solid-State Electron.* 42, 1335 (1998).
108. M. Bayer, T. Gubrod, A. Forchel, V. D. Kulakovskii, A. Gorbunov, M. Michel, R. Steffen, and K. H. Wand, *Phys. Rev. B* 58, 4740 (1998).
109. E. Dekel, D. Gershoni, E. Ehrenfreund, J. M. Garcia, and P. M. Petroff, *Phys. Rev. B* 61, 11009 (2000).
110. F. Findeis, A. Zrenner, G. Böhm, and G. Abstreiter, *Solid State Commun.* 114, 227 (2000).
111. J. J. Finley, P. W. Fry, A. D. Ashmore, A. Lemaitre, A. I. Tartakovskii, R. Oulton, D. J. Mowbray, M. S. Skolnick M. Hopkinson, P. D. Buckle, P. A. Maksym, *Phys. Rev. B* 63, 161305 (2001).
112. B. Urbaszek, R. J. Warburton, K. Karrai, B. D. Gerardot, P. M. Petroff, and J. M. Garcia, *Phys. Rev. Lett.* 90, 247403 (2003).
113. L. Besombes, J. J. Baumberg, and J. Motohisa, *Phys. Rev. Lett.* 90, 257402 (2003).
114. M. Lomascolo, A. Vergine, T. K. Johal, R. Rinaldi, A. Passaseo, R. Cingolani, S. Patanè, M. Labardi, M. Allegrini, F. Troiani, E. Molinari, *Phys. Rev. B* 66, 041302 (2002).
115. M. A. Paesler and P. J. Moyer, "Near-Field Optics: Theory, Instrumentation, and Applications." Wiley, New York, 1996.
116. B. Hecht, B. Sick, U. P. Wild, V. Deckert, R. Zenobi, O. J. F. Martin, and D. W. Pohl, *J. Chem. Phys.* 112, 7761 (2000).
117. B. Hanewinkel, A. Knorr, P. Thomas, and S. W. Koch, *Phys. Rev. B* 55, 13 715 (1997).
118. A. Liu and G. W. Bryant, *Phys. Rev. B* 59, 2245 (1999).
119. A. Liu and G. W. Bryant, *Phys. Rev. B* 59, 2245 (1999).
120. K. Matsuda, T. Saiki, S. Nomura, M. Mihara, and Y. Aoyagi, *Appl. Phys. Lett.* 81, 2291 (2002).
121. O. Mauritz, G. Goldoni, F. Rossi, and E. Molinari, *Phys. Rev. Lett.* 82, 847 (1999).
122. O. Mauritz, G. Goldoni, E. Molinari, and F. Rossi, *Phys. Rev. B* 62, 8204 (2000).
123. S. Savasta, O. Di Stefano, and R. Girlanda, *Phys. Rev. A* 65, 043801 (2002).
124. G. Pistone, S. Savasta, O. Di Stefano, and R. Girlanda, *Appl. Phys. Lett.* 84, 2971 (2004).
125. J. Shah, "Ultrafast Spectroscopy of Semiconductors and Semiconductor Nanostructures." Springer, Berlin, 1996.
126. N. H. Bonadeo, G. Chen, D. Gammon, and D. G. Steel, *Phys. Status Solidi* (b), 221, 5 (2000).
127. H. A. Rabitz, M. M. Hsieh, and C. M. Rosenthal, *Science* 303, 1998 (2004).
128. L. Viola and S. Lloyd, *Phys. Rev. A* 58, 2733 (1998).
129. L. E. Reichl, "Statistical Physics." Wiley, New York, 1998.
130. I. I. Rabi, *Phys. Rev.* 51, 652 (1937).
131. H. Htoon, T. Takagahara, D. Kulik, O. Baklenov, A. L. Holmes, Jr., and C. K. Shih, *Phys. Rev. Lett.* 88 087401 (2002).
132. P. Borri, W. Langbein, S. Schneider, U. Woggon, R. L. Sellin, D. Ouyang, and D. Bimberg, *Phys. Rev. B* 66, 081306(R) (2002).
133. H. Kamada, H. Gotoh, and J. Temmyo, *Phys. Rev. Lett.* 87, 246401 (2001).
134. H.-P. Breuer, B. Kappler, and F. Petruccione, *Phys. Rev. A* 59, 1633 (1999).
135. H.-P. Breuer and F. Petruccione, "Open Quantum Systems." Oxford University Press, New York, 2002.

136. S. Nakajima, *Prog. Theor. Phys.* 20, 948 (1958).
137. R. Zwanzig, *J. Chem. Phys.* 33, 1338 (1960).
138. R. Zwanzig, *Phys. Rev.* 124, 983 (1961).
139. E. Fick and G. Sauermann, "The Quantum Statistics of Dynamic Processes," Springer, Berlin, 1990.
140. A. O. Calderia and A. J. Leggett, *Phys. Rev. Lett.* 46, 211 (1981).
141. A. O. Caldeira and A. J. Leggett, *Ann. Phys. (N.Y.)* 149, 374 (1983).
142. A. J. Leggett, S. Chakravarty, A. T. Dorsey, M. P. A. Fisher and A. Garg, and W. Zwerger, *Rev. Mod. Phys.* 59, 1 (1987).
143. G. Lindblad, *Commun. Math. Phys.* 48, 119 (1976).
144. R. Dum, A. S. Parkins, P. Zoller, and C. W. Gardiner, *Phys. Rev. A* 46, 4382 (1992).
145. J. Dalibard, Y. Castin, and K. Molmer, *Phys. Rev. Lett.* 68, 580 (1992).
146. M. B. Plenio and P. L. Knight, *Rev. Mod. Phys.* 70, 101 (1998).
147. U. Hohenester, C. Sifel, and P. Koskinen, *Phys. Rev. B* 68, 245304 (2003).
148. M. Lax, *Phys. Rev.* 129, 2342 (1963).
149. M. Lax, *Rev. Mod. Phys.* 38, 541 (1966).
150. P. Borri, W. Langbein, S. Schneider, U. Woggon, R. L. Sellin, D. Ouyang, and D. Bimberg, *Phys. Rev. Lett.* 89, 187401 (2002).
151. V. Zwiller, T. Aichele, and O. Benson, *Phys. Rev. B* 69, 165307 (2004).
152. J. M. Gérard and B. Gayral, *J. Lightwave Technol.* 17, 2089 (1999).
153. Y. Yamamoto, F. Tassone, and H. Cao, "Semiconductor Cavity Quantum Electrodynamics," Springer, Berlin, 2000.
154. P. Michler, A. Kiraz, C. Becher, W. V. Schoenfeld, P. M. Petroff, L. Zhang, E. Hu, and A. Imamoglu, *Science* 290, 2282 (2000).
155. G. C. Hegerfeldt, *Phys. Rev. A* 47, 449 (1993).
156. C. Santori, M. Pelton, G. S. Solomon, Y. Dale, and Y. Yamamoto, *Phys. Rev. Lett.* 86, 1502 (2001).
157. M. Pelton, C. Santori, J. Vuckovic, B. Zhang, G. S. Solomon, J. Plant, and Y. Yamamoto, *Phys. Rev. Lett.* 89, 233602 (2002).
158. C. Santori, G. S. Solomon, M. Pelton, and Y. Yamamoto, *Phys. Rev. B* 65, 073310 (2002).
159. E. Waks, K. Inoue, C. Santori, D. Fattal, J. Vuckovic, G. S. Solomon, and Y. Yamamoto, *Nature* 420, 6917 (2002).
160. M. H. Baier, E. Pelucchi, E. Kapon, S. Varoutsis, M. Gallart, I. Robert-Philip, and I. Abram, *Appl. Phys. Lett.* 84, 648 (2004).
161. Z. Yuan, B. E. Kardynal, R. M. Stevenson, A. J. Shields, C. J. Lobo, K. Cooper, N. S. Beattie, D. A. Ritchie, and M. Pepper, *Science* 295, 102 (2002).
162. O. Benson, C. Santori, M. Pelton, and Y. Yamamoto, *Phys. Rev. Lett.* 84, 2513 (2000).
163. D. V. Regelman, U. Mizrahi, D. Gershoni, E. Ehrenfreund, W. V. Schoenfeld, and P. M. Petroff, *Phys. Rev. Lett.* 87, 257401 (2001).
164. E. Moreau, I. Robert, L. Manin, V. Thierry-Mieg, J. M. Gérard, and I. Abram, *Phys. Rev. Lett.* 87, 183601 (2001).
165. O. Gywat, G. Burkard, and D. Loss, *Phys. Rev. B* 65, 205329 (2002).
166. T. M. Stace, G. J. Milburn, and C. H. W. Barnes, *Phys. Rev. B* 67, 085317 (2003).
167. C. Sifel and U. Hohenester, *Appl. Phys. Lett.* 83, 153 (2003).
168. U. Bockelmann and G. Bastard, *Phys. Rev. B* 42, 8947 (1990).
169. H. Benisty, C. M. Sotomayor-Torres, and C. Weisbuch, *Phys. Rev. B* 44, 10945 (1991).
170. U. Bockelmann, *Phys. Rev. B* 48, 17637 (1993).
171. U. Bockelmann, *Phys. Rev. B* 50, 17271 (1994).
172. B. Ohnesorge, M. Albrecht, J. Oshinowo, A. Forchel, and Y. Arakawa, *Phys. Rev. B* 54, 11532 (1996).
173. R. Heitz, M. Veit, N. N. Ledentsov, A. Hoffmann, D. Bimberg, V. M. Ustinov, P. S. Kopev, and Zh. I. Alferov, *Phys. Rev. B* 56, 10435 (1997).
174. S. Grosse, J. H. Sandmann, G. von Plessen, J. Feldmann, H. Lipsanen, M. Sopanen, J. Tulkki, and J. Ahopelto, *Phys. Rev. B* 55, 4473 (1997).
175. C. B. Duke and D. Mahan, *Phys. Rev.* 139, A1965 (1965).
176. L. Jacak, P. Machnikowski, J. Krasnyj, and P. Zoller, *Eur. Phys. J. D* 22, 319 (2003).
177. C. Uchiyama and M. Aihara, *Phys. Rev. A* 66, 032313 (2002).
178. A. Yu. Smirnov, *Phys. Rev. B* 67, 155104 (2003).
179. A. Vagov, V. M. Axt, and T. Kuhn, *Phys. Rev. B* 67, 115338 (2003).
180. K. Král and Z. Khás, *Phys. Rev. B* 57, 2061 (1998).
181. T. Stauber, R. Zimmermann, and H. Castella, *Phys. Rev. B* 62, 7336 (2000).
182. A. Vagov, V. M. Axt, and T. Kuhn, *Phys. Rev. B* 66, 165312 (2002).
183. U. Hohenester and G. Stadler, *Phys. Rev. Lett.* 92, 196801 (2004).
184. M. Bonitz, "Quantum Kinetic Theory," Teubner, Stuttgart, 1998.
185. R. Balescu, "Statistical Mechanics of Charged Particles," Interscience, New York, 1963.
186. A. V. Uskov, A. P. Jauho, B. Tromborg, J. Mork, and R. Lang, *Phys. Rev. Lett.* 85, 1516 (2000).
187. S. Hameau, Y. Guldner, O. Verzelen, R. Ferreira, G. Bastard, J. Zeman, A. Lemaitre, and J. M. Gérard, *Phys. Rev. Lett.* 83, 4152 (1999).
188. K. Oshiro, K. Akai, and M. Matsura, *Phys. Rev. B* 59, 10 850 (1999).

189. M. Bissiri, G. Baldassarri Höger von Högersthal, A. S. Bhatti, M. Capizzi, A. Frova, P. Frigeri, and S. Franchi, *Phys. Rev. B* 62, 4642 (2000).
190. R. Ferreira and G. Bastard, *Appl. Phys. Lett.* 74, 2818 (1999).
191. O. Verzelen, R. Ferreira, and G. Bastard, *Phys. Rev. B* 62, R4809 (2000).
192. O. Verzelen, R. Ferreira, and G. Bastard, *Phys. Rev. Lett.* 88, 146803 (2002).
193. O. Verzelen, G. Bastard, and R. Ferreira, *Phys. Rev. B* 66, 081308 (2002).
194. L. Jacak, J. Krasnyj, D. Jacak, and P. Machinowski, *Phys. Rev. B* 65, 113305 (2002).
195. R. Heitz, H. Born, F. Guffarth, O. Stier, A. Schliwa, A. Hoffmann, and D. Bimberg, *Phys. Rev. B* 64, 241305 (2001).
196. J. Urayama, T. B. Norris, J. Singh, and P. Bhattacharya, *Phys. Rev. Lett.* 86, 4930 (2001).
197. P. Zanardi and F. Rossi, *Phys. Rev. Lett.* 81, 4752 (1998).
198. P. Zanardi and F. Rossi, *Phys. Rev. B* 59, 8170 (1999).
199. H. Kamada, H. Gotoh, H. Ando, J. Temmyo, and T. Tamamura, *Phys. Rev. B* 60, 5791 (1999).
200. S. Cortez, O. Krebs, S. Laurent, M. Senes, X. Marie, P. Voisin, R. Ferreira, G. Bastard, J.-M. Gérard, and T. Amand, *Phys. Rev. Lett.* 89, 207401 (2002).
201. A. V. Khaetskii and Y. V. Nazarov, *Phys. Rev. B* 61, 12 639 (2000).
202. W. Pötz and W. A. Schroeder (Eds.), "Coherent Control in Atoms, Molecules, and Semiconductors." Kluwer, Dordrecht, 1999
203. H. Rabitz, R. de Vivie-Riedle, M. Motzkus, and K. Kompka, *Science* 288, 824 (2000).
204. A. P. Heberle, J. J. Baumberg, and K. Köhler, *Phys. Rev. Lett.* 75, 2598 (1995).
205. K. Bergmann, H. Theuer, and B. W. Shore, *Rev. Mod. Phys.* 70, 1003 (1998).
206. U. Hohenester, F. Troiani, E. Molinari, G. Panzarini, and C. Macchiavello, *Appl. Phys. Lett.* 77, 1864 (2000).
207. M. Lindberg and R. Binder, *Phys. Rev. Lett.* 75, 1403 (1995).
208. W. Pötz, *Phys. Rev. Lett.* 79, 3262 (1997).
209. R. Binder and M. Lindberg, *Phys. Rev. Lett.* 81, 1477 (1998).
210. M. Artoni, G. C. La Rocca, and F. Bassani, *Europhys. Lett.* 49, 445 (2000).
211. T. Brandes and F. Renzoni, *Phys. Rev. Lett.* 85, 4148 (2000).
212. T. Brandes, F. Renzoni, and R. H. Blick, *Phys. Rev. B* 64, 035319 (2001).
213. F. Troiani, E. Molinari, and U. Hohenester, *Phys. Rev. Lett.* 90, 206802 (2003).
214. C. P. Slichter, "Principles of Magnetic Resonance," 3rd Edn. Springer, Berlin, 3, 1996.
215. M. H. Levitt, "Spin Dynamics, Basics of Nuclear Magnetic Resonance." Wiley, Chichester, 2003.
216. P. Chen, C. Piermarocchi, and L. J. Sham, *Phys. Rev. Lett.* 87, 067401 (2001).
217. C. Piermarocchi, P. Chen, Y. S. Dale, and L. J. Sham, *Phys. Rev. B* 65, 075307 (2002).
218. Pochung Chen, C. Piermarocchi, L. J. Sham, D. Gammon, and D. G. Steel, *Phys. Rev. B* 69, 075320 (2004).
219. A. P. Peirce, M. A. Dahleh, and H. Rabitz, *Phys. Rev. A* 37, 4950 (1988).
220. A. Borzì, G. Stadler, and U. Hohenester, *Phys. Rev. A* 66, 053811 (2002).
221. D. Zeidler, S. Frey, K.-L. Kompa, and M. Motzkus, *Phys. Rev. A* 64, 023420 (2001).
222. G. Chen, T. H. Stievater, E. T. Batteh, X. Li, D. G. Steel, D. Gammon, D. S. Katzer, D. Park, and L. J. Sham, *Phys. Rev. Lett.* 88, 117901 (2002).
223. A. Imamoglu, D. D. Awschalom, G. Burkard, D. P. DiVincenzo, D. Loss, M. Sherwin, and A. Small, *Phys. Rev. Lett.* 83, 4204 (1999).
224. E. Pazy, E. Biolatti, T. Calarco, I. D'Amico, P. Zanardi, F. Rossi, and P. Zoller, *Europhys. Lett.* 62, 175 (2003).
225. U. Hohenester, F. Troiani, and E. Molinari, in "Radiation-Matter Interaction in Confined Systems" (L. C. Andreani, G. Benedek, and E. Molinari, Eds.), p. 25. Società Italiana di Fisica, Bologna, 2002.
226. F. Troiani, U. Hohenester, and E. Molinari, *Phys. Rev. B* 62, R2263 (2000).
227. E. Pazy, I. D'Amico, P. Zanardi, and F. Rossi, *Phys. Rev. B* 64, 195320 (2001).
228. U. Boscain, G. Charlot, J.-P. Gauthier, S. Guérin, and H.-R. Jauslin, *J. Math. Phys.* 43, 2107 (2002).
229. Z. Kis and S. Stenholm, *J. Mod. Opt.* 49, 111 (2002).
230. K. G. Kim and M. D. Girardeau, *Phys. Rev. A* 52, R891 (1995).
231. R. S. Judson and H. Rabitz, *Phys. Rev. Lett.* 68, 1500 (1992).
232. W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, "Numerical Recipes in C++: The Art of Scientific Computing," 2nd Edn. Cambridge University Press, Cambridge, UK, 2002.
233. L. C. Andreani, G. Panzarini, A. V. Kavokin, and M. R. Vladimirova, *Phys. Rev. B* 57, 4670 (1998).
234. S. L. McCall and E. L. Hahn, *Phys. Rev. Lett.* 18, 908 (1967).
235. S. L. McCall and E. L. Hahn, *Phys. Rev.* 183, 457 (1969).
236. H. M. Gibbs and R. E. Slusher, *Appl. Phys. Lett.* 18, 505 (1971).
237. R. E. Slusher and H. M. Gibbs, *Phys. Rev. A* 5, 1634 (1972).
238. L. Allen and J. H. Eberly, "Optical Resonance and Two-level Atoms." Wiley, New York, 1975.
239. G. Panzarini, U. Hohenester, and E. Molinari, *Phys. Rev. B* 65, 165322 (2002).
240. U. Hohenester, *Phys. Rev. B* 66, 245323 (2002).
241. S. Schneider, P. Borri, W. Langbein, U. Woggon, J. Förstner, A. Knorr, R. L. Sellin, D. Ouyang, and D. Bimberg, *Appl. Phys. Lett.* 83, 3668 (2003).
242. P. Shor, in "Proceedings of the 35th Annual Symposium on Foundations of Computer Science" (S. Goldwasser, Ed.), p. 124. IEEE Computer Society, Los Alamitos, 1994.
243. R. Laflamme, C. Miquel, J. P. Paz, and W. H. Zurek, *Phys. Rev. Lett.* 77, 198 (1996).
244. D. P. DiVincenzo, *Fortschritte der Physik* 48, 771 (2000).
245. D. P. DiVincenzo, D. Bacon, J. Kempe, G. Burkard, and K. B. Whaley, *Nature* 408, 339 (2000).

246. I. L. Chuang, L. M. K. Vandersypen, X. Zhou, D. W. Leung, and S. Lloyd, *Nature* 393, 143 (1998).
247. J. I. Cirac and P. Zoller, *Phys. Rev. Lett.* 74, 4091 (1995).
248. A. Barenco, D. Deutsch, A. Ekert, and R. Jozsa, *Phys. Rev. Lett.* 74, 4083 (1995).
249. E. Biolatti, R. C. Iotti, P. Zanardi, and F. Rossi, *Phys. Rev. Lett.* 85, 5647 (2000).
250. E. Biolatti, I. D'Amico, P. Zanardi, and F. Rossi, *Phys. Rev. B* 65, 075306 (2002).
251. P. Bianucci, A. Muller, C. K. Shih, Q. Q. Wang, Q. K. Xue, and C. Piermarocchi, *Phys. Rev. B* 69, 161303 (2004).
252. S. De Rinaldis, I. D'Amico, E. Biolatti, R. Rinaldi, R. Cingolani, and F. Rossi, *Phys. Rev. B* 65, 081309 (2002).
253. S. De Rinaldis, I. D'Amico, and F. Rossi, *Appl. Phys. Lett.* 81, 4236 (2002).
254. L. Quiroga and N. F. Johnson, *Phys. Rev. Lett.* 83, 2270 (1999).
255. B. W. Lovett, J. H. Reina, A. Nazir, and G. A. D. Briggs, *Phys. Rev. B* 68, 205319 (2003).
256. S. Sungu, K. Kobayashi, A. Shojiguchi, and M. Ohtsu, *Phys. Rev. B* 69, 115334 (2004).
257. D. Loss and D. P. DiVincenzo, *Phys. Rev. A* 57, 120 (1998).
258. M. S. Sherwin, A. Imamoglu, and T. Montroy, *Phys. Rev. A* 60, 3508 (1999).
259. M. Feng, I. D'Amico, P. Zanardi, and F. Rossi, *Phys. Rev. A* 67, 014306 (2003).
260. C. Piermarocchi, P. Chen, L. J. Sham, and D. G. Steel, *Phys. Rev. Lett.* 89, 167402 (2002).
261. I. Calarco, A. Datta, P. Fedichev, E. Pazy, and P. Zoller, *Phys. Rev. A* 68, 012310 (2003).
262. F. Troiani, *Solid State Commun.* 128, 147 (2003).
263. R. G. Unanyan, N. V. Vitanov, and K. Bergmann, *Phys. Rev. Lett.* 87, 137902 (2001).
264. Z. Kis and F. Renzoni, *Phys. Rev. A* 65, 032318 (2002).
265. P. Zhang, C. K. Chan, Q.-K. Xue, and X.-G. Zhao, *Phys. Rev. A* 67, 012312 (2003).
266. C. M. Tesch and R. de Vivie-Riedle, *Phys. Rev. Lett.* 89, 157901 (2002).
267. L. P. Kadanoff and G. Baym, "Quantum Statistical Mechanics." Benjamin, New York, 1962.
268. A. L. Fetter and J. D. Walecka, "Quantum Theory of Many-Particle Systems." McGraw-Hill, New York, 1971.
269. H. Haug and A. P. Jauho, "Quantum Kinetics in Transport and Optics of Semiconductors." Springer, Berlin, 1996.
270. V. M. Axt and S. Mukamel, *Rev. Mod. Phys.* 70, 145 (1998).
271. U. Hohenester, *Phys. Rev. B* 64, 205305 (2001).
272. R. McWeeny, "Methods of Molecular Quantum Mechanics." Academic, London, 1992.
273. M. Braskén, M. Lindberg, D. Sundholm, and J. Olsen, *Phys. Rev. B* 61, 7652 (2000).
274. S. Corni, M. Braskén, M. Lindberg, J. Olsen, and D. Sundholm, *Phys. Rev. B* 67, 085314 (2003).
275. S. M. Barnett and P. M. Radmore, "Methods in Theoretical Quantum Optics." Clarendon, Oxford, 1997.

CHAPTER 8

Nanometer-Scale Electromagnetic Field Fluctuations

C. Henkel
Institut für Physik, Universität Potsdam, Potsdam, Germany

CONTENTS

1. INTRODUCTION

1.1. Fields and Matter on the Nanometer Scale

The electromagnetic field is the basic medium that induces interactions between charged particles. At distances on the order of the atomic size, the interactions are dominated by the instantaneous Coulomb interaction and give rise to atomic orbitals, chemical bonding, and the formation of solid matter. Retardation becomes important on the wavelength scale and determines the power emitted via electromagnetic radiation.

ISBN: 1-58883-049-7

Handbook of Theoretical and Computational Nanotechnology
Edited by Michael Rieth and Wolfram Schommers
Volume 7; Pages (463–501)

In this chapter, we focus on intermediate spatial scales of at least a few nanometers. On this scale, solid matter can, with increasing accuracy, be described by continuum theory. We consider, in addition, globally neutral systems like atoms or molecules that probe solid nanostructures via the nearby electromagnetic field. To the lowest nonvanishing order in a multipole expansion, the coupling to the field is provided by the electric and magnetic dipole moments of the corresponding atomic and molecular charge and current densities. The overlap of electron wave functions is negligible on the nanometer scale, with the latter being much larger than the extension of atomic and molecular orbitals. The electromagnetic interactions are often dominated by resonant processes in the visible frequency range, where typical electronic transitions for atoms and molecules occur. The corresponding wavelengths are hundreds of nanometers.

On the nanoscale, electromagnetic interactions thus occur via the near field, the properties of which are distinct from the propagating plane waves that prevail at distances beyond the transition wavelength (in the far field). The fact that retardation is negligible does not mean, however, that the near field is free from dynamical processes. We show, on the contrary, that it fluctuates at a level that can be enhanced orders of magnitude above the far-field level at which the Planck formula for blackbody radiation applies.

In fact, the near field is also determined by evanescent or nonpropagating fields that are bound to the solid structures. The field fluctuations contain both a quantum and a thermal component, depending on the ratio between frequency and temperature. They induce transitions and loss processes in atoms or molecules that are brought into the near field.

Typical examples we focus on in this contribution occur in near-field optics and integrated-atom optics. In near-field optics, atomic or molecular particles provide a local probe of a nanostructured substrate with an exquisite spatial resolution. In integrated-atom optics, ultracold matter waves are trapped and manipulated in (sub)micrometer potentials close to nanostructured substrates. Field fluctuations perturb these traps and determine their stability and coherence time.

1.2. Quantum Electrodynamics of Mesoscopic Media

In this chapter, we review the electromagnetic interactions between atomic or molecular dipoles and a nanostructured material. Adopting a description on the superatomic scale, the material is modelled in terms of a locally homogeneous medium characterized by a permittivity ε and a permeability μ. This assumes a linear response of the material to an external field, which is valid for sufficiently weak fields. A nonlocal or anisotropic response is not excluded, however, and can be taken into account by material response functions that are wave-vector dependent or tensors.

This material description is sufficient to describe the dynamics of the electromagnetic field on the nanoscale, starting from the so-called macroscopic Maxwell equations [1]. Even for the quantization of this field theory, no additional material data are required; the natural assumption suffices that the dielectric and magnetic susceptibilities are causal. This also covers the realistic, but theoretically challenging, case of materials with a nonzero absorption. The quantum theory of these lossy systems becomes consistent when a reservoir is allowed into which the energy is dissipated. An effective model for this reservoir can be constructed from the imaginary part of the material's response functions [2].

In many situations, however, the reservoir is not explicitly needed because the relevant statistical information about the quantum and thermal fluctuations of the electromagnetic field can be formulated from the field equations themselves, using equilibrium thermodynamics. This is achieved by means of the fluctuation–dissipation theorem, which relates the correlation function to the Green function of the field [3, 4]. As an alternative description, one can enlarge the theory to include a more detailed model for the current densities in the material. Dissipation then occurs as a result of coupling to a phonon reservoir, for example [5].

Although matter can be adequately described by a continuum theory on the nanometer scale, atoms or molecules are small on this scale and can be modeled in terms of pointlike electric or magnetic dipole moments [6]. Their interaction with the electromagnetic field,

for example, via the scattered radiation power, thus provides local information about the field. This idea has fostered the development of single-molecule probes in near-field optics that potentially offer a very high spatial resolution compared to solid tips. The fluorescence of single molecules allows us to extract information about the local density states (LDOS) if the molecule is scanned in a controlled way through the near field of a nanostructured sample [7].

We give here a basic introduction to the dynamics of fluorescent emission and show that the LDOS gives access to properties of the sample itself. The spatial resolution is limited only by the distance of the scanning particle. Integrated-atom optics provides our second application [8–11]. We describe miniaturized electromagnetic traps for ultracold atomic gases, where the atomic magnetic dipoles realize a sensitive magnetometer for near-field fluctuations. These fluctuations determine the lifetime and coherence time of the trapped atoms.

1.3. Domains of Relevance

1.3.1. Subwavelength Optical Microscopy

As mentioned above, electromagnetic field fluctuations on the nanometer scale play an important role for scanning near-field optical microscopy (SNOM) with single-molecule probes [7]. In the optical frequency range and at room temperature, these fluctuations are dominated by quantum or vacuum noise, and quantum electrodynamics is required. A generic example is the calculation of the spontaneous decay rate of a two-level molecule [12–14]. For many applications, however, the problem can be reduced to a classical calculation using arguments from equilibrium thermodynamics. It then suffices to know the total radiation emitted by a pointlike oscillating dipole source [15].

This does not mean, however, that the position dependence of the fluorescence signal in a given experimental context allows us to recover in a simple way information about a nanostructured substrate. Indeed, if the molecule is optically excited, the intensity distribution of the excitation light shows spatial variations in addition to the spatially modulated decay rate that involves the LDOS [16]. Another difficulty is that close to an absorbing nanostructure, a molecule has separate radiative and nonradiative decay channels that have to be disentangled using different experimental signals. We review the conditions and observables that allow for a simple signal processing and outline analytical arguments that provide a link between the LDOS and the local optical properties of the substrate on the nanometer scale.

1.3.2. Integrated-Atom Optics

The field of atom optics has emerged in the last two decades [17]. It transposes basic concepts of light optics (beamsplitters, mirrors, diffraction gratings, interferometers, and lasers) to atomic matter or de Broglie waves. A twofold motivation continues to drive the rapid evolution of the field: the exploration of fundamental wave mechanics, and the demonstration of devices in which atoms perform better to photons or other particles. Atom interferometers, for example, are sensitive to rotations and gravitational forces. They give phase shifts that are larger by a factor $mc\lambda_L/h \sim 10^9 m[\text{amu}]\lambda_L[\mu\text{m}]$ compared to light interferometers with the same enclosed area. The accuracy achieved in current experiments indicates that tiny relativistic corrections or the gravitational constant G can be measured in the near future [18].

As in light optics, a major step toward practical applications is integration: Miniaturize the device and combine it with source and detection system in a single, preferably solid setup. A natural solution, inspired by electromagnetic optics, is to replace free space with a waveguide for the propagation of an atomic beam. In matter wave optics, this corresponds to a confining potential and can be implemented easily with electromagnetic fields [8–10, 19, 20]. One type of atom guide that is particularly relevant for miniaturization below the micron scale is based on the magnetic field minimum formed when the azimuthal field from a current filament is superimposed on a homogeneous field, as put forward by Frisch and Segré [21]. A whole network of integrated waveguides can be realized, with static magnetic fields generated by nanostructures written onto a solid substrate. This concept offers strong

confinement and highly versatile structure design and is potentially scalable when using nanofabrication technologies. For a review, see Refs. [9–11].

Electromagnetic fluctuations arise in these devices from thermal charge and current fluctuations inside the nanostructures. They perturb the cold trapped matter waves by inducing spin flip transitions and transitions between stationary trap eigenstates, which leads to heating and a loss of matter wave coherence. Building on the relevant electromagnetic noise spectra, we summarize here methods to compute the spin flip and heating timescales in typical integrated-atom traps near nanostructured surfaces.

1.3.3. Other Scanning Probe Techniques

Finally, electromagnetic field fluctuations are present in other areas of probe microscopy at the nanometer scale. They have been discussed as being an explanation of viscous forces observed in tunnelling and atomic force microscopy. Quantum fluctuations have been known for a long time to lead to conservative, attractive forces of the Van der Waals and Casimir type [22]. The transfer of heat on the nanoscale is an associated issue that is of significant technological interest for submicron information storage [23, 24]. Magnetic noise determines the signal-to-noise ratio and the ultimate resolution in magnetic resonance microscopy, which is essentially a local version of nuclear magnetic resonance imaging that strives towards the observation of single molecules; for example, in biological samples [25]. We briefly discuss issues related to mechanical forces on the nanoscale in Section 5.

2. MESOSCOPIC MODEL FOR FIELD FLUCTUATIONS

2.1. Macroscopic Maxwell Equations

Given a continuous description of matter, the electromagnetic field dynamics are governed by the macroscopic Maxwell equations. In terms of the electric field $\mathbf{E}$, the magnetic flux density $\mathbf{B}$, the electric displacement field $\mathbf{D}$, and the magnetic field $\mathbf{H}$, these equations read in the time domain (and in SI units) [1]

$$\nabla \cdot \mathbf{D} = \rho \tag{1}$$

$$\nabla \times \mathbf{H} - \partial_t \mathbf{D} = \mathbf{j} \tag{2}$$

$$\nabla \cdot \mathbf{B} = 0 \tag{3}$$

$$\nabla \times \mathbf{E} + \partial_t \mathbf{B} = \mathbf{0} \tag{4}$$

The sources of the field are the density of free electric charges ρ and the free current density $\mathbf{j}$. They are distinct from the bound charge and current densities inside the material, which are described by the fields $\mathbf{D}$ and $\mathbf{H}$. We focus in this contribution on materials that respond linearly to the field.

In this case, one has the material equations

$$\mathbf{D}(\mathbf{x}, t) = \varepsilon_0 \int_{-\infty}^{+\infty} \mathrm{d}\tau \chi_e(\mathbf{x}, \tau)\mathbf{E}(\mathbf{x}, t - \tau) \tag{5}$$

$$\mathbf{H}(\mathbf{x}, t) = \mu_0^{-1} \int_{-\infty}^{+\infty} \mathrm{d}\tau \chi_m(\mathbf{x}, \tau)\mathbf{B}(\mathbf{x}, t - \tau) \tag{6}$$

where ε_0 and μ_0 are the vacuum permittivity and permeability (with $\varepsilon_0\mu_0 c^2 = 1$) and $\chi_e(\mathbf{x}, \tau)$, $\chi_m(\mathbf{x}, \tau)$ denote the electric and magnetic susceptibilities [equal to $\delta(\tau)$ in vacuum]. The time convolution is a generic feature in a dispersive material. In the frequency domain using the convention

$$\mathbf{E}(\mathbf{x}, t) = \int_{-\infty}^{+\infty} \frac{\mathrm{d}\omega}{2\pi} \mathbf{E}(\mathbf{x}; \omega) \mathrm{e}^{-i\omega t} \tag{7}$$

one has

$$\mathbf{D}(\mathbf{x};\omega) = \varepsilon_0\varepsilon(\mathbf{x};\omega)\mathbf{E}(\mathbf{x};\omega) \tag{8}$$

$$\mathbf{H}(\mathbf{x};\omega) = \frac{\mathbf{B}(\mathbf{x};\omega)}{\mu_0\mu(\mathbf{x};\omega)} \tag{9}$$

where $\varepsilon(\mathbf{x};\omega)$ and $\mu(\mathbf{x};\omega)$ are the relative dielectric function (permittivity) and the permeability, respectively. In free space, $\varepsilon = \mu = 1$. For non-magnetic materials, $\mu = 1$. In an inhomogeneous material, one has to include the position dependence of the dielectric function. We recall that the material Eqs. (8) and (9) are actually valid only on sufficiently large spatial scales, where a nonlocal dielectric or magnetic response can be neglected. We discuss the more general case below, where examples for the dielectric function are also given.

In frequency space, the Maxwell Eqs. (1)–(4) can be written in the form

$$\varepsilon_0\nabla\cdot\varepsilon\mathbf{E} = -\nabla\cdot\mathbf{P} \tag{10}$$

$$\nabla\times\frac{\mathbf{B}}{\mu_0\mu} + \mathrm{i}\omega\varepsilon_0\varepsilon\mathbf{E} = -\mathrm{i}\omega\mathbf{P} + \nabla\times\mathbf{M} \tag{11}$$

$$\nabla\cdot\mathbf{B} = 0 \tag{12}$$

$$\nabla\times\mathbf{E} - \mathrm{i}\omega\mathbf{B} = \mathbf{0} \tag{13}$$

We have assumed that the sources of the field are globally neutral systems whose charge and current densities can be characterized by polarization and magnetization fields $\mathbf{P}$ and $\mathbf{M}$. These fields have two distinct contributions. For the pointlike electric ($\mathbf{d}$) and magnetic ($\boldsymbol{\mu}$) dipole moments of atoms and molecules, we have

$$\mathbf{P}_{\mathrm{ext}}(\mathbf{x},t) = \mathbf{d}(t)\delta(\mathbf{x}-\mathbf{r}) \tag{14}$$

$$\mathbf{M}_{\mathrm{ext}}(\mathbf{x},t) = \boldsymbol{\mu}(t)\delta(\mathbf{x}-\mathbf{r}) \tag{15}$$

where $\mathbf{r}$ is the position of the atom. The thermal excitations in the material can be modeled as random fields or as operators $\mathbf{P}_{\mathrm{fl}}$ and $\mathbf{M}_{\mathrm{fl}}$. The existence of these fluctuations is essential to establish thermal equilibrium between the material and the field. Their fluctuation spectrum can be expressed in terms of the material response functions with the help of the fluctuation–dissipation theorem, see Sections 2.2 and 2.4.3.

In the following text, we restrict ourselves to nonmagnetic materials and put $\mu = 1$.

2.2. Material Response Functions

In this section, we give typical models for the dielectric function $\varepsilon = \varepsilon(\mathbf{x};\omega)$ that are widely used for some classes of materials. It has to be noted that these forms are often approximations that are valid only in a limited range of frequencies.

At optical frequencies, the permittivity is related to the index of refraction by

$$\varepsilon = n^2 = n'^2 - \kappa^2 + 2\mathrm{i}n'\kappa \tag{16}$$

where n' and κ are the real and imaginary parts of the index, respectively. Both typically depend on frequency. In a homogeneous medium, the Maxwell equations can be solved by plane waves with a dispersion relation

$$\sqrt{\mathbf{k}^2} = \frac{\omega}{c}n = \frac{\omega}{c}\sqrt{\varepsilon(\omega)} \tag{17}$$

so that $\mathrm{Re}(c/n)$ is the phase velocity of light in the medium and $(\omega/c)\mathrm{Im}\,n = (\omega/c)\kappa$ is the absorption coefficient.

The dielectric function for a metal can often by described by an electron plasma model. In terms of the electron density N_e, the plasma frequency is given by $\Omega_p^2 = e^2 N_e/(m_e \varepsilon_0)$, and one has the Drude model

$$\varepsilon(\omega) = 1 - \frac{\Omega_p^2}{\omega(\omega + i\Gamma_e)} = \frac{\omega^2 + \Gamma_e^2 - \Omega_p^2}{\omega^2 + \Gamma_e^2} + i\frac{\sigma(\omega)}{\varepsilon_0 \omega} \tag{18}$$

where Γ_e is the damping rate of electron motion and the conductivity is $\sigma(\omega) = \varepsilon_0 \Gamma_e \Omega_p^2/(\omega^2 + \Gamma_e^2)$. The current density induced by an electric field inside the material is then given by the Ohm law $\mathbf{j}_{ind} = \sigma \mathbf{E}$. Only for $\omega > \Omega_p$ is the permittivity Eq. (18) positive and the metal transparent; this typically occurs in the ultraviolet. At lower frequencies, ε is complex with a negative real part so that the refractive index is essentially imaginary: The electromagnetic field is strongly damped inside the metal and cannot enter it. As a consequence, metals are good reflectors for light in the visible range. At frequencies well below the damping rate, the permittivity is essentially imaginary, and one often uses the approximation $\varepsilon(\omega) = 1 + i\sigma/(\varepsilon_0 \omega)$ with the direct current conductivity $\sigma = \varepsilon_0 \Omega_p^2/\Gamma_e$.

Metals also provide an example of a material with a nonlocal dielectric function. Only on a sufficiently large spatial scale can one consider that the motion of the electrons is instantaneously damped so that the dielectric response becomes local again. On a scale comparable to the mean free path or lower, however, the electron plasma reacts ballistically, and hence in a nonlocal way, to an applied electric field. The material Eq. (8) is then generalized to

$$\mathbf{D}(\mathbf{x}; \omega) = \varepsilon_0 \int d^3x' \varepsilon(\mathbf{x}, \mathbf{x} - \mathbf{x}'; \omega) \mathbf{E}(\mathbf{x}'; \omega) \tag{19}$$

This is the most general relationship allowed for linear materials. Note the dependence on both $\mathbf{x}$ and $\mathbf{x} - \mathbf{x}'$: this can describe different materials as $\mathbf{x}$ varies.

If we focus for the moment on the dependence on $\mathbf{x} - \mathbf{x}'$, a spatial Fourier transformation leads to

$$\mathbf{D}(\mathbf{k}, \omega) = \varepsilon(\mathbf{k}, \omega) \mathbf{E}(\mathbf{k}, \omega) \tag{20}$$

featuring a $\mathbf{k}$-dependent dielectric function (spatial dispersion). A simple diffusion model involves the mean free path l_{mf} and the Fermi velocity $v_F = (\hbar/m_e)(3N_e/\pi^2)^{1/3}$ via the diffusion coefficient $D_e = v_F l_{mf}/3$

$$\varepsilon(\mathbf{k}, \omega) = 1 + \frac{i\sigma/\varepsilon_0}{\omega + iD_e k^2} \tag{21}$$

For an inhomogeneous sample, one can allow for a position-dependent $\varepsilon(\mathbf{x}; \mathbf{k}, \omega)$; this is the Fourier transform of $\varepsilon(\mathbf{x}, \mathbf{x} - \mathbf{x}'; \omega)$ in Eq. (19) with respect to $\mathbf{x} - \mathbf{x}'$. Because mean free path and damping rate are related by $l_{mf}\Gamma_e \sim v_F$, the wave-vector-dependent term in Eq. (21) is a small correction compared to ω, provided $(kl_{mf})^2 \ll \omega/\Gamma_e$. At typical frequencies ($\omega \leq \Gamma_e$); this occurs on a spatial scale much larger than the mean free path. Note that the mean free path for typical metals is in the range of a few nanometers; it can be much larger in high-quality semiconductor samples. More details on the dielectric function of metals can be found in the review paper by Ford and Weber [26].

For nonmetallic, crystalline materials with oppositely charged ions in the unit cell, the dielectric function shows a resonance near the frequency of a lattice vibration where positive and negative ions move with opposite velocities (an "optical phonon"). This is described by the Lorentz–Drude model

$$\varepsilon(\omega) = \varepsilon_{opt} + \frac{\Omega_p^2}{\Omega_{res}^2 - \omega(\omega + i\Gamma_{res})} \tag{22}$$

that shows a Lorentzian absorption resonance in $\mathrm{Im}\,\varepsilon(\omega)$, centered at Ω_{res} and with a width Γ_{res}. In Eq. (22), the strength Ω_p^2 of the resonance is analogous to the plasma frequency, and ε_{opt} gives the permittivity at frequencies well above the resonance (typically, in the visible). Optical phonon resonances mainly occur in the infrared region, where Eq. (22) is a good approximation.

2.3. Properties of the Macroscopic Equations

2.3.1. Energy Conservation

From the Maxwell Eqs. (10)–(13), one obtains, by multiplication with the complex conjugate fields $\mathbf{E}^*$ and $\mathbf{H}^*$ and by simple manipulations, the energy conservation law

$$2\omega \,\mathrm{Im}[\mathbf{P}^* \cdot \mathbf{E}] + 2\mathrm{Re}[(\nabla \times \mathbf{M}^*) \cdot \mathbf{E}]$$
$$= 2\nabla \cdot \mathrm{Re}[\mathbf{E}^* \times \mathbf{H}] + 2\omega\{\varepsilon_0(\mathrm{Im}\,\varepsilon)\mathbf{E}^* \cdot \mathbf{E} + \mu_0(\mathrm{Im}\,\mu)\mathbf{H}^* \cdot \mathbf{H}\} \tag{23}$$

The left-hand side gives the mechanical work per unit time and unit volume done by the electric field on the polarization and magnetization currents. On the right-hand side, the first term contains the Poynting vector averaged over one oscillation period, $\mathbf{S} = 2\,\mathrm{Re}[\mathbf{E}^* \times \mathbf{H}]$, which describes the energy emitted per unit time and unit area; the second term involves the imaginary parts of the material response functions and gives the energy dissipated per unit time and unit volume in the material.

The relation Eq. (23) thus characterizes quantitatively how energy is not conserved, but transferred between different subsystems, and is eventually lost in the absorbing material. It is clear that such a theory cannot be formulated in a straightforward way in terms of a Hamiltonian. In addition, we note that the theory is consistent with equilibrium thermodynamics only if the loss of energy in the material is compensated for by an energy flux radiated by the medium. We shall show below that the material fluctuations $\mathbf{P}_{\mathrm{fl}}$ and $\mathbf{M}_{\mathrm{fl}}$ provide this energy flux in a straightforward way.

To illustrate the conversion of energy in macroscopic electrodynamics, consider the example of an oscillating electric dipole moment at position $\mathbf{r}$. Integrating Eq. (23) over all space, one gets

$$2\omega \,\mathrm{Im}[\mathbf{d}^* \cdot \mathbf{E}(\mathbf{r})] = P_{\mathrm{em}} + P_{\mathrm{abs}} \tag{24}$$

$$P_{\mathrm{em}} = \oint \mathrm{d}^2\mathbf{a} \cdot \mathbf{S} \tag{25}$$

$$P_{\mathrm{abs}} = 2\omega \int \mathrm{d}^3V\{\varepsilon_0(\mathrm{Im}\,\varepsilon)\mathbf{E}^* \cdot \mathbf{E} + \mu_0(\mathrm{Im}\,\mu)\mathbf{H}^* \cdot \mathbf{H}\} \tag{26}$$

The imaginary part of the local electric field (i.e., evaluated at the position of the dipole) determines the total rate of emitted energy. The flux of the Poynting vector through a closed surface (with surface element $\mathrm{d}^2\mathbf{a}$) provides the radiation power P_{em} emitted into the far field. The integral over the volume occupied by the material gives the rate of energy absorption. This separation also applies in quantum theory, as we shall see below, and allows us to compute the radiative and nonradiative decay rates of a two-level system.

2.3.2. Reciprocity

Another property of the macroscopic Maxwell equations is their symmetry under the exchange of source and detector. This symmetry is also known as Onsager's reciprocity relations [27]. In optics, it corresponds to the reversal of the light path [1, 28]. It holds as long as permittivity and permeability are symmetric tensors, which is the case for isotropic media. An exception are Faraday-active media in a static magnetic field, which show polarization rotation.

Consider two monochromatic polarization sources $\mathbf{P}_1$ and $\mathbf{P}_2$ that are bounded in space and that generate the fields $\mathbf{E}_s$ and $\mathbf{H}_s$ ($s = 1, 2$). We assume that the sources and fields correspond to the same surroundings [i.e., the same medium response functions $\varepsilon(\mathbf{x})$ and $\mu(\mathbf{x})$ are used]. For simplicity, we assume that no magnetization sources are present. Simple manipulations starting from the Maxwell Eqs. (11) and (13) in frequency space then lead to

$$-\mathrm{i}\omega\varepsilon_0(\mathbf{E}_1 \cdot \varepsilon\mathbf{E}_2 - \mathbf{E}_2 \cdot \varepsilon\mathbf{E}_1) - \mathrm{i}\omega\mu_0(\mathbf{H}_1 \cdot \mu\mathbf{H}_2 - \mathbf{H}_2 \cdot \mu\mathbf{H}_1) + \nabla \cdot (\mathbf{E}_1 \times \mathbf{H}_2 - \mathbf{E}_2 \times \mathbf{H}_1)$$
$$= \mathrm{i}\omega(\mathbf{P}_1 \cdot \mathbf{E}_2 - \mathbf{P}_2 \cdot \mathbf{E}_1) \tag{27}$$

The first line vanishes provided ε and μ are symmetric. One can show as well that the space integral of the remaining terms on the left-hand side vanishes: integrate both sides over a large spherical volume enclosing the polarization sources and the dielectric material. By Gauss' theorem, the integral reduces to a surface integral with outward unit normal $\hat{\mathbf{x}}$, say. Now take into account that in the far field, the electric and magnetic fields are mutually perpendicular and perpendicular to the observation direction: $\hat{\mathbf{x}} \times \mathbf{E}_s = c\mu_0\mu\mathbf{H}_s$ $(s = 1, 2)$. Using again the symmetry of μ, we find that the integrand of the surface integral $(\hat{\mathbf{x}} \times \mathbf{E}_1) \cdot \mathbf{H}_2 - (\hat{\mathbf{x}} \times \mathbf{E}_2) \cdot \mathbf{H}_1$ vanishes. The reciprocity principle finally takes the form

$$0 = \int \mathrm{d}V(\mathbf{P}_1 \cdot \mathbf{E}_2 - \mathbf{P}_2 \cdot \mathbf{E}_1) \tag{28}$$

Note that no complex conjugation is required here, in contrast to the energy conversion law (26). A useful special case is provided by dipole sources $\mathbf{d}_s$ located at positions $\mathbf{r}_s$. The volume integral Eq. (28) then reduces to

$$\mathbf{d}_2 \cdot \mathbf{E}_1(\mathbf{r}_2) = \mathbf{d}_1 \cdot \mathbf{E}_2(\mathbf{r}_1) \tag{29}$$

The field radiated by a source dipole $\mathbf{d}_1$ observed at a position $\mathbf{r}_2$, $\mathbf{E}_1(\mathbf{r}_2)$, is thus identical to the field $\mathbf{E}_2(\mathbf{r}_1)$, such that a "reciprocal" dipole located at the observation point would create at the position $\mathbf{r}_1$ of the source dipole. For example, the field emitted by a dipole into the far field can be computed by putting a fictitious source into the far field and calculating the field it would create at the dipole. Sometimes one of these two calculations is easier to perform, and reciprocity thus reduces the computational effort.

2.3.3. Green Tensor

A widespread technique to solve the macroscopic Maxwell equations for an arbitrary polarization source is based on the Green function. Because the Maxwell equations are linear in the source fields, the radiated field can be written as the integral (we have again put $\mathbf{M} = 0$ here)

$$\mathbf{E}(\mathbf{x}; \omega) = \int \mathrm{d}^3x' \, \mathcal{G}(\mathbf{x}, \mathbf{x}'; \omega) \cdot \mathbf{P}(\mathbf{x}'; \omega) \tag{30}$$

In general, the Green function $\mathcal{G}$ is a tensor and depends separately on the source and observation points $\mathbf{x}'$, $\mathbf{x}$. By construction, we can interpret $\mathcal{G}(\mathbf{x}, \mathbf{x}'; \omega) \cdot \mathbf{d}$ as the electric field created at the position $\mathbf{x}$ by a monochromatic point dipole source $\mathbf{d}$ located at $\mathbf{x}'$. The Green tensor must therefore satisfy the same boundary conditions as any physical solution of the wave equations: It reduces at infinity to an outgoing spherical wave. We shall peruse the same shorthand notation as before and suppress the frequency argument ω whenever no confusion is possible.

The reciprocity theorem Eq. (29) implies that the Green tensor is symmetric. In terms of cartesian components, we have

$$\mathcal{G}_{ij}(\mathbf{x}, \mathbf{x}') = \mathcal{G}_{ji}(\mathbf{x}', \mathbf{x}) \tag{31}$$

In the time domain, the Green tensor provides a linear response function that connects the dipole moment, a real quantity, to the real electric field (for this reason, some authors call it the field susceptibility). In the frequency domain, it is therefore constrained to satisfy the relation

$$[\mathcal{G}_{ij}(\mathbf{x}, \mathbf{x}'; \omega)]^* = \mathcal{G}_{ij}(\mathbf{x}, \mathbf{x}'; -\omega^*) \tag{32}$$

where * means complex conjugation and we have allowed for complex ω.

Finally, consider again an oscillating electric point dipole at position $\mathbf{x}$ and express the local field $\mathbf{E}(\mathbf{x})$ in the energy conversion law Eq. (24) in terms of the Green tensor. The sum of emitted and absorbed power can then be written as

$$P_{\mathrm{em}} + P_{\mathrm{abs}} = 2\omega \, \mathrm{Im}\,[\mathbf{d}^* \cdot \mathcal{G}(\mathbf{x}, \mathbf{x}; \omega) \cdot \mathbf{d}] \tag{33}$$

Taking a linearly polarized dipole along the x_i coordinate, we thus observe that $\omega \operatorname{Im} \mathcal{G}_{ii}(\mathbf{x}, \mathbf{x}; \omega) \geq 0$. We can also normalize the power lost by the dipole by the energy quantum $\hbar\omega$ and introduce a rate of photon emission. Equation (33) then becomes very similar to Fermi's golden rule for the spontaneous decay rate of a two-level atom with dipole matrix element $\mathbf{d}$ (see Section 3.1 for a detailed discussion), and this motivates the following definition for the projected LDOS (PLDOS) of the electro-magnetic field

$$\text{PLDOS} = C \operatorname{Im} [\mathbf{u}^* \cdot \mathcal{G}(\mathbf{x}, \mathbf{x}; \omega) \cdot \mathbf{u}] \tag{34}$$

where the unit vector $\mathbf{u}$ specifies the polarization of the photon states (complex $\mathbf{u}$ corresponds to circular polarization) and C is a normalization constant. The local density of states LDOS and the global DOS (per unit volume) are then given by

$$\text{LDOS} = C \operatorname{Im} \operatorname{tr} \mathcal{G}(\mathbf{x}, \mathbf{x}; \omega) \tag{35}$$

$$\text{DOS} = \frac{C}{V} \int_V \mathrm{d}^3 x \operatorname{Im} \operatorname{tr} \mathcal{G}(\mathbf{x}, \mathbf{x}; \omega) \tag{36}$$

where tr denotes the trace and V is the volume of the sample. We give these definitions here because they are widely used.

Some subtle points are related to the fact that even in a dielectric without absorption, the Green tensor solves a wave equation with a nonhermitean differential operator, are discussed by Sprik, van Tiggelen, and Lagendijk [29] and by Sakoda [30]. Note also that the definition [Eq. (36)] of the DOS gives the electric field a distinguished role relative to the magnetic field. In free space, this does not lead to inconsistencies because the electric and magnetic energy densities are the same. (The magnetic Green tensor leads to the same DOS.) A different behavior occurs in the near field of a metallic structure, where the spectrum of the electromagnetic energy density is dominantly magnetic for some range of subwavelength distances (see Joulain et al. [31] and Section 2.5).

Let us illustrate the above properties with the aid of the Green tensor in free space. From the Maxwell equations, we find that it solves the wave equation

$$\nabla \times \nabla \times \mathcal{G}(\mathbf{x}, \mathbf{x}'; \omega) - \frac{\omega^2}{c^2} \mathcal{G} = -\frac{\omega^2 \mathbb{I}}{\varepsilon_0 c^2} \delta(\mathbf{x} - \mathbf{x}') \tag{37}$$

where $\mathbb{I}$ is the unit tensor. As a result of translation symmetry in free space, the Green tensor can only depend on $\mathbf{r} \equiv \mathbf{x} - \mathbf{x}'$. The solution that asymptotically reduces to an outgoing spherical wave is given by

$$\mathcal{G}(\mathbf{r}; \omega) = \frac{\mathbb{I}}{3\varepsilon_0} \delta(\mathbf{r}) + \frac{\omega^2 \, \mathrm{e}^{\mathrm{i}\omega r/c}}{4\pi\varepsilon_0 c^2} \left[\frac{\mathbb{I} - \hat{\mathbf{r}}\hat{\mathbf{r}}}{r} + (\mathbb{I} - 3\hat{\mathbf{r}}\hat{\mathbf{r}}) \left(\frac{\mathrm{i}}{\omega r^2/c} - \frac{1}{\omega^2 r^3/c^2} \right) \right] \tag{38}$$

where $\hat{\mathbf{r}}$ is the unit vector along $\mathbf{r}$, $r = |\mathbf{r}|$, and $\hat{\mathbf{r}}\hat{\mathbf{r}}$ denotes a tensor product.

We observe that this particular Green tensor is symmetric [Eq. (31)] and changes into its complex conjugate when the sign of the (real) frequency ω is flipped [Eq. (32)]. To compute the local density of states [Eq. (34)], one has to be careful when taking the limit $r \to 0$. De Vries, van Coevorden, and Lagendijk discuss a regularization of the Green tensor in terms of cutoffs in momentum space [32]. A simpler procedure, used, for example, by Tomaš and Lenac [33], uses angular averaging. Rotational symmetry means that we expect $\mathcal{G}(\mathbf{r} \to \mathbf{0})$ not to depend on the direction of $\mathbf{r}$. Performing an angular average of Eq. (38), we observe that $\hat{\mathbf{r}}\hat{\mathbf{r}}$ averages to $\frac{1}{3}\mathbb{I}$ so that the singular $1/r^2$ and $1/r^3$ terms drop out. Taking the imaginary part, the δ singularity drops out as well. Expanding the spherical wave $\mathrm{e}^{\mathrm{i}\omega r/c}$ to first order in $\omega r/c$, we finally get [32, 33]

$$\lim_{r \to 0} \operatorname{Im} \mathcal{G}(\mathbf{r}; \omega) = \frac{\omega^3 \mathbb{I}}{6\pi\varepsilon_0 c^3} \tag{39}$$

This LDOS does not depend on the observation point, of course. Compared to the well-known DOS in free space $\omega^2/(\pi^2 c^3)$, we can thus fix the normalization factor in Eqs. (34)–(36) to $C = 2\varepsilon_0/(\pi\omega)$. Note that different normalizations occur in the literature, depending on the convention used for the prefactor of the source term in the wave Eq. (37).

2.3.4. Heisenberg Uncertainty Relations

Spatial resolution is a key question for electrodynamics on the subwavelength scale, and its fundamental limits are given by the uncertainty relations. The standard diffraction limit applies to fields formed by propagating plane waves, which is a consequence of the well-known relation

$$\Delta x \Delta k \sim 1 \tag{40}$$

between the resolution Δx and the width Δk of the field in reciprocal space. Because $\Delta k \leq 2\pi n/\lambda$ for plane waves in a homogeneous medium with refractive index n, Eq. (40) implies a resolution Δx on the order of the medium wavelength λ/n [28].

Near-field optics break the diffraction limit by using subwavelength detectors or scatterers that are sensitive to evanescent waves as well [34]. For example, the field

$$\mathbf{E}(\mathbf{x}; \omega) = \mathbf{e} \exp(ik_\parallel x - z/l_z) \tag{41}$$

with $k_\parallel^2 - l_z^{-2} = (\omega/c)^2$ and $\mathbf{e} \cdot (k_x \mathbf{e}_x + il_z^{-1}\mathbf{e}_z) = 0$ is a solution to the Maxwell equations in the vacuum half space $z \geq 0$ that is bound to a solid structure located in $z < 0$. The wavevector k_x is not limited by $2\pi/\lambda$, and a field formed by evanescent waves (41) can therefore contain spatial information on much finer scales. For example, at a nanostructured interface, the boundary conditions can only be satisfied by including evanescent waves in the field. Because of their exponential decay with distance z, however, the subwavelength information they carry is lost as the detector is withdrawn from the near field of the structure, and the standard diffraction limit is recovered.

The relation Eq. (40) corresponds to the Heisenberg uncertainty relations between the noncommuting observables position and momentum in quantum mechanics. For the electromagnetic field in second quantization, another uncertainty relation involves the electric and magnetic fields or, given a single field mode, two orthogonal field quadratures [12, 14]. The commutator between the electric and magnetic field operators is given by (summation over k is understood)

$$[E_i(\mathbf{x}, t),\ B_j(\mathbf{x}', t)] = -\frac{i\hbar}{\varepsilon_0} \epsilon_{ijk} \frac{\partial}{\partial x_k} \delta(\mathbf{x} - \mathbf{x}') \tag{42}$$

so that electric and magnetic fields cannot be measured simultaneously. This commutator applies also to the macroscopic Maxwell equations, irrespective of the gauge chosen for the quantization procedure [2]. In the visible frequency range, however, one is actualy not sensitive to the field components that occur in Eq. (42). Depending on the type of detector or scatterer used, the relevant observables are rather the time-averaged intensities $\overline{\mathbf{E}^2}$, $\overline{\mathbf{B}^2}$ of the electric or magnetic field in some finite frequency band and smeared out over some spatial region. For these, an estimate based on Eq. (42) yields a fundamental limit that corresponds to about one photon energy per spatially resolved volume when the field intensities are expressed in units of energy densities. This reflects the detection of the quantized field in terms of discrete energy packets.

2.4. Field Fluctuation Spectra

The fluctuations of the electromagnetic fields can be characterized by statistical electrodynamics, which result from the application of equilibrium thermodynamics and quantum theory to the macroscopic Maxwell equations [35]. We introduce in this section the basic definitions for the field fluctuation spectra and review how they can be calculated. A key result in this context is the fluctuation dissipation theorem derived by Callen and Welton [4], whose proof is sketched here. We conclude with some remarks on how to handle nonequilibrium situations.

2.4.1. Basic Definition

We shall assume that at thermodynamic equilibrium at temperature T, the field and the solid medium can be described by a Gibbs ensemble: each state with energy E is weighted with the Boltzmann factor $\exp(-E/k_B T) \equiv e^{-\beta E}$. In the quantized version of the theory, these states are stationary states: they are eigenstates with energy E of the corresponding Hamilton operator $\widehat{H}$. We shall work in the Heisenberg picture where the field variables (called observables) evolve in time while the state of the system is fixed. The Gibbs ensemble is then characterized by the density operator

$$\hat{\rho}_{\rm eq} = \frac{\exp(-\widehat{H}/k_B T)}{\mathrm{Tr}\,\exp(-\widehat{H}/k_B T)} \tag{43}$$

This is an operator on the Hilbert space of the system that can be represented by a (infinite-dimensional) matrix, the density matrix. The trace of the operator-valued Boltzmann factor is $\mathrm{Tr}\,\exp(-\widehat{H}/k_B T)$; it is also called the partition function [3].

With respect to this equilibrium ensemble, we can define average values for the observables of interest. This average combines the quantum expectation value in a given stationary state with the corresponding statistical ensemble weights. The average electric field, for example, is given by

$$\langle \mathbf{E}(\mathbf{x}, t)\rangle \equiv \mathrm{Tr}[\mathbf{E}(\mathbf{x}, t)\hat{\rho}_{\rm eq}] = \mathrm{Tr}[\hat{\rho}_{\rm eq}\mathbf{E}(\mathbf{x}, t)] \tag{44}$$

where we have made use of the cyclic permutation under the trace. The time-dependence of the field is generated by the Hamilton operator, so that we have

$$\langle \mathbf{E}(\mathbf{x}, t)\rangle = \mathrm{Tr}[\exp(\mathrm{i}\widehat{H}t/\hbar)\mathbf{E}(\mathbf{x})\exp(-\mathrm{i}\widehat{H}t/\hbar)\hat{\rho}_{\rm eq}] = \mathrm{Tr}[\mathbf{E}(\mathbf{x})\hat{\rho}_{\rm eq}] \tag{45}$$

where $\mathbf{E}(\mathbf{x})$ is the electric field operator at time zero and we have used the fact that the Gibbs density operator (43) is invariant under time evolution. The average can now be computed in the Schrödinger picture and is found to vanish at equilibrium. In the classical theory, this is because the phase of the field is uniformly distributed. In the quantum theory, the stationary states for each mode of the field (labelled by κ) are eigenstates of the photon number operator $a_\kappa^\dagger a_\kappa$. The field observable is a linear combination of the annihilation and creation operators a_κ and $a_\kappa^\dagger$ that lower or raise the photon number: Their quantum expectation values thus vanish in a photon number eigenstate. More details can be found in Section 2.4.2 and in the textbooks by Loudon [12] and Mandel and Wolf [14].

The relevant information about the field fluctuations is thus encoded in the correlation function

$$\langle \mathbf{E}(\mathbf{x}, t)\mathbf{E}(\mathbf{x}', t')\rangle \equiv \mathrm{Tr}[\mathbf{E}(\mathbf{x}, t)\mathbf{E}(\mathbf{x}', t')\hat{\rho}_{\rm eq}] = \langle \mathbf{E}(\mathbf{x}, 0)\mathbf{E}(\mathbf{x}', t' - t)\rangle \tag{46}$$

In the second step, we have shifted the time arguments, using the fact that the time evolution commutes with the equilibrium density operator [Eq. (43)]. As expected from stationarity, this correlation function only depends on the time difference $\tau = t' - t$. In the limit $|\tau| \to \infty$, one expects the fields $\mathbf{E}(\mathbf{x}, t)$ and $\mathbf{E}(\mathbf{x}', t')$ to decorrelate and the correlation function [Eq. (46)] to vanish. The timescale on which this happens gives the correlation or coherence time of the field.

The spectrum of the field fluctuations can be defined by the Fourier expansion of the correlation function

$$\langle \mathbf{E}(\mathbf{x}, t)\mathbf{E}(\mathbf{x}', t')\rangle = \int_{-\infty}^{+\infty} \frac{\mathrm{d}\omega}{2\pi} \mathrm{e}^{-\mathrm{i}\omega(t'-t)}\mathscr{E}(\mathbf{x}, \mathbf{x}'; \omega) \tag{47}$$

This relation is also known as the Wiener–Khintchine theorem [14]. The spectrum is actually a tensor use: the cross-spectral density tensor. For $\mathbf{x} = \mathbf{x}'$, its components specify the local polarization state of the field. As a function of $\mathbf{x} - \mathbf{x}'$, it characterizes the field's spatial coherence (i.e., the contrast of interference fringes in a double-slit experiment with slits placed by $\mathbf{x}$ and $\mathbf{x}'$; see, e.g., the textbooks by Mandel and Wolf [14] and by Goodman [36]).

As a function of the frequency ω, the spectrum specifies the strength of the field fluctuations. This can be seen from the equivalent relation for the Fourier transforms of the field

$$\langle \mathbf{E}(\mathbf{x}, -\omega)\mathbf{E}(\mathbf{x}', \omega')\rangle = 2\pi\delta(\omega - \omega')\,\mathcal{E}(\mathbf{x}, \mathbf{x}'; \omega') = \langle [\mathbf{E}(\mathbf{x}, \omega)]^{\dagger}\mathbf{E}(\mathbf{x}', \omega')\rangle \tag{48}$$

In the last step, we have made use of the reality of the electric field that leads to the analog of Eq. (32). This way of writing also shows that the diagonal elements of $\mathcal{E}(\mathbf{x}, \mathbf{x}; \omega)$ are positive. More generally, $0 \le \mathbf{u}^* \cdot \mathcal{E}(\mathbf{x}, \mathbf{x}; \omega) \cdot \mathbf{u}$ for any complex vector $\mathbf{u}$. We note that the Fourier transforms of the fields strictly speaking do not exist as ordinary functions. The Fourier calculus nevertheless applies symbolically for the corresponding operator-valued distributions.

The fluctuation spectrum of the electric field plays a key role for spontaneous and stimulated decay on the electric dipole transitions of an atom or molecule. The corresponding spectrum for the magnetic field characterizes the perturbation the field exerts on an atomic magnetic moment or spin. This is discussed in detail in Sections 3 and 4.

2.4.2. Blackbody Fluctuations

Let us illustrate the correlation function introduced above with the example of the blackbody radiation field in free space. The electric field operator can, in that case, be expanded in plane wave modes [12, 14]

$$\mathbf{E}(\mathbf{x}, t) = \sum_{\kappa} \sqrt{\frac{\hbar\omega(k)}{2\varepsilon_0 V}}\,[a_{\kappa}(t)\varepsilon \exp(\mathrm{i}\mathbf{k}\cdot\mathbf{x}) + \mathrm{HC}] \tag{49}$$

where the mode label $\kappa = (\mathbf{k}, \varepsilon)$ combines the wave vector $\mathbf{k}$ and the polarization vector $\varepsilon \perp \mathbf{k}$, and V is the quantization volume. We assume periodic boundary conditions so that the allowed wave vectors are discrete. The mode frequency is given by $\omega(k) = ck = c\sqrt{\mathbf{k}^2}$. The term HC denotes the hermitean conjugate operator, so that the electric field is globally hermitean. In the absence of any sources, the annihilation and creation operators evolve according to

$$a_{\kappa}(t) = a_{\kappa}\exp[-\mathrm{i}\omega(k)t] \qquad a_{\kappa}^{\dagger}(t) = a_{\kappa}^{\dagger}\exp[\mathrm{i}\omega(k)t] \tag{50}$$

The Schrödinger operators a_{κ} and $a_{\kappa}^{\dagger}$ satisfy the bosonic commutation relations

$$[a_{\kappa}, a_{\kappa'}^{\dagger}] \equiv a_{\kappa}a_{\kappa'}^{\dagger} - a_{\kappa'}^{\dagger}a_{\kappa} = \delta_{\kappa,\kappa'} \equiv \delta_{\mathbf{k},\mathbf{k}'}\delta_{\varepsilon,\varepsilon'} \tag{51}$$

The equilibrium expectation value of products of the mode operators decorrelates for different modes because the density operator factorizes into a product over all modes. Therefore, the expectation value $\langle a_{\kappa'}^{\dagger}a_{\kappa}\rangle$ vanishes for $\kappa \neq \kappa'$.

For a given mode κ, one gets from the Gibbs ensemble the Bose–Einstein occupation number [with $\beta \equiv 1/(k_{\mathrm{B}}T)$]

$$\langle a_{\kappa}^{\dagger}a_{\kappa}\rangle = \bar{n}[\omega(k)] \equiv \frac{1}{\mathrm{e}^{\beta\hbar\omega(k)} - 1} \tag{52}$$

$$\langle a_{\kappa}a_{\kappa}^{\dagger}\rangle = 1 + \bar{n}[\omega(k)] = \frac{1}{1 - \mathrm{e}^{-\beta\hbar\omega(k)}} \tag{53}$$

whereas the products $a_{\kappa}a_{\kappa}$ and $a_{\kappa}^{\dagger}a_{\kappa}^{\dagger}$ have zero average. To prove Eq. (52), we evaluate the trace in the eigenbasis $|n_{\kappa}\rangle$ of the photon number operator $a_{\kappa}^{\dagger}a_{\kappa}$ for the given mode. The energy eigenvalue of $|n_{\kappa}\rangle$ is $\hbar\omega(k)(n_{\kappa} + \frac{1}{2})$, and the summation over the Boltzmann weights gives

$$\begin{aligned}\langle a_{\kappa}^{\dagger}a_{\kappa}\rangle &= \frac{\sum_{n_{\kappa}=0}^{\infty} n_{\kappa}\mathrm{e}^{-\beta\hbar\omega(k)(n_{\kappa}+\frac{1}{2})}}{\sum_{n_{\kappa}=0}^{\infty}\mathrm{e}^{-\beta\hbar\omega(k)(n_{\kappa}+\frac{1}{2})}} \\ &= -\frac{\partial}{\partial\xi}\log\left(\sum_{n=0}^{\infty}\mathrm{e}^{-\xi n}\right)\Bigg|_{\xi=\beta\hbar\omega(k)} = \frac{\partial}{\partial\xi}\log(1 - \mathrm{e}^{-\xi})\big|_{\xi=\beta\hbar\omega(k)} \\ &= \frac{1}{\mathrm{e}^{\beta\hbar\omega(k)} - 1}\end{aligned} \tag{54}$$

Using this information, a straightforward calculation in the continuum limit $\sum_{\mathbf{k}} \mapsto V \int \mathrm{d}^3k/(2\pi)^3$ leads to

$$\mathcal{K}(\mathbf{x}, \mathbf{x}+\mathbf{r}; \omega) = \frac{\hbar\omega^3 \bar{n}(\omega)}{2\pi\varepsilon_0 c^3} \left\{ (\mathbb{I} - \hat{\mathbf{r}}\hat{\mathbf{r}}) \frac{\sin(\omega r/c)}{\omega r/c} + (\mathbb{I} - 3\hat{\mathbf{r}}\hat{\mathbf{r}}) \left(\frac{\cos(\omega r/c)}{(\omega r/c)^2} - \frac{\sin(\omega r/c)}{(\omega r/c)^3} \right) \right\} \tag{55}$$

which only depends on the difference vector $\mathbf{r} = \mathbf{x}' - \mathbf{x}$, as expected. This expression is very similar to the Green tensor [Eq. (38)], which is not a coincidence, but a special case of the fluctuation–dissipation (FD) theorem introduced in Section 2.4.3.

In the limit $\mathbf{r} \to \mathbf{0}$, one can check that $\mathcal{K}(\mathbf{x}, \mathbf{x}; \omega)$ is proportional to the unit tensor and positive. Taking the trace, one gets, up to a factor $\varepsilon_0/2$, the spectrum of the electric energy density $u_{\mathrm{e}}(\omega)$. Summing the contributions of positive and negative frequencies, the electric energy density is given by the Planck formula

$$U_{\mathrm{bb,e}} = \int_0^\infty \frac{\mathrm{d}\omega}{2\pi} u_{\mathrm{bb,e}}(\omega) \tag{56}$$

$$u_{\mathrm{bb,e}}(\omega) = \frac{\hbar\omega^3(\bar{n}(\omega) + \frac{1}{2})}{\pi c^3} = 2\pi \mathrm{DOS}(\omega) \frac{\hbar\omega}{2} \left(\bar{n}(\omega) + \frac{1}{2} \right) \tag{57}$$

In the last expression, we have made use of the free space DOS $= \omega^2/\pi^2c^3$. This result has an intuitive explanation: The electric energy density is the density of modes per unit volume $\mathrm{DOS}(\omega)\mathrm{d}\omega$ multiplied by one half of the average equilibrium energy $\hbar\omega(\bar{n}(\omega) + \frac{1}{2})$ of a harmonic oscillator (the other half contributes to the magnetic energy).

At zero temperature, the electric plus magnetic energy per mode is given by the ground-state oscillator energy $\hbar\omega/2$, which leads to a divergent integral in the UV. From this divergent zero-point energy, one can extract a finite, measurable energy difference, called the Casimir energy, when the mode functions are changed by the boundary conditions imposed by material structures. See Section 5 and the textbook by Mostepanenko and Trunov [22] for more details. The energy measured by a photodetector does not diverge because it is proportional to average of the photon number operator $\langle a_\kappa^\dagger a_\kappa \rangle = \bar{n}(\omega(k))$ [12, 14]. If the creation and annihilation operators are ordered in this way, the divergent contribution of the zero-point energy disappears. The blackbody spectrum then shows an exponential decrease beyond the thermal wavelength $\lambda_{\mathrm{th}} = 2\pi\hbar c/(k_{\mathrm{B}}T)$ (the Wien displacement law), and the frequency integral [Eq. (56)] becomes convergent.

2.4.3. Fluctuation–Dissipation Theorem

The previous example indicates that there is a relation between the field fluctuation spectrum and the Green function. With the previously introduced notation, we have

$$\mathcal{K}(\mathbf{x}, \mathbf{x}'; \omega) = \frac{2\hbar}{\mathrm{e}^{\beta\hbar\omega} - 1} \mathrm{Im}\, \mathcal{G}(\mathbf{x}, \mathbf{x}'; \omega) \tag{58}$$

where the Green tensor is defined by Eq. (30) and the fluctuation spectrum by Eq. (47). Equation (58) is actually true under more general conditions and is known as a fluctuation-dissipation (FD) theorem [4]. It holds for linear systems and their fluctuations around the thermal equilibrium state. The dissipation is encoded in the imaginary part of the response function that characterizes the linear response of the system to an external perturbation. The FD theorem is of the form given here provided the Green tensor satisfies the symmetry condition specified in Eq. (31).

The FD theorem will be our basic tool to compute field fluctuation spectra in the near field of nanostructures. We give in this section an introduction for linear systems and summarize a general proof in the context of statistical electrodynamics. A generalization to nonequilibrium situations is discussed in Section 2.4.4.

2.4.3.1. Johnson-Nyquist Noise in Metals Consider a resistance at temperature T. One observes a thermal fluctuation of the current through the resistance, called Johnson noise, whose variance in a given bandwidth $\Delta f = \Delta\omega/(2\pi)$ is given by the Nyquist formula

$$\langle \delta I^2 \rangle_{\Delta\omega} = \frac{\Delta\omega}{2\pi} \frac{4k_B T}{R(\omega)} \tag{59}$$

We now show that this formula can be related to a fluctuation dissipation theorem for the current density $\mathbf{j}(\mathbf{x})$ of the resistance. Consider first the z-component of the current density and a small volume element $\Delta V = \Delta z \Delta A$. From the current noise along the z-direction, one then has

$$\langle \delta j_z(\mathbf{x})^2 \rangle_{\Delta\omega} = \frac{\Delta\omega}{2\pi} \frac{4k_B T}{R(\omega)\Delta A^2} = \frac{\Delta\omega}{2\pi} \frac{4k_B T\sigma(\omega)}{\Delta V} \tag{60}$$

where $\sigma(\omega) = \Delta z/[R(\omega)\Delta A]$ is the conductivity (the inverse of the specific resistance). In the low-frequency range where the Nyquist formula is valid, the conductivity can be expressed via the dielectric function of the resistance $\varepsilon_0\varepsilon(\omega) = \varepsilon_0 + i\sigma/\omega$.

The result [Eq. (60)] can thus be obtained by averaging the following relation over the volume element

$$\langle \delta\mathbf{j}(\mathbf{x})\delta\mathbf{j}(\mathbf{x}') \rangle_{\Delta\omega} = \frac{\Delta\omega}{2\pi} 4k_B T\omega \, \mathbb{I}\,\mathrm{Im}\, \varepsilon_0\varepsilon(\mathbf{x};\omega)\delta(\mathbf{x}-\mathbf{x}') \tag{61}$$

We have assumed that neighboring volume elements have uncorrelated current noise; hence the spatial delta function. We use here the convention that the current noise [Eq. (61)] is given by the integral of the noise spectrum $\mathcal{J}(\mathbf{x},\mathbf{x}';\omega)$ over intervals $\Delta\omega/(2\pi)$ centered at positive and negative frequencies $\pm\omega$. Because $\omega\,\mathrm{Im}\,\varepsilon(\mathbf{x};\omega)$ is an even function of ω, the noise spectrum is given by

$$\mathcal{J}(\mathbf{x},\mathbf{x}';\omega) = 2k_B T\omega \, \mathbb{I}\,\mathrm{Im}\, \varepsilon_0\varepsilon(\mathbf{x};\omega)\delta(\mathbf{x}-\mathbf{x}') \tag{62}$$

This result already has the structure of the FD theorem [Eq. (58)]. Because the dielectric function gives the polarization induced by an electric field, a more natural formulation is in terms of the polarization noise spectrum [writing $\mathbf{j}(\omega) = -i\omega\mathbf{P}(\omega)$]

$$\mathcal{P}(\mathbf{x},\mathbf{x}';\omega) = \frac{2k_B T}{\omega} \, \mathbb{I}\,\mathrm{Im}\, \varepsilon_0\varepsilon(\mathbf{x};\omega)\delta(\mathbf{x}-\mathbf{x}') \tag{63}$$

The temperature-dependent prefactor is the low-frequency limit of $2\hbar/(e^{\hbar\omega/k_B T} - 1)$ occurring in Eq. (58). The spatial δ-function applies to a local dielectric response. The generalization to a nonlocal medium is immediate: the fluctuations are then correlated on some characteristic scale (typically the mean free path).

We thus find that the strength of the thermal Johnson noise at a given frequency in an absorbing material is related to the amount of dissipation, as encoded in the imaginary part $\mathrm{Im}\,\varepsilon(\omega)$. This permits us to characterize the thermal polarization and magnetization fluctuations that appear in the macroscopic Maxwell Eqs. (10)–(13). The noise spectrum of the polarization noise $\mathbf{P}_{\mathrm{fl}}(\mathbf{x};\omega)$ is given by the Johnson–Nyquist formula [Eq. (63)], with the factor $k_B T$ replaced by $\hbar\omega/(e^{\beta\hbar\omega} - 1)$ to be valid at all frequencies. If the material is magnetic with a (local) permeability $\mu(\mathbf{x};\omega)$, it contains magnetization fluctuations $\mathbf{M}_{\mathrm{fl}}(\mathbf{x};\omega)$ with a spectrum

$$\mathcal{M}(\mathbf{x},\mathbf{x};\omega) = -\frac{2\hbar \, \mathbb{I}}{e^{\beta\hbar\omega} - 1} \mathrm{Im} \frac{1}{\mu_0\mu(\mathbf{x};\omega)} \delta(\mathbf{x}-\mathbf{x}') \tag{64}$$

The fluctuating material polarization radiates an electromagnetic field that, in thermodynamic equilibrium, compensates for the loss of electromagnetic energy inside the material. Only in this way is it possible to enforce the equipartition law of thermal equilibrium, with every degree of freedom (here the polarization field) carrying an energy $k_B T/2$.

Consider the average of the energy conservation law [Eq. (23)] for a system without external polarization sources. On the left-hand side, we get

$$\omega \operatorname{Im}\langle \mathbf{P}_{\mathrm{fl}}^{\dagger}(\mathbf{x};\omega)\cdot\mathbf{E}(\mathbf{x};\omega')\rangle = 2\pi\omega\delta(\omega-\omega')\int \mathrm{d}^3x'\,\mathcal{P}_{ij}(\mathbf{x},\mathbf{x}';\omega)\operatorname{Im}\mathcal{G}_{ij}(\mathbf{x},\mathbf{x}';\omega)$$

$$= 2\pi\delta(\omega-\omega')\frac{2\hbar\omega}{\mathrm{e}^{\beta\hbar\omega}-1}\operatorname{Im}\varepsilon_{\mathrm{fl}}\varepsilon(\mathbf{x};\omega)\operatorname{Im}\operatorname{Tr}\mathcal{G}(\mathbf{x},\mathbf{x};\omega) \quad (65)$$

We have taken into account that only the part of the field radiated by the polarization fluctuation is correlated with this fluctuation and have expressed that field in terms of the Green tensor [Eq. (30)]. A similar result holds for the magnetic contribution. On the right-hand side, the average over the electric losses leads to the same expression

$$\omega \operatorname{Im}[\varepsilon_0\varepsilon(\mathbf{x};\omega)]\langle \mathbf{E}^{\dagger}(\mathbf{x};\omega)\cdot\mathbf{E}(\mathbf{x};\omega')\rangle$$

$$= 2\pi\delta(\omega-\omega')\operatorname{Im}[\varepsilon_0\varepsilon(\mathbf{x};\omega)]\frac{2\hbar\omega}{\mathrm{e}^{\beta\hbar\omega}-1}\operatorname{Im}\operatorname{Tr}\mathcal{G}(\mathbf{x},\mathbf{x};\omega) \quad (66)$$

using the FD theorem [Eq. (58)]. As a result, the energy the polarization emits into the field [Eq.(65)] is exactly compensated for by the field energy lost by absorption, Eq. (66). This also implies that the average Poynting vector $\langle \mathbf{E}^{\dagger}\times\mathbf{H}\rangle$ vanishes, as there is no net energy transfer between medium and field.

We shall see that in the quantized theory, the polarization fluctuations of the material also contribute to the quantum fluctuations of the field. Otherwise, the dissipation present in the macroscopic Maxwell equations would force the field operators to decay to zero, including their commutators. These are preserved because of the quantum fluctuations of the material polarization. This picture indicates as well the existence of an FD theorem: the material loss that forces the fields to decay must be balanced by the fluctuations inside the material.

2.4.3.2. Properties of Quantum Field Fluctuations Before giving a general proof of the FD theorem [Eq. (58)], let us summarize some of the properties it implies for the equilibrium fluctuations of quantized fields.

The FD theorem allows us to compute the quantum and thermal fluctuations of the electromagnetic field once the Green tensor is known. This quantity can be computed by solving the macroscopic Maxwell equations with pointlike dipole sources. As long as the medium responds linearly to the field, the classical version of the theory is sufficient, and quantum and thermal fluctuations are handled self-consistently using the theorem.

The noise spectrum of a quantized field is not symmetric. It is proportional to the Bose–Einstein occupation number $\bar{n}(\omega)$ for positive frequencies and decays to zero for $\hbar\omega \gg k_{\mathrm{B}}T$. At negative frequencies, one finds, given that $\operatorname{Im}\mathcal{G}(\omega)$ is an odd function [see Eq. (32)], that the spectrum is proportional to $1+\bar{n}(|\omega|)$. The zero-point fluctuations appear here. The asymmetric frequency spectrum of zero-point or vacuum fluctuations can be understood qualitatively by noting that in the ground state, a system can only fluctuate via a virtual transition toward a state with higher energy. The corresponding Bohr frequencies are all positive. {That this leads to a spectral weight at negative ω is related to our—conventional—choice of the exponential factor in the noise spectrum [Eq. (47)].} In the high-temperature limit or, equivalently, for classical systems, upward and downward transitions occur with equal probability, and the fluctuation spectrum is symmetric: $\bar{n}(\omega)\approx\bar{n}(|\omega|)+1\approx k_{\mathrm{B}}T/\hbar\omega \gg 1$. For intermediate temperatures, we show below that the principle of detailed balance is satisfied, with upward and downward transition rates differing by a factor $\mathrm{e}^{\beta\hbar\omega}$.

At positive frequencies, Eq. (47) shows that the spectrum $\mathcal{E}(\mathbf{x},\mathbf{x}';\omega)$ picks out that part $\mathbf{E}^{(+)}(\mathbf{x}',t')$ of the electric field operator that evolves like $\mathrm{e}^{-\mathrm{i}\omega t'}$. By analogy to time-dependent wave functions in quantum mechanics, this part is called the positive frequency part of the field. In the mode expansion [Eq. (49)], it corresponds to the sum over the annihilation operators a_κ [see also Eq. (50)]. Similarly, only the negative frequency part $\mathbf{E}^{(-)}(\mathbf{x},t)=[\mathbf{E}^{(+)}(\mathbf{x},t)]^{\dagger}$ of the field operator contributes in the first factor of Eq. (47). It follows that in the vacuum state, the expectation value

$$\operatorname{Tr}\left[\mathbf{E}^{(-)}(\mathbf{x},t)\mathbf{E}^{(+)}(\mathbf{x}',t')\hat{\rho}_{\mathrm{vac}}\right]=0 \quad (67)$$

vanishes because the annihilation operators, by definition, give zero when acting on the vacuum state. This operator order (annihilation operators acting first) is usually called "normal" order. A typical example is the intensity measured by a photodetector [12, 14]. A nonzero vacuum expectation value occurs with the reverse operator order (creation operators acting first). This antinormal order is picked out for negative frequencies in the fluctuation spectrum $\mathcal{E}(\mathbf{x}, \mathbf{x}'; \omega)$. It gives nonzero results even in the vacuum state (at zero temperature), as we have seen in Eq. (53) and in the FD theorem [Eq. (58)].

2.4.3.3. Proof of the FD Theorem with Linear Response Theory As mentioned in Section 2.4.3.1, the macroscopic Maxwell equations in an absorbing medium have to be supplemented by material fluctuations to be consistent with thermodynamics and quantum theory. We thus split the polarization and magnetization operators into

$$\begin{aligned} \mathbf{P}(\mathbf{x}, t) &\mapsto \mathbf{P}_{\mathrm{fl}}(\mathbf{x}, t) + \mathbf{P}_{\mathrm{ext}}(\mathbf{x}, t) \\ \mathbf{M}(\mathbf{x}, t) &\mapsto \mathbf{M}_{\mathrm{fl}}(\mathbf{x}, t) + \mathbf{M}_{\mathrm{ext}}(\mathbf{x}, t) \end{aligned} \tag{68}$$

where the terms with the subscript fl describe the fluctuations in the material, and with the subscript ext fluctuations in all other sources, like the dipole moments of atoms or molecules. In thermal equilibrium, the fluctuations average to zero, as we found after Eq. (45). In the following, more explicit information about the polarization noise is not needed. We shall assume that a Hamilton operator $\hat{H}$ exists that generates the macroscopic Maxwell equations as the Heisenberg equations of motion for the electric and magnetic field operators. (We are actually adopting a quantum Langevin picture; see Mandel and Wolf [14] and Gardiner [37].) A similar demonstration has been given by Wylie and Sipe [38].

The field fluctuation spectrum, from Eq. (47), is given by the expectation value

$$\mathcal{E}(\mathbf{x}, \mathbf{x}'; \omega) = \int_{-\infty}^{+\infty} \mathrm{d}\tau \, \mathrm{e}^{\mathrm{i}\omega\tau} \langle \mathbf{E}(\mathbf{x}, 0)\mathbf{E}(\mathbf{x}', \tau) \rangle \tag{69}$$

We assume thermal equilibrium without external sources and have used the stationarity of the correlation function. We now connect this spectrum to the linear response of the field to an external dipole oscillator, following Callen and Welton [4].

The solution for the electric field operator in the presence of a polarization source can be represented in terms of the Green tensor in the quantum theory as well because the substitution [Eq. (68)] preserves the linearity of the macroscopic Maxwell equations. We thus get Eq. (30) with an operator-valued source term, plus a term describing the free evolution of the field

$$\mathbf{E}(\mathbf{x}, t) = \mathbf{E}_{\mathrm{free}}(\mathbf{x}, t) + \int \frac{\mathrm{d}\omega}{2\pi} \mathrm{e}^{-\mathrm{i}\omega t} \int_V \mathrm{d}^3 x' \, \mathcal{G}(\mathbf{x}, \mathbf{x}'; \omega) \cdot [\mathbf{P}_{\mathrm{fl}}(\mathbf{x}'; \omega) + \mathbf{P}_{\mathrm{ext}}(\mathbf{x}'; \omega)] \tag{70}$$

There is a similar contribution from the magnetization that we do not need for the present discussion. In equilibrium, the free field operator has zero average, and we get the expectation value

$$\langle \mathbf{E}(\mathbf{x}, t) \rangle = \int \frac{\mathrm{d}\omega}{2\pi} \mathrm{e}^{-\mathrm{i}\omega t} \int_V \mathrm{d}^3 x' \, \mathcal{G}(\mathbf{x}, \mathbf{x}'; \omega) \cdot \langle \mathbf{P}_{\mathrm{ext}}(\mathbf{x}'; \omega) \rangle \tag{71}$$

The Green tensor can thus be identified with the linear response of the average field to a classical external polarization source (where $\mathbf{P}_{\mathrm{ext}}(\mathbf{x}'; \omega)$ is c-number valued).

The linear response of the field can also be calculated directly from the Heisenberg equations. This provides us with an alternative expression for the Green tensor, where equilibrium correlations will become apparent. For simplicity, in the following text we focus on the response to an electric point dipole at the position $\mathbf{x}'$. The coupling of the field to the dipole is described by adding to the Hamiltonian the term

$$H_{\mathrm{int}}(t) = -\mathbf{d}(t) \cdot \mathbf{E}(\mathbf{x}', t) \tag{72}$$

and the Heisenberg equation reads

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{E}(\mathbf{x},t) = -\frac{\mathrm{i}}{\hbar}[\mathbf{E}(\mathbf{x},t), \widehat{H}] + \frac{\mathrm{i}}{\hbar}[\mathbf{E}(\mathbf{x},t), E_j(\mathbf{x}',t)]d_j(t) \tag{73}$$

where summation over j is understood in the last term. The first term generates the free evolution of the field.

Solving to first order in $\mathbf{d}$ and taking the average, we identify the field response function in the time domain as

$$\langle E_i(\mathbf{x},t)\rangle = \int_{-\infty}^{+\infty} \mathrm{d}\tau\, \chi_{ij}(\mathbf{x},\mathbf{x}',\tau)\, d_j(t-\tau) \tag{74}$$

$$\chi_{ij}(\mathbf{x},\mathbf{x}',\tau) = \begin{cases} \dfrac{\mathrm{i}}{\hbar}\langle[E_i(\mathbf{x},t), E_j(\mathbf{x}',t-\tau)]\rangle & \text{for } \tau \geq 0 \\ 0 & \text{for } \tau < 0 \end{cases} \tag{75}$$

where the time dependence of the field operators is that of the evolution under $\widehat{H}$. The response function is thus itself a correlation function of the field. Because of the stationarity of equilibrium, Eq. (75) does not depend on t. It can be checked directly that $\chi_{ij}(\mathbf{x},\mathbf{x}',\tau)$ is real as in the classical theory.

Taking the Fourier transform of χ_{ij}, we thus get an expression for the Green tensor in terms of a field correlation spectrum

$$\mathcal{G}_{ij}(\mathbf{x},\mathbf{x}';\omega) = \frac{\mathrm{i}}{\hbar}\int_0^{+\infty} \mathrm{d}\tau\, \mathrm{e}^{\mathrm{i}\omega\tau}\langle[E_i(\mathbf{x},\tau), E_j(\mathbf{x}',0)]\rangle \tag{76}$$

By causality, the time integral is running over one half of the real axis only. We can make an integral over all τ appear, as it occurs in the fluctuation spectrum [Eq. (69)], by forming the combination

$$\frac{1}{2\mathrm{i}}\{\mathcal{G}_{ij}(\mathbf{x}',\mathbf{x};\omega) - [\mathcal{G}_{ij}(\mathbf{x},\mathbf{x}';\omega)]^*\} = -\frac{1}{2\hbar}\int_{-\infty}^{+\infty} \mathrm{d}\tau\, \mathrm{e}^{\mathrm{i}\omega\tau}\langle[E_i(\mathbf{x},0), E_j(\mathbf{x}',\tau)]\rangle \tag{77}$$

The following relation allows us to permute operators occurring in equilibrium correlation functions

$$\int_{-\infty}^{+\infty} \mathrm{d}\tau\, \mathrm{e}^{\mathrm{i}\omega\tau}\langle A(\tau)B(0)\rangle = \mathrm{e}^{\beta\hbar\omega}\int_{-\infty}^{+\infty} \mathrm{d}\tau\, \mathrm{e}^{\mathrm{i}\omega\tau}\langle B(0)A(\tau)\rangle \tag{78}$$

In the classical theory, $\hbar = 0$ and operator ordering is irrelevant. Using this identity in the second term of the commutator in Eq. (77), we find the FD theorem

$$\mathcal{E}_{ij}(\mathbf{x}',\mathbf{x};\omega) = \frac{2\hbar}{\mathrm{e}^{\beta\hbar\omega}-1}\,\frac{\mathcal{G}_{ji}(\mathbf{x}',\mathbf{x};\omega) - [\mathcal{G}_{ij}(\mathbf{x},\mathbf{x}';\omega)]^*}{2\mathrm{i}} \tag{79}$$

The form from Eq. (58) is recovered when the Green tensor satisfies the symmetry relation of Eq. (31). This requires the additional assumption that permittivity and permeability are symmetric, an assumption that we shall make in this contribution.

Equation (78) can be proved using the Gibbs density operator from Eq. (43) and the solution for the Heisenberg operator $A(\tau)$ [see also Eq. (45)]

$$\mathrm{e}^{\mathrm{i}\omega\tau}\langle A(\tau)B(0)\rangle = \mathrm{e}^{\mathrm{i}\omega\tau}\,\frac{\mathrm{Tr}[\exp(\mathrm{i}\widehat{H}\tau/\hbar)A\exp(-\mathrm{i}\widehat{H}\tau/\hbar)B\exp(-\beta\widehat{H})]}{\mathrm{Tr}\exp(-\beta\widehat{H})} \tag{80}$$

One shifts the integration path in the complex τ-plane to the line $-\infty - i\hbar\beta \cdots + \infty - i\hbar\beta$ and assumes that for $|\tau| \to \infty$, the correlation function vanishes (otherwise this limiting value can be subtracted). Along the shifted path, Eq. (80) becomes

$$\begin{aligned} & e^{\beta\hbar\omega} e^{i\omega\tau} \frac{\mathrm{Tr}[\exp(\beta\widehat{H}) \exp(i\widehat{H}\tau/\hbar) A \exp(-i\widehat{H}\tau/\hbar) \exp(-\beta\widehat{H}) B \exp(-\beta\widehat{H})]}{\mathrm{Tr}\, \exp(-\beta\widehat{H})} \\ & = e^{\beta\hbar\omega} e^{i\omega\tau} \frac{\mathrm{Tr}[B \exp(i\widehat{H}\tau/\hbar) A \exp(-i\widehat{H}\tau/\hbar) \exp(-\beta\widehat{H})]}{\mathrm{Tr}\, \exp(-\beta\widehat{H})} \\ & = e^{\beta\hbar\omega} e^{i\omega\tau} \langle B(0) A(\tau) \rangle \end{aligned} \tag{81}$$

using cyclic permutation under the trace. The τ-integral now yields the right-hand side of Eq. (78).

2.4.4. Nonequilibrium Situations

A typical nonequilibrium situation that occurs in physics on the nanometer scale is a temperature gradient inside a nanostructure. In thermal scanning probe microscopy, to quote another example, structures are held at different temperatures, as they are in contact with different reservoirs. These kinds of settings can be described by a slight generalization of the present theory, provided one assumes that each volume element of the solid structure is locally in thermal equilibrium at temperature $T(\mathbf{x})$.

In this case, we can write down the fluctuation dissipation theorem for the thermal polarization field by generalizing Eq. (63)

$$\mathcal{P}(\mathbf{x}, \mathbf{x}'; \omega) = \frac{2\hbar\mathbb{I}\mathrm{m}\, \varepsilon_0 \varepsilon(\mathbf{x}; \omega)}{\exp[\hbar\omega/k_B T(\mathbf{x})] - 1} \delta(\mathbf{x} - \mathbf{x}') \tag{82}$$

We have assumed a local dielectric response for simplicity. The corresponding field fluctuation spectrum can be computed from the field operator [Eq. (70)], where the freely evolving field (with material damping, but without material fluctuations) and the Green tensor appear. Without external sources, one gets

$$\mathcal{E}_{ij}(\mathbf{x}, \mathbf{x}'; \omega) = \mathcal{E}_{ij}^{(\mathrm{free})}(\mathbf{x}, \mathbf{x}'; \omega) + \int_V \mathrm{d}^3 r [G_{ik}(\mathbf{x}, \mathbf{r}; \omega)]^* G_{jk}(\mathbf{x}', \mathbf{r}; \omega) \frac{2\hbar \mathrm{Im}\, \varepsilon_0 \varepsilon(\mathbf{r}; \omega)}{\exp[\hbar\omega/k_B T(\mathbf{r})] - 1} \tag{83}$$

See Henry and Kazarinov for a similar approach [39]. The first term is nonzero for a bounded material surrounded by a nonabsorbing dielectric (like free space) and describes the photons incident from infinity toward the observation points $\mathbf{x}$, $\mathbf{x}'$. It accounts for all of the field fluctuations when there is no material absorption at all. If the field in the surrounding medium as assumed to be at zero temperature (as in the visible frequency range), this term is zero for $\omega > 0$. Even at finite temperature, however, this term is typically negligible at subwavelength distances from an absorbing structure. Under these conditions, the second one dominates [i.e., the radiation resulting from the polarization noise from Eq. (82)]. We summarize explicit examples above planar substrates in the next section.

2.5. Example: Planar Surface

We review here results for the electric and magnetic field fluctuations at nanometer distances from a solid surface. We first outline the corresponding Green tensors and then discuss the electromagnetic noise spectra.

2.5.1. Electric and Magnetic Green Tensors

Consider a nonmagnetic solid with permittivity ε and permeability $\mu = 1$ that fills the half-space $z \leq 0$. We shall be interested in the fluctuation spectrum of the electromagnetic field in the vacuum half-space, in particular for subwavelength distances $0 < z < \lambda \equiv 2\pi c/\omega$. Up to hundreds of nanometers from the surface, this regime is relevant even at optical frequencies.

For the Green tensor $\mathcal{G}(\mathbf{x}, \mathbf{x}'; \omega)$, we can make the *ansatz*, provided both $\mathbf{x}$ and $\mathbf{x}'$ are located outside the solid

$$\mathcal{G}(\mathbf{x}, \mathbf{x}'; \omega) = \mathcal{G}^{(\mathrm{vac})}(\mathbf{x}, \mathbf{x}'; \omega) + \mathcal{G}^{(\mathrm{refl})}(\mathbf{x}, \mathbf{x}'; \omega) \tag{84}$$

where $\mathcal{G}^{(\mathrm{vac})}$ is the vacuum Green tensor [Eq. (38)]. $\mathcal{G}^{(\mathrm{refl})}$ describes the electric field reflected from the solid and is determined from the boundary conditions for the electric and magnetic fields at $z = 0$. The decomposition [Eq. (84)] is convenient to compute the field fluctuation spectra via the FD theorem [Eq. (58)] because it exhibits clearly the additional contribution resulting from the scattering from the surface [38, 40]. This statement remains true for scatterers of arbitrary shape, with more a complicated expression for the reflected or scattered field, of course.

Above a planar solid, the so-called Weyl expansion or angular spectrum representation provides a natural plane-wave basis for the incident and reflected fields (see the textbook by Nieto-Vesperinas [41]). The xy plane naturally plays a distinguished role here. Introducing two-dimensional in-plane wave vectors $\mathbf{Q} = (q_x, q_y)$, we shall use the notation

$$\mathbf{q}(\pm) = \mathbf{Q} \pm \mathbf{n} q_z \tag{85}$$

where $\mathbf{n}$ is the unit normal. One then has the following Fourier expansion for the reflected Green tensor [38]

$$z, z' > 0: \quad \mathcal{G}^{(\mathrm{refl})}(\mathbf{x}, \mathbf{x}'; \omega) = \frac{\mathrm{i}\omega^2}{2\varepsilon_0 c^2} \int \frac{\mathrm{d}^2 Q}{(2\pi)^2} \frac{\mathrm{e}^{\mathrm{i}\mathbf{q}(+)\mathbf{x} - \mathrm{i}\mathbf{q}(-)\mathbf{x}'}}{q_z} \sum_{\mu = \mathrm{s,p}} r_\mu \mathbf{e}_\mu(+) \mathbf{e}_\mu(-) \tag{86}$$

where the integral runs over all in-plane wave vectors $\mathbf{Q}$. The wave vectors for downward and upward waves (both propagating and evanescent) are given by $\mathbf{q}(-)$ and $\mathbf{q}(+)$, respectively.

The disc $|\mathbf{Q}| \leq \omega/c$ corresponds to propagating waves where

$$q_z = \sqrt{(\omega/c)^2 - Q^2} \qquad \mathrm{Im}\, q_z \geq 0 \tag{87}$$

is real, and $|\mathbf{Q}| > \omega/c$ describes evanescent waves that decay or increase exponentially with distance. Evanescent waves are required in the Green tensor to describe correctly the near field of a point dipole. For the reflected field, they provide the dominant contribution at subwavelength distances from the solid. All elementary plane waves satisfy the vacuum dispersion relation $[\mathbf{q}(\pm)]^2 = (\omega/c)^2$. Their polarization vectors are given by

$$\begin{aligned} \mathbf{e}_\mathrm{s}(\pm) &= \hat{\mathbf{Q}} \times \mathbf{n} \\ \mathbf{e}_\mathrm{p}(\pm) &= \frac{\mathbf{q}(\pm) \times \mathbf{e}_\mathrm{s}}{\omega/c} = \frac{\pm \hat{\mathbf{Q}} q_z - \mathbf{n} Q}{\omega/c} \end{aligned} \tag{88}$$

where $\hat{\mathbf{Q}}$ is the unit vector along $\mathbf{Q}$. We have normalized the polarization vectors in Eq. (88) such that $\mathbf{e}_\mu^2(\pm) = 1$; note that no complex conjugation is involved, although the vectors are complex in general. The conventional polarizations s and p are also called TE and TM in the literature (TE = electric field transverse to the plane of incidence spanned by $\mathbf{n}$ and $\hat{\mathbf{Q}}$).

The reflection from the solid mixes downward with upward waves and is characterized by the Fresnel reflection coefficients [1, 28]

$$r_\mathrm{s} = \frac{q_z - \sqrt{\varepsilon(\omega)(\omega/c)^2 - Q^2}}{q_z + \sqrt{\varepsilon(\omega)(\omega/c)^2 - Q^2}} \tag{89}$$

$$r_\mathrm{p} = \frac{\sqrt{\varepsilon(\omega)(\omega/c)^2 - Q^2} - \varepsilon(\omega) q_z}{\sqrt{\varepsilon(\omega)(\omega/c)^2 - Q^2} + \varepsilon(\omega) q_z} \tag{90}$$

Note that the permittivity of the solid only enters via the Fresnel coefficients. As long as the planar symmetry is not broken, Eq. (86) can also be used above a multilayer medium [42, 43]. Note that sign conventions differ for the Fresnel coefficients and polarization vectors; only the product of r_μ and the polarization vectors appearing under the sum in Eq. (86) has an unambiguous meaning.

We define the magnetic Green tensor by analogy to Eq. (30) as the magnetic induction field radiated by a point magnetic moment

$$B_i(\mathbf{x};\omega) = \mathscr{H}_{ij}(\mathbf{x},\mathbf{x}';\omega)\mu_j \tag{91}$$

From the Maxwell Eq. (13) we find that in terms of its electric counterpart, the magnetic Green tensor in the vacuum above the solid is given by the double curl

$$\mathscr{H}_{ij}(\mathbf{x},\mathbf{x}';\omega) = \frac{1}{\omega^2}\epsilon_{ikl}\epsilon_{jmn}\frac{\partial}{\partial x_k}\frac{\partial}{\partial x'_m}\mathscr{G}_{ln}(\mathbf{x},\mathbf{x}';\omega) \tag{92}$$

For the free space Green tensor, this leads to an expression similar to Eq. (38). For the reflected field, as given by the Green tensor in Eq. (86), we observe that the curl exchanges the polarization vectors from Eq. (88) according to

$$\mathbf{q}\times\mathbf{e}_\mathrm{s} = \frac{\omega}{c}\mathbf{e}_\mathrm{p} \qquad \mathbf{q}\times\mathbf{e}_\mathrm{p} = -\frac{\omega}{c}\mathbf{e}_\mathrm{s} \tag{93}$$

because $\mathbf{q}$, $\mathbf{e}_\mathrm{s}$, and $\mathbf{e}_\mathrm{p}$ form an orthogonal *Dreibein*. Hence, up to a factor of $1/c^2$, we obtain the magnetic field reflected from the solid by exchanging the reflection coefficients $r_\mathrm{s} \leftrightarrow r_\mathrm{p}$ in Eq. (86).

2.5.2. Short-Distance Expansions

To illustrate the behavior of the field at short distances, we review here asymptotic expansions for the electric and magnetic Green tensors.

2.5.2.1. Electric Field As a first step, we show that in the near field, the reflected part [Eq. (86)] of the electric Green tensor takes a simple, electrostatic form. The integral over the wave vector $\mathbf{Q}$ involves the factor $\mathrm{e}^{\mathrm{i}q_z(z+z')}$, which provides a natural cutoff for large $\mathbf{Q}$ as soon as q_z becomes imaginary with $|q_z| \geq 1/(z+z') \gg \omega/c$. Analyzing the integrand, we notice that it peaks around the cutoff value. We thus get the leading order asymptotics by using an expansion for Q, $|q_z|$ much larger than ω/c under the integral.

Let us first assume the more stringent condition $Q \gg |\sqrt{\varepsilon}|\omega/c$, which corresponds to a distance much shorter than the medium wavelength $z \ll \lambda/|\sqrt{\varepsilon}|$. The reflection coefficients and polarization vectors then behave like

$$r_\mathrm{s}\mathbf{e}_\mathrm{s}(+)\mathbf{e}_\mathrm{s}(-) \approx (\varepsilon-1)\frac{\omega^2}{4Q^2c^2}(\mathbf{n}\times\hat{\mathbf{Q}})(\mathbf{n}\times\hat{\mathbf{Q}}) \tag{94}$$

$$r_\mathrm{p}\mathbf{e}_\mathrm{p}(+)\mathbf{e}_\mathrm{p}(-) \approx \frac{\varepsilon-1}{\varepsilon+1}\frac{Q^2c^2}{\omega^2}(\mathbf{n}-\mathrm{i}\hat{\mathbf{Q}})(\mathbf{n}+\mathrm{i}\hat{\mathbf{Q}}) \tag{95}$$

where $\mathbf{n}$ is the surface normal. The polarizations behave very differently, with the p-polarized part dominating for large Q by a factor $(Qc/\omega)^4$. The reflection coefficient r_p tends toward the electrostatic value $r_\mathrm{stat} = (\varepsilon-1)/(\varepsilon+1)$ and becomes independent of Q. This means that the reflection from the surface is nondispersive and can be modeled in terms of image theory [1]: The reflected field corresponds to the well-known field of an image dipole $\tilde{\mathbf{d}} = (-d_x, -d_y, d_z)r_\mathrm{stat}$ located at the position $\mathbf{X}' - z'\mathbf{n}$ below the surface. Performing the integrals over $\mathbf{Q}$, we indeed find the short-distance asymptotics

$$\begin{aligned}&\mathscr{G}^{(\mathrm{refl})}(\mathbf{X}'+\mathbf{R}, z, \mathbf{X}', z';\omega)\\ &\quad\approx \frac{\varepsilon(\omega)-1}{\varepsilon(\omega)+1}\,\frac{(R^2+\bar{z}^2)\mathbb{1} - 3\mathbf{R}\mathbf{R} + (\bar{z}^2-2R^2)\mathbf{n}\mathbf{n} + 3\bar{z}\,(\mathbf{R}\mathbf{n}-\mathbf{n}\mathbf{R})}{4\pi\varepsilon_0(R^2+\bar{z}^2)^{3/2}}\end{aligned} \tag{96}$$

where $\bar{z} = z + z'$ and $(R^2 + \bar{z}^2)^{1/2}$ is the distance between the observation point and the image dipole. This expression depends on frequency only via the electrostatic reflection coefficient. Note also the broken isotropy of the correlation tensor where the coordinates that are parallel and perpendicular to the surface appear in nonequivalent ways. Nevertheless, the symmetry relations from Eqs. (31) and (32) are satisfied.

In a similar way, an asymptotic expression for distances larger than $\lambda/|\sqrt{\varepsilon}|$ can be worked out. This is particularly interesting above metallic surfaces, where $\varepsilon \approx i\sigma/\varepsilon_0\omega$ can be very large. In this case, the expansion [Eq. (96)] is valid for $z \ll \delta_\omega$, where the skin depth

$$\delta_\omega \equiv \sqrt{\frac{2\varepsilon_0 c^2}{\sigma\omega}} \tag{97}$$

can be much smaller than the wavelength (typically, at frequencies below the infrared). For the complementary range $\delta_\omega \ll z$, one finds, when repeating the analysis leading to the asymptotics in Eqs. (94) and (95) (see Ref. [44])

$$\operatorname{Im} \mathcal{G}(\mathbf{X}, z, \mathbf{X}, z; \omega) \approx \frac{\omega^2 \delta_\omega \mathbb{1}}{32\pi\varepsilon_0 c^2 z^2} \tag{98}$$

For simplicity we give only the imaginary part of the tensor with coinciding positions. Note the different power law with distance z and also the isotropic noise strength regarding the field polarization.

To summarize, the electric field fluctuation spectrum derived from the Green tensor from Eq. (96) using the FD theorem from Eq. (58) is

$$\mathcal{K}(\mathbf{x}, \mathbf{x}; \omega) = \frac{\hbar(\omega\delta_\omega/c)^2}{16\pi\varepsilon_0(e^{\beta\hbar\omega} - 1)} \begin{cases} \dfrac{\mathbb{1} + \mathbf{nn}}{z^3} & \text{for } z \ll \delta_\omega \\ \dfrac{\mathbb{1}}{\delta_\omega z^2} & \text{for } \delta_\omega \ll z \ll (\delta_\omega\lambda)^{1/2} \\ \dfrac{16\omega\mathbb{1}}{3c\delta_\omega^2} & \text{for } (\delta_\omega\lambda)^{1/2} \le z \end{cases} \tag{99}$$

It is only at distances $z \gtrsim (\delta_\omega\lambda)^{1/2}$ that the blackbody spectrum, originating from the free-space contribution $\mathcal{G}^{(\mathrm{free})}$, becomes the dominant contribution. Closer to the surface, the electric field fluctuations have a noise spectrum that can exceed the Planck formula by several orders of magnitude. Note also that the low-frequency limit is given by a constant spectrum $\propto k_B T/(\sigma z^3)$. The electric field fluctuations thus behave like white noise on the nanometer scale.

2.5.2.2. Magnetic Field Analogous calculations give for the Green tensor of the magnetic field at distances below the skin depth, $z, z' \ll \delta_\omega$

$$\begin{aligned} \mathcal{H}^{(\mathrm{refl})}(\mathbf{X}' + \mathbf{R}, z, \mathbf{X}', z'; \omega) \approx \frac{\omega^2\mu_0}{4\pi c^2} \Bigg\{ & \frac{\varepsilon(\omega) - 1}{4\bar{r}(\bar{r} + \bar{z})} [\bar{r}\mathbb{1} - \mathbf{RR}/(\bar{r} + \bar{z}) + \bar{z}\mathbf{nn} + (\mathbf{Rn} - \mathbf{nR})] \\ & - \frac{\varepsilon(\omega) - 1}{\varepsilon(\omega) + 1} \frac{\bar{z}\mathbb{1} + \mathbf{RR}/(\bar{r} + \bar{z}) - \bar{z}\mathbf{nn}}{\bar{r}(\bar{r} + \bar{z})} \Bigg\} \end{aligned} \tag{100}$$

where $\bar{r} = (R^2 + \bar{z}^2)^{1/2}$. From the FD theorem, we get the fluctuation spectrum of the magnetic field. We give here also the regimes of larger distances

$$\mathcal{H}(\mathbf{x}, \mathbf{x}; \omega) = \frac{\hbar\mu_0}{8\pi\delta_\omega^2(e^{\beta\hbar\omega} - 1)} \begin{cases} \dfrac{\mathbb{1} + \mathbf{nn}}{z} & \text{for } z \ll \delta_\omega \\ \dfrac{\delta_\omega^3(\mathbb{1} + \mathbf{nn})}{3z^4} & \text{for } \delta_\omega \ll z \ll (\delta_\omega\lambda^3)^{1/4} \\ \dfrac{8\delta_\omega^2\omega^3\mathbb{1}}{3c^3} & \text{for } (\delta_\omega\lambda^3)^{1/4} \le z \end{cases} \tag{101}$$

Note the different exponents for the power laws with distance compared to the electric field and the larger cross-over distance to the blackbody radiation spectrum. The low-frequency limit of the magnetic noise spectrum is $\propto \mu_0^2 k_B T\sigma/z$; it is frequency independent as well. Similar expressions have been derived in Refs. [25, 44, 45].

2.5.3. Discussion

2.5.3.1. Electromagnetic Energy Density Some of the results summarized above have been discussed by Joulain, Carminati, Mulet, and Greffet in a recent paper on the definition and measurement of the LDOS close to planar surfaces [31]. These authors analyze the spectrum of the electric and magnetic energy densities $u_e(z;\omega)$, $u_m(z;\omega)$ as a function of distance and point out that the definition from Eq. (35) of the LDOS should be taken with care given the nonequivalent role played by the electric and magnetic fields in the near field. To illustrate this, we plot in Fig. 1 the ratio $u_m(z;\omega)/u_e(z;\omega)$. Notice the strong dominance of the magnetic energy throughout the near field range up to $z \sim \lambda$ for a metallic surface. The asymptotic formulas (99), (101) provide good agreement with a numerical calculation based on the exact plane wave expansion (86) for the electric Green functions and its magnetic equivalent.

For a dielectric surface with an essentially real permittivity, the near field energy is dominantly electric, as shown in Fig. 1 (right). This behavior is not covered by the asymptotics from Eqs. (99) and (101) because the assumption $|\varepsilon| \gg 1$ breaks down but can be found from Eq. (96) and its magnetic counterpart. At distances comparable to the wavelength, oscillations appear for both metallic and dielectric surfaces that correspond to the standing waves formed by the partial reflection of the field. In the far field regime $z \gg \lambda$, the symmetry between electric and magnetic energy is restored, as expected from the Planck formula.

2.5.3.2. Macroscopic Field Quantization The thermodynamical approach to field fluctuations that we followed here is based on the fluctuation electrodynamics pioneered by Rytov and co-workers [35], also called source theory by Schwinger and coworkers (see, e.g., Ref. [46]). Applications to quantum optical problems can be found in the series of papers by Agarwal [40, 47–49] and by Wylie and Sipe [38, 50]. The quantization of the macroscopic Maxwell equations has been performed by Barnett and coworkers [51–53], who studied a quantum model for the medium excitations; see Ref. [54] for a review. Polarization and magnetization fluctuation operators have been introduced by Knöll, Welsch, and coworkers [55–58] and by Tip [59, 60]. A review is given in Ref. [2], and the equivalence of both approaches is proven in [61]. A formalism similar to that followed here, where quantum and thermal field fluctuations are treated on an equal footing, has been developed in the review of Henry and Kazarinov [39].

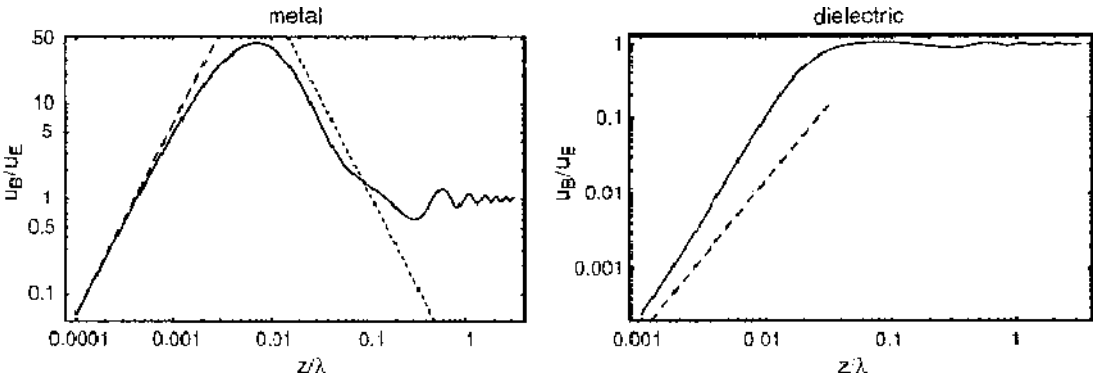


Figure 1. Ratio of magnetic to electric energy density versus distance from a medium-filled half-space, normalized to the wavelength λ. Left panel: metal with $\varepsilon = 1 + 8001i$ (skin depth $\delta = 0.008\lambda$). Right panel: dielectric ($\varepsilon = 2.3 + 0.1i$, $\delta > \lambda$). The dashed and dotted lines correspond to the asymptotic expansion from Eqs. (99) and (101).

2.5.3.3. Layered Media The Green tensor for planar multilayer structures has been extensively studied in the context of dielectric mirrors (or Bragg mirrors) [42]. Formulas similar to those reviewed here have been derived to study fluorescence from dye molecules or atoms, either placed in a layer on top of a planar structure or embedded inside (see, e.g., Refs. [15, 43, 62, 63]). (We give an introduction to fluorescence dynamics in Section 3.) As mentioned above, the expression from Eq. (86) describes the Green tensor above a multilayer structure as well, provided the correct reflection coefficients are used. Perturbative techniques, combined with a statistical description, have been developed to describe the reflection from layer structures with interface roughness [64–68]. In the field of surface physics, it is of great interest to calculate reflection coefficients with an emphasis on a microscopic description of the excitations in the solid (see, e.g., Refs. [69, 70] and the review papers by Feibelman [71] and Ford and Weber [26]). On short scales, this requires us to take into account the nonlocal character of the permittivity [see Eq. (20)]. The most versatile concept generalizing the reflection coefficients is that of a surface impedance: it applies even to superconductors, where the concept of a permittivity breaks down because the field does not penetrate into the solid.

The reflection coefficients show poles when the multilayer supports a guided mode. This can be a waveguide mode in a high-index layer [42] or a surface plasmon polariton on a planar metallic surface [72]. In the near-field asymptotics [Eq. (96)], for example, a pole occurs for $\varepsilon(\omega) = -1$, which corresponds to the large wave vector limit of the surface plasmon dispersion relation. The surface polariton mode typically gives a significant contribution to the electromagnetic energy spectrum in a narrow frequency range. This leads to a dramatic change in the frequency spectrum of thermal radiation as one approaches a metallic surface at finite temperature, as shown by Greffet and coworkers [73, 74]. Such a situation is most naturally modeled by the nonequilibrium approach outlined in Section 2.4.4: The surface is supposed to be held at some temperature T and radiates into the cold vacuum above. The fluorescent emission of molecules is also drained in narrow frequency bands by polariton and guided modes (see, e.g., Refs. [33, 63, 75] and the review by Barnes [76]).

2.5.3.4. Near-Field Coherence Finally, we point out that the electromagnetic near field shows unusual properties regarding polarization and its spatial coherence. We have already noted the breakdown of the isotropy of the polarization fluctuations in Eq. (99). The usual Jones vector formalism for partially polarized beams [14, 28] has been systematically generalized by the group of Friberg: While in a beam, only field components transverse to the wave vector occur; all three components have to be taken into account in the near field [77, 78]. The broken isotropy entails a nonzero degree of polarization, which can reach quite high values in the regime where thermally excited surface plasmon polaritons dominate the field energy (i.e., observation distance comparable to the wavelength).

The spatial coherence length l_{coh} of the field can be defined by analogy to the coherence time as the scale on which the correlation spectrum $\ell(\mathbf{x}, \mathbf{x}'; \omega)$ decays to zero as the distance $\mathbf{x} - \mathbf{x}'$ increases. Carminati and Greffet [73] have pointed out that in the near field, l_{coh} in general depends on the observation distance and the frequency. If l_{coh} is measured in a plane parallel to a radiating surface, it is determined at short distances $z \ll \lambda$ by the observation distance itself, as has been shown by Henkel et al. [79] using asymptotic expansions similar to Eqs. (96) and (100). The thermal near field is thus spatially much less coherent than the blackbody field where $l_{\mathrm{coh}} \sim \lambda$. A very different behavior is found at distances $z \sim \lambda$ from surfaces that support plasmon polariton modes. These modes can have very weak damping and create long-range coherence $l_{\mathrm{coh}} \gg \lambda$, that is determined by the attenuation length of the polariton mode (related to the imaginary part of the wave vector).

3. NEAR-FIELD VACUUM FLUCTUATIONS

In this section, we review the theory of spontaneous emission for atoms or molecules at distances of at least a few nanometers from solid structures. We restrict ourselves to electromagnetic interactions and neglect the tunnelling of the atomic or molecular electrons into the solid. We also put for simplicity the quantum efficiency of the transition to unity,

assuming that the electronic excitation can only be lost via radiative decay (conversion into photons that propagate into the far field) or nonradiative decay (field absorbed in the solid). In this section, we primarily deal with electromagnetic fields in the visible. Excluding very high temperatures ($T \ll 10^4$ K), we can safely assume that the field is in the vacuum state.

3.1. Molecular Fluorescence Dynamics

Consider the three-level system shown in Fig. 2. It captures many of the features of real molecules, where an optical field excites first the level $|a\rangle$. This level then decays rapidly, on the picosecond scale or faster, toward the level $|e\rangle$ from which only radiative decay is possible in free space, with a characteristic timescale of nanoseconds. In the following text, we summarize the equations of motion for the two levels $|e\rangle$, $|g\rangle$.

3.1.1. Optical Bloch Equations

Because of the coupling to the electromagnetic vacuum field, the three-level system is actually open, and its dynamics do not reduce to a simple Schrödinger equation. When the vacuum field fluctuations are averaged over, one gets an effective description in terms of correlation functions of the system's probability amplitudes $c_{e,g}(t)$. These correlations are the level populations

$$p_{e,g} = \langle |c_{e,g}(t)|^2 \rangle \tag{102}$$

and the coherence or average dipole moment

$$\mathbf{d}(t) = \langle \mathbf{d}\, c_g^*(t) c_e(t) \rangle + \text{CC} = \langle \mathbf{d}\, \sigma(t) + \mathbf{d}^*\, \sigma^\dagger(t) \rangle \tag{103}$$

where $\mathbf{d} = \langle g|\hat{\mathbf{d}}|e\rangle$ is the matrix element of the electric dipole operator between the two levels and CC denotes the complex conjugate term. In the second equality, we have introduced the annihilation and creation operators σ and $\sigma^\dagger$ for the two-level system. In the Schrödinger picture, $\sigma = |g\rangle\langle e|$ is the lowering operator and $\sigma^\dagger$ the raising operator. Note that $\sigma^2 = (\sigma^\dagger)^2 = 0$, which is characteristic for two-level systems.

For the level populations, one has the pair of rate equations [13, 14]

$$\dot{p}_e = -\gamma p_e + R p_g + \frac{\mathrm{i}}{\hbar}\left[\mathbf{d}^* \cdot \mathbf{E}_L^{(+)}(t)\langle \sigma^\dagger \rangle - \mathbf{d} \cdot \mathbf{E}_L^{(-)}(t)\langle \sigma \rangle\right] \tag{104}$$

$$\dot{p}_g = +\gamma p_e - R p_g - \frac{\mathrm{i}}{\hbar}\left[\mathbf{d}^* \cdot \mathbf{E}_L^{(+)}(t)\langle \sigma^\dagger \rangle - \mathbf{d} \cdot \mathbf{E}_L^{(-)}(t)\langle \sigma \rangle\right] \tag{105}$$

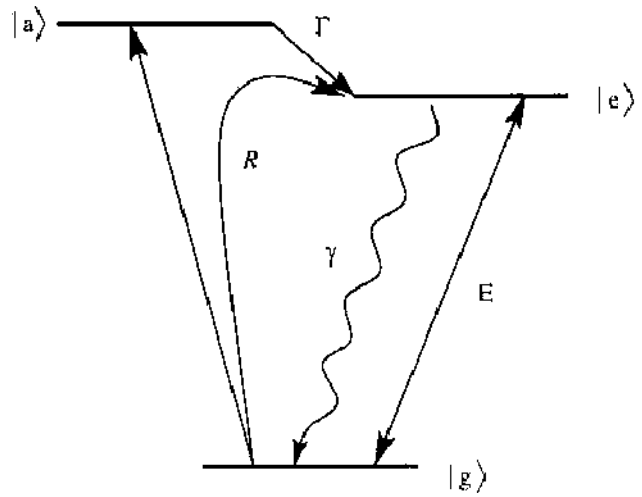


Figure 2. Three-level model for excitation of a molecule. Spontaneous emission is analyzed for the $|e\rangle \rightarrow |g\rangle$ transition, which is also probed by a near-resonant laser field **E**. The molecule is pumped to the excited state $|e\rangle$ via the short-lived state $|a\rangle$. γ, spontaneous decay rate for transition $|e\rangle \rightarrow |g\rangle$; R, effective excitation rate $|g\rangle \rightarrow |a\rangle \rightarrow |e\rangle$.

where the excitation rate R and the spontaneous decay rate γ describe the transitions shown in Fig. 2. We give their expressions in terms of the field in Eqs. (107) and (108), below. The average dipole operator evolves according to the optical Bloch equation

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle\sigma\rangle = -\mathrm{i}(\omega_{\mathrm{eg}} - \mathrm{i}\gamma/2)\langle\sigma\rangle + \frac{\mathrm{i}}{\hbar}\mathbf{d}^* \cdot \mathbf{E}_{\mathrm{L}}^{(+)}(t)(p_{\mathrm{g}} - p_{\mathrm{e}}) \tag{106}$$

where $\hbar\omega_{\mathrm{eg}} = E_{\mathrm{e}} - E_{\mathrm{g}}$ is the Bohr frequency of the transition. We have allowed in Eqs. (104)–(106), for later use, the interaction with a coherent laser field $\mathbf{E}_{\mathrm{L}}(t) = \mathbf{E}_{\mathrm{L}}\,\mathrm{e}^{-\mathrm{i}\omega t} + \mathrm{CC} \equiv \mathbf{E}_{\mathrm{L}}^{(+)}(t) + \mathbf{E}_{\mathrm{L}}^{(-)}(t)$ with the slowly varying, complex amplitude $\mathbf{E}_{\mathrm{L}} = \mathbf{E}_{\mathrm{L}}(\mathbf{r})$ at the molecule position. Note that the Bloch equations that describe the dissipative dynamics of the two-level system are analogous to the macroscopic Maxwell equations when the polarization fluctuations of the material are averaged over.

For the excitation and spontaneous emission rates, a simple calculation starting from Fermi's Golden Rule leads to

$$R = \frac{|\mathbf{d}_{\mathrm{ga}} \cdot \mathbf{E}_{\mathrm{ex}}(\mathbf{r};\, \omega_{\mathrm{ag}})|^2}{\hbar^2\Gamma} \tag{107}$$

where $\mathbf{d}_{\mathrm{ga}}$ is the dipole moment for the $|\mathrm{g}\rangle \leftrightarrow |\mathrm{a}\rangle$ transition, $\mathbf{E}_{\mathrm{ex}}(\mathbf{r};\, \omega_{\mathrm{ag}})$ is the corresponding complex electric field amplitude, and Γ is the relaxation rate between the excited states. The excitation rate R is thus proportional to the local excitation intensity. The spontaneous decay rate γ is determined by the field correlation spectrum at the negative Bohr frequency:

$$\gamma = \frac{1}{\hbar^2}\mathbf{d}^* \cdot \mathcal{C}(\mathbf{r}, \mathbf{r};\, -\omega_{\mathrm{eg}}) \cdot \mathbf{d} = \frac{2}{\hbar}\mathbf{d}^* \cdot \mathrm{Im}\,\mathcal{G}(\mathbf{r}, \mathbf{r};\, \omega_{\mathrm{eg}}) \cdot \mathbf{d} \tag{108}$$

With our sign conventions, it is thus nonzero at zero temperature as expected for spontaneous decay. The fluctuation dissipation theorem (58) provides the link to the imaginary part of the Green tensor, viz., the LDOS of Eq. (35). In free space, we get from Eqs. (39) and (108) the well-known Einstein A coefficient $\gamma = |\mathbf{d}|^2\omega_{\mathrm{eg}}^3/(3\pi\varepsilon_0\hbar c^3)$.

Both rates from Eqs. (107) and (108) show that the atom or molecule is sensitive to the field at its position $\mathbf{r}$ and can thus measure the excitation intensity or vacuum fluctuation spectrum with high spatial resolution. This observation provides one of the motivations for combining near-field scanning optical microscopy with single-molecule spectroscopy [7]. The molecule realizes a detector with dimensions that can be much smaller than typical probe tips. In addition, it is a passive probe because the field dynamic is not significantly perturbed by the presence of a single dipole. We can thus use the field equilibrium fluctuation spectrum to compute the decay rate and get the excitation rate from the local field intensity in the absence of the molecule. (The scanning tip holding the molecule may provide some perturbation, of course.)

The spontaneous decay rate γ can be measured quite directly with pulsed excitation. After the rapid relaxation from the state $|\mathrm{a}\rangle$, this prepares the molecule in the excited state $|\mathrm{e}\rangle$. The temporal evolution of the subsequently emitted fluorescence light provides γ. If single-photon pulses are detected, their time delay τ_{d} after the excitation pulse is distributed according to $\gamma\exp(-\gamma\tau_{\mathrm{d}})$. This assumes that the photon escape time from the sample is negligibly small, or at least constant. For a discussion of the emitted fluorescence power in a stationary situation, see Section 3.1.3.

3.1.2. Properties of the Bloch Equations

In the rate Eqs. (104) and (105), we observe that the laser field drives the populations on the transition $|\mathrm{g}\rangle \leftrightarrow |\mathrm{e}\rangle$ with a rate given by the out-of-phase quadrature $\mathrm{Im}[\langle\sigma^+(t)\rangle\mathbf{d}^* \cdot \mathbf{E}_{\mathrm{L}}^{(+)}(t)]$ of the dipole moment with respect to the laser. This is as expected, because only this quadrature is associated with mechanical energy dissipation, as we already saw in the energy conservation law in Eq. (23).

From Eq. (106), the average dipole moment induced by the laser field depends on the population inversion $p_e - p_g$. In the stationary state, its positive frequency part can be written in terms of the polarizability tensor

$$\mathbf{d}\langle\sigma\rangle_{st} = \langle\alpha(\omega)\rangle \cdot \mathbf{E}_L e^{-i\omega t}$$

$$\langle\alpha(\omega)\rangle = \frac{\mathbf{d}\mathbf{d}^* \langle p_g - p_e\rangle_{st}}{\hbar(\omega_{eg} - \omega - i\gamma/2)} \tag{109}$$

The small-signal absorption turns into gain ($\mathrm{Im}\,\alpha < 0$) when the level populations are inverted, $p_e > p_g$. In this case, stimulated emission dominates over absorption. When the laser field on the $|g\rangle \leftrightarrow |e\rangle$ transition is weak, the stationary populations are determined by the "incoherent" excitation rate R and the decay rate γ [Eqs. (107) and (108)]

$$\langle p_g\rangle_{st} = \frac{\gamma}{R+\gamma} \qquad \langle p_e\rangle_{st} = \frac{R}{R+\gamma} \tag{110}$$

For a large laser power, the stationary populations also depend on $\mathbf{E}_L$ (saturation). These populations are sometimes modeled by an effective temperature defined by $e^{-\beta_{eff}\hbar\omega_{eg}} = \langle p_e\rangle_{st}/\langle p_g\rangle_{st}$, which becomes negative in the case of inversion [39].

The rate Eqs. (104) and (105) imply the constant of motion $p_g + p_e \equiv 1$, as they should, because otherwise the total probability would not be conserved. Note that this is similar to energy conservation in the macroscopic Maxwell equations when the thermal radiation by the medium fluctuations is taken into account.

We recall that the Bloch equations (104)–(106) are valid only on time scales much longer than the optical period [14]. The dynamics indeed happens on the timescale $1/\gamma \gg 1/\omega_{eg}$ for spontaneous decay. The exponential decay implicit in Eq. (104) is a good approximation when the vacuum fluctuation spectrum does not show any strong frequency dependence on the scale γ around the transition frequency ω_{eg}. Otherwise, the vacuum field has to represented as a reservoir with a finite memory time (non-Markovian bath). The Bloch equations are then replaced by integro-differential equations (see Van Kampen's textbook [80] and the review by Lambropoulos, Nikolopoulos, Nielsen, and Bay [81]).

Finally, we have assumed here that the decay rates of the population p_e and of the dipole $\langle\sigma\rangle$ (also denoted by $1/T_1$ and $1/T_2$) only differ by a factor 1/2. This it true as long as the coupling to the radiation field provides the only decay mechanism, which is the case for isolated atoms in vacuum [13]. In particular, for molecules embedded in a solid matrix, the dipole $\langle\sigma\rangle$ relaxes faster because vibrations of the surroundings randomly shift the transition frequency. This leads to a much smaller dephasing time T_2: Superpositions of the states $|g\rangle$ and $|e\rangle$ are then more fragile than these states themselves.

3.1.3. Emitted Fluorescence Spectrum

We consider now the light emitted by the molecule in a steady-state situation. As mentioned before, we assume that the excited state $|e\rangle$ is populated incoherently via the third level $|a\rangle$. Excluding the case of "resonance fluorescence" [14], we put the laser field $\mathbf{E}_L(t)$ near-resonant with the $|g\rangle \leftrightarrow |e\rangle$ transition to zero.

For this incoherent excitation scheme, Eq. (106) implies that the average dipole moment vanishes in the stationary state. This does not mean, however, that the molecule does not emit at all, because the dipole has a nonzero fluctuation spectrum

$$\mathcal{S}(\omega) = \int_{-\infty}^{+\infty} d\tau\, e^{i\omega\tau} \langle \mathbf{d}(t)\mathbf{d}(t+\tau)\rangle_{st} \tag{111}$$

To get the power fed by the molecule into radiation in the emission band around ω_{eg}, we start from the energy conservation law in Eq. (23). The local electric field operator is found from the macroscopic Maxwell equations, using the Heisenberg operator from Eq. (103) for the molecular dipole as source term. One takes the average of the lost power $2\omega\,\mathrm{Im}[\mathbf{P}\cdot\mathbf{E}]$ in the stationary state of the two-level system with the populations [Eq. (110)]. The spectrum

of the total dissipated power, containing both emitted radiation and nonradiative loss, is then given [see Eq. (65)] by

$$P_{\rm tot}(\mathbf{r};\omega) = 2\omega^2 \mathscr{I}_{jj}(\omega)\,\mathrm{Im}\,\mathscr{G}_{jj}(\mathbf{r},\mathbf{r};\omega) \tag{112}$$

The radiation emitted into the far field has a spectrum given by the Poynting vector. Using the far field behavior of the electric and magnetic fields, the emission can be represented by the Green tensor as well

$$P_{\rm em}(\mathbf{r};\omega) = \frac{2}{c\mu_0}\int \mathrm{d}^2a(\mathbf{x})\,[\mathscr{G}_{ij}(\mathbf{x},\mathbf{r};\omega)]^*\,\mathscr{G}_{ik}(\mathbf{x},\mathbf{r};\omega)\mathscr{I}_{jk}(\omega) \tag{113}$$

where the integration variable $\mathbf{x}$ runs over a sphere in the far field with surface element $\mathrm{d}^2a(\mathbf{x})$. (We assume the system to be completely surrounded by vacuum.) Finally, the spectrum of the absorbed power is

$$P_{\rm abs}(\mathbf{r};\omega) = 2\omega\int \mathrm{d}^3V\{\varepsilon_0\,\mathrm{Im}\,\varepsilon(\mathbf{x};\omega)[\mathscr{G}_{ij}(\mathbf{x},\mathbf{r};\omega)]^*\,\mathscr{G}_{jk}(\mathbf{x},\mathbf{r};\omega)\mathscr{I}_{jk}(\omega) + (\mathrm{Im}\,\mu\text{ term})\} \tag{114}$$

and involves the overlap between the dipole emission spectrum and the medium absorption bands. The magnetic absorption is proportional to the square of the magnetic field, see Eq. (23), which can be written as the curl of the electric Green tensor.

We compute the dipole fluctuation spectrum from Eq. (111) from the fluctuation dissipation theorem [cf. Eq. (79)]

$$\mathscr{I}_{ij}(\omega) = \frac{2\hbar}{\mathrm{e}^{\beta_{\rm eff}\hbar\omega}-1}\frac{\langle\alpha_{ji}(\omega)\rangle_{\rm st} - \langle[\alpha_{ij}(\omega)]^*\rangle_{\rm st}}{2\mathrm{i}} \tag{115}$$

where the effective temperature $1/\beta_{\rm eff}$ is defined after Eq. (110). The polarizability in the stationary state, Eq. (109), shows that the spectrum is a narrow Lorentzian centered at $\omega_{\rm eg}$ with a width $\gamma \ll \omega_{\rm eg}$. We can thus replace the Boltzmann factor $\mathrm{e}^{-\beta_{\rm eff}\hbar\omega}$ by its value R/γ at resonance and get

$$\mathscr{I}(\omega) = \langle p_{\rm e}\rangle_{\rm st}\frac{\mathbf{d}^*\mathbf{d}\,\gamma}{(\omega-\omega_{\rm eg})^2+\gamma^2/4} \tag{116}$$

This expression shows that the spontaneous decay rate γ determines the width of the fluorescence spectrum, provided the Green function in Eq. (113) does not present a strong frequency dependence on the scale γ around $\omega_{\rm eg}$. Recall that we had to assume this when using the Bloch equations in the present form for the dynamics of the two-level system.

The total power dissipated by the molecule can be computed from the frequency integral of Eq. (112). Using the same approximation of a narrow-band emission,

$$P_{\rm tot}(\mathbf{r}) = \langle p_{\rm e}\rangle_{\rm st} 2\omega_{\rm eg} d_i^* d_j\,\mathrm{Im}\,\mathscr{G}_{ij}(\mathbf{r},\mathbf{r};\omega_{\rm eg}) \tag{117}$$

Using the fluctuation dissipation theorem for the field Eq. (58) and Eq. (108) for the spontaneous decay rate γ, this can also be written in the intuitive form

$$P_{\rm tot}(\mathbf{r}) = \langle p_{\rm e}\rangle_{\rm st}\hbar\omega_{\rm eg}\gamma \tag{118}$$

showing that the dissipated power, in units of the photon energy $\hbar\omega_{\rm eg}$, is given by the steady-state probability to find the molecule in the excited state times the spontaneous decay rate γ. The comparison between Eqs. (117) and (118) shows that the spontaneous decay rate can be computed classically: The imaginary part of the Green tensor gives the total power dissipated by a classical dipole with frequency $\omega_{\rm eg}$ and amplitude $\mathbf{d}$, and γ equals this power when normalized to the photon energy $\hbar\omega_{\rm eg}$ [15, 33].

To the same approximation, the power emitted as far-field radiation can be written as

$$P_{\rm em}(\mathbf{r}) = \frac{2R(\mathbf{r})}{R(\mathbf{r})+\gamma(\mathbf{r})}\int \mathrm{d}^2a(\mathbf{x})|\mathscr{G}(\mathbf{x},\mathbf{r};\omega_{\rm eg})\cdot\mathbf{d}|^2 \tag{119}$$

and the radiative decay rate is given by

$$\frac{\gamma_{\mathrm{rad}}(\mathbf{r})}{\gamma(\mathbf{r})} = \frac{P_{\mathrm{em}}(\mathbf{r})}{P_{\mathrm{tot}}(\mathbf{r})} \tag{120}$$

In Eq. (119), we have made explicit that the excited state population generally depends on the molecule position via both the excitation rate and the decay rate. The dependence of the emitted radiation power on $\mathbf{r}$ is thus quite intricate because it involves both the spatial profile of the excitation field and, via the Green tensor, the emission efficiency as a function of the dipole source position. All these dependencies, however, can be calculated with the classical macroscopic Maxwell equations. To the order of perturbation theory we consider here, the exciting field and the Green tensor are those of the dielectric surroundings and do not involve the backaction of the molecule. We note that the physical quantity with the simplest position dependence is the width $\gamma(\mathbf{r})$ of the fluorescence spectrum given by Eq. (108). It provides a passive probe of the LDOS, as defined in Eq. (35).

A similar simplification occurs for the emitted power [Eq. (119)] when the excited state shows a fast nonradiative decay (low quantum efficiency). This corresponds to a large, position-independent rate $\gamma \gg R(\mathbf{r})$ so that $P_{\mathrm{em}}(\mathbf{r}) \propto R(\mathbf{r})$, which is proportional to the excitation intensity (see, e.g., Refs. [82]). However, the integral involving the Green tensor will still contribute an additional dependence on $\mathbf{r}$.

To summarize, we have outlined the quantum theory of molecule fluorescence in arbitrary dielectric surroundings. The total and radiative decay rates have been connected to a classical electrodynamics problem, whose solution is encoded in the Green tensor for the given material geometry. The molecular decay gives access to the LDOS or, more precisely, to the spectrum of electric vacuum fluctuations at the position of the molecule and its transition frequency. For weak excitation strength, the decay rate is independent of the local excitation intensity, so that the position dependence is only determined by the LDOS. It can be measured either with pulsed excitation or from the width of the emission spectrum. The power emitted into the far field yields the radiative contribution to the decay rate. It also depends on the local intensity of the excitation field, this linear dependence being for weak excitation.

3.2. Spontaneous Decay Close to Nanostructures

We review in this section calculations showing that the spontaneous decay rate of a molecule can be used as a local probe for a nanostructured substrate. It is clear from the above discussion that $\gamma(\mathbf{r})$ is determined by the fluctuation spectrum of the electromagnetic vacuum field. The link to the properties of the substrate thus remains to be investigated. It is of particular interest to analyze the spatial resolution with which the properties of a nanostructure are mapped to the near field above.

3.2.1. Weakly Corrugated Structures

Let us focus first on dielectric structures with vanishing or small absorption and $\mathrm{Re}\,\varepsilon(\omega_{eg}) \geq 1$. The spontaneous decay rate in the near field of objects with simple shapes has been able to be calculated since the 1970s (see, e.g., Ref. [15]). More complicated shapes require numerical methods that became available with increased computer power in the 1990s [83, 84]. Insight into the resolution limit, however, is provided by analytical treatments as well.

Henkel and Sandoghdar [85] have reported a perturbative calculation of the strength of the vacuum fluctuations as a function of the molecule position and the orientation of its transition dipole moment. The price to pay for the analytical approach is the requirement of an essentially planar substrate, whose corrugation amplitude is small compared to both the wavelength and the molecular distance. The field scattered from the nanoscale corrugation is then calculated perturbatively to first order in the corrugation, using the so-called Rayleigh expansion [86]. Similar approaches have been reported for scattering problems from rough surfaces [65, 87–89]. One finds a position-dependent correction to the decay rate whose spatial Fourier transform is proportional to the corrugation spectrum of the substrate. This

correction answers the question of resolution, as the wavenumber-dependent ratio between the decay rate and the corrugation provides, in the language of optical imaging, the transfer function of the imaging process. Its knowledge allows us to deconvolve the spatial modulations of the fluorescence rate and to retrieve the surface profile. We note that a spatially varying refractive index is indistinguishable, in this approach, from a corrugated topography [90]. Topographic information, however, is routinely accessible in scanning near-field devices from the feedback loop that monitors the probe-sample distance [7]. Henkel and Sandoghdar have found simple asymptotic formulas for the limit of small-scale corrugations that allow them to show analytically that the resolution is only limited by the molecule–substrate distance. Introducing the wavenumber $k_0 = \omega_{eg}/c$ of the fluorescence light and the spectrum $s(\mathbf{Q})$ of the surface corrugation, the modulation of the decay rate can be written as [85]

$$\frac{\delta\gamma(\mathbf{R}, z)}{\gamma_{vac}} = \int \frac{d^2Q}{(2\pi)^2} e^{i\mathbf{Q}\cdot\mathbf{R}} k_0 s(\mathbf{Q}) \sum_i |d_i|^2 F_i(\mathbf{Q}; z) \tag{121}$$

$$F_i(\mathbf{Q}; z) \approx \frac{3(Q/k_0)^{\alpha_i} e^{-Qz}}{2} \mathrm{Im}\left[\frac{\varepsilon - 1}{\varepsilon + 1} f_i(k_0 z; \varepsilon)\right] \quad \text{for } Q \gg \frac{\omega_{eg}}{c} \tag{122}$$

where the substrate permittivity ε is to be taken at the transition frequency ω_{eg}, and the exponents are $\alpha_i = 1$ for a dipole component d_i parallel to $\mathbf{e}_z$ or $\widehat{\mathbf{Q}}$ and $\alpha_i = 0$ for $\mathbf{d} \| (\mathbf{e}_z \times \widehat{\mathbf{Q}})$. The dimensionless functions $f_i(k_0 z; \varepsilon)$ are given in Eqs. (30–32) of Ref. [85]. The main feature of Eq. (122) is the exponential cutoff for large wave vectors $Q \gg 1/z$. According to the uncertainty relation (40), this width in reciprocal space implies that one can retrieve features of the substrate with a lateral resolution on the order of the scanning height z by measuring the spatial modulation of the fluorescence rate.

Van Labeke's group proposed a similar approach [91, 92] and computed two-dimensional fluorescence images that demonstrate numerically a subwavelength resolution. de Fornel's group has developed numerical calculations including scattering to all orders to demonstrate the validity of this way of imaging a nanostructure [93, 94]. The concept of a transfer function has been discussed by Greffet and coworkers for the images obtained by a SNOM (Refs. [95, 96], see [34] for a review). The physics is very similar in both cases, with the difference coming only from the illuminating field used in a SNOM. A direct comparison of Eqs. (121) and (122) to experiments in which the fluorescence lifetime is measured as a function of molecule position is not straightforward because of the influence of the nearby probe tip (see Ref. [84] and the review by Dunn [7]). Promising results with single molecules as pointlike detectors or sources have been achieved by Sandoghdar's group [97, 98].

3.2.2. Fluorescence Near Nano-objects

The fluorescence decay rate at nanometer distance from regularly shaped objects like ellipsoids has been recently studied analytically in the electrostatic approximation [99, 100]. A review is given by Klimov and co-workers in Ref. [101]. On this small scale, retardation is negligible, and the field scattered by the object acquires an imaginary part (leading to nonradiative damping) only because of the complex object permittivity. Elliptical metallic objects, for example, allow us to study the coupling to plasmon resonances whose dispersion relation depends on the object curvature. For a sphere, the resonance occurs at $\varepsilon(\omega) = -2$, and this condition shifts to $\varepsilon(\omega) = -1$ for a cylinder or a plane [72, 101].

The characterization of fluorescent decay near objects of arbitrary shape requires the numerical solution of the Maxwell equations that have been developed in recent years following different techniques. Popular approaches use a finite element discretization of the wave equations, a reformulation as a volume integral equation [83, 102], or a direct integration in the time domain (FDTD, finite-difference time-domain [103]). We focus this discussion on a semianalytical method, the multiple multipole expansion (MMP) that has been developed by Hafner [104] and applied to near-field optics by Hecht, Novotny, Pohl, and coworkers [84, 105]. The main idea is to expand the electromagnetic field in multipoles centered at different positions in the dielectric structure. The multipole coefficients are determined by an optimal matching of the fields at the boundaries between otherwise homogeneous dielectric

objects. The method also covers metallic objects and its convergence is sped up when the multipoles are centered at high-symmetry positions in the structure. The dipole moment of an excited molecule is naturally included by fixing the corresponding coefficient in the multipole expansion. A planar substrate can be taken into account by adding the appropriate reflected part to the multipolar basis fields. Once the expansion coefficients are found numerically from a least squares algorithm, the field can be computed everywhere. More details about the MMP approach are given by Hafner [104].

An example of fluorescence decay rates computed with this method is illustrated by Fig. 3, reproduced from a paper by Novotny [84]. A single molecule is placed on a planar substrate, and a cylindrical object is scanned above it. This models a typical situation in SNOM. The radiative lifetime of the molecule shows the largest reduction when the object is metallic and the distance is in the range of a few nanometers. At slightly larger scanning distances, nonradiative decay becomes less effective and the lifetime can even be larger than in free space. A strong dependence on the orientation of the dipole is observed as well, with a behavior qualitatively similar to that observed above a planar metallic surface [15]. A dielectric object gives less dramatic changes of the lifetime, but a strong polarization dependence as

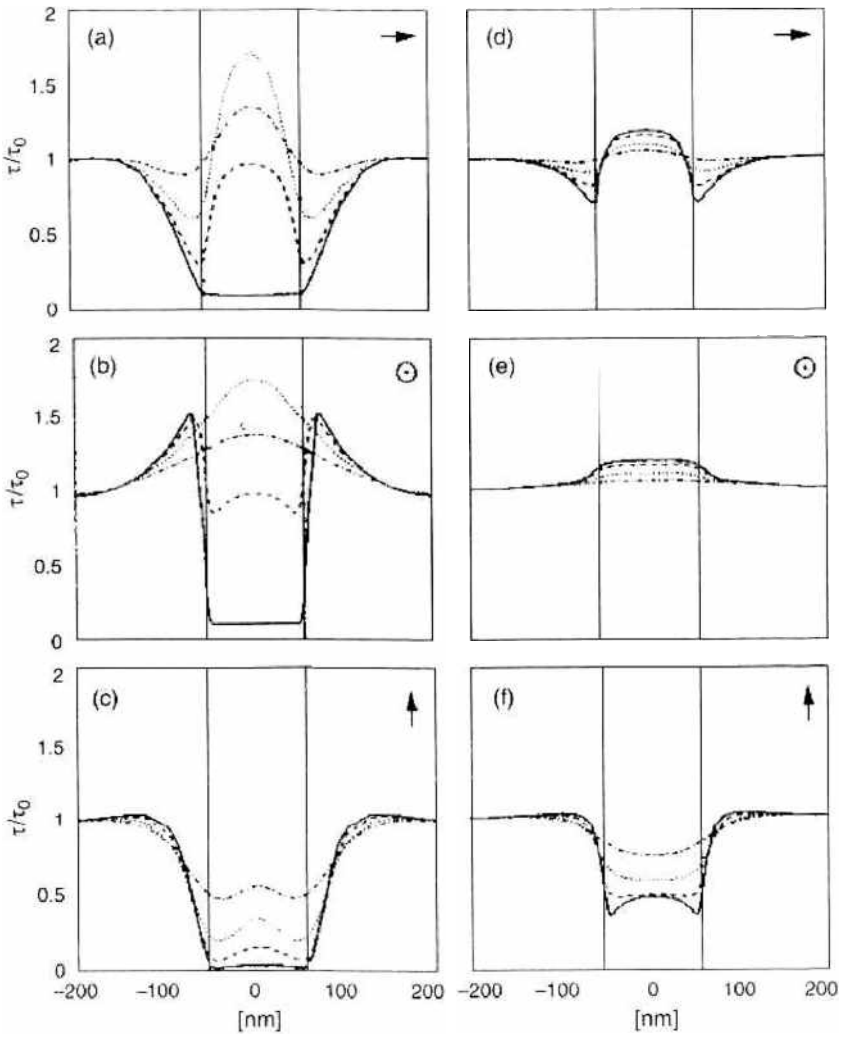


Figure 3. Lifetime $1/\gamma(x)$ of a molecule on a dielectric surface as a nanometer-sized object is scanned above it. The lifetime is normalized to its vacuum value. A quantum yield of unity is assumed (no intrinsic decay channel). Emission wavelength $\lambda = 488$ nm, dipole orientation as indicated by the symbols. The object is cylindrical and 110 nm in diameter and with a 40-nm height. The vertical lines indicate the lateral size of the object. (a–c): aluminum object ($\varepsilon = -34.5 + 18.5i$); (d–f): dielectric object ($\varepsilon = 3$). The different curves correspond to different scanning heights: 5 nm (solid curves), 10 nm (dashed curves), 20 nm (dotted curves), and 40 nm (dash-dotted curves). Reprinted with permission from [84], L. Novotny, Appl. Phys. Lett. **69** 3806 (1996). © 1996 American Institute of Physics.

well. From the viewpoint of SNOM, the sharp features around the borders of the object are interesting because they may lead to enhancements in image contrast and spatial resolution.

The angular distribution of the fluorescent emission is changed as well when an object approaches the molecule at a subwavelength distance, as predicted by Novotny [84] and as observed experimentally by Van Hulst's group [106, 107]. One observes the somewhat counterintuitive behavior that emission directions pointing toward a metallic object tend to be favored.

As a final example, we quote the work by Barnes, Sambles, and coworkers on fluorescence close to a patterned metallic substrate, where a two-dimensional band structure for the surface plasmon resonances is created. At distances comparable to the wavelength, an excited molecule predominantly emits into these surface resonances. If a surface plasmon band gap coincides with the molecular emission frequency, the emission is inhibited and one observes an increased lifetime [108, 109].

3.2.3. Molecules Embedded in Nanostructures

We have restricted the discussion so far to a molecule located in vacuum and outside the solid structure. This simplifies the analysis because the finite distance provides a natural limit for nonradiative decay. In an absorbing matrix, however, the Green tensor for the homogeneous dielectric that generalizes the vacuum part in Eq. (84) has a singular imaginary part at its origin. This divergence can be regularized by introducing a small empty cavity around the molecule, but the result depends sensitively on the cavity size [33, 110, 111] and also changes with the cavity shape [112, 113]. Inside a solid, one also has to take into account local field factors that link the field described by the macroscopic Maxwell equations to the excitation field at the site of the molecule. This correction is well-known for the model of a spherical cavity filled with dielectric material that is supposed to be small compared to the wavelength, so that the local field can be computed in the electrostatic approximation [1]. A different value for the local field is obtained with the model of an empty, spherical cavity [114].

It has been argued that for the power emitted into the far field, the local field correction is equivalent to a suitably renormalized dipole moment [33, 115]. In practice, one can thus avoid the explicit calculation of the local field correction by normalizing the radiation power emitted in a finite nanostructure to the power emitted in a homogeneous dielectric. It has been shown both experimentally and theoretically that a (nonabsorbing) nanoparticle reduces the spontaneous decay rate of an embedded molecule compared to its bulk value [116–118]. Even metallic nanoparticles can enhance the fluorescence scattering cross section if plasmon or cavity resonances amplify the local excitation intensity and overcompensate for the nonradiative loss [82, 119].

For the frequency shift of the molecular transition, similar regularization procedures are required. Even in free space, this shift is formally infinite and must be renormalized by replacing a bare transition frequency with the physically observed one [6]. This divergence also occurs close to a solid structure, but it can be disposed of using the same renormalization argument. It turns out that only the reflected or scattered part of the Green tensor in Eq. (84) provides the interesting shift of the transition frequency as a result of the solid environment. At a finite distance from the solid, the scattered Green tensor does not show any divergence, and finite results are directly obtained [50].

4. ATOM CHIPS AND THERMAL NEAR-FIELD NOISE

In this section, we consider cold atoms trapped in tightly confining electromagnetic fields a few micrometers above a nanostructured substrate (an "atom chip" [10, 11]). As has become clear from the previous sections, the electromagnetic field shows enhanced fluctuations in the near field. We outline here the loss and decoherence mechanisms these fluctuations imply. They are particularly important to estimate when atoms in microtraps are used as an implementation of a register of qubits in a quantum computer. We focus on the coupling between the field and the permanent magnetic dipole moment that many atoms with nonzero spin possess.

4.1. Atom-Field Coupling

4.1.1. Zeeman Interaction

For weak magnetic fields, the interaction between the magnetic dipole moment $\boldsymbol{\mu}$ of an atom and the magnetic field is given by the Zeeman potential

$$V = -\boldsymbol{\mu} \cdot \mathbf{B}(\mathbf{r}, t) \tag{123}$$

where $\mathbf{r}$ is the atomic position. The magnetic moment is linked to the (dimensionless) total atomic spin operator $\mathbf{J}$ by the Bohr magneton and the Landé factor

$$\boldsymbol{\mu} = \mu_B g \hbar \mathbf{J} \tag{124}$$

A typical electromagnetic trap above a structured substrate is based on a static magnetic quadrupole field created by the superposition of fields generated by wires deposited on the substrate, and possibly by an external homogeneous field [10, 19]. Pure quadrupole fields are actually avoided in atomic microtraps because in the vicinity of the field zero, the magnetic moment easily flips its direction with respect to the local field, leading to trap losses.

The magnetic field in Eq. (123) contains a static part $\mathbf{B}_0(\mathbf{r})$ that describes the quadrupole trap, in addition to a time-dependent part when quantum or thermal fluctuations are allowed for. The relevant frequency scale is set by the energy splitting $\hbar\omega_L = \mu_B g \hbar |\mathbf{B}_0(\mathbf{r}_t)|$ at the trap center $\mathbf{r}_t$. Typical values for the Larmor frequency $\omega_L/(2\pi)$ are in the range of 1 . . . 100 MHz, corresponding to vacuum wavelengths on the order of meters. From Section 2.5, it is clear that near-field fluctuations dominate in these traps over the blackbody field. In addition, the field fluctuations are essentially thermal because the high-temperature limit $k_B T \gg \hbar\omega$ applies in this frequency range unless the solid substrate is cooled to very low temperature.

4.1.2. Magnetic Trapping Potential

The depth of typical magnetic quadrupole traps is of the order of 1 mK, so that laser cooling is required to load them with cold atoms. If the atoms move sufficiently slowly in the magnetic field, their magnetic moment adiabatically follows the locally changing field direction. The magnetic quantum number m_J is conserved, and we can introduce the effective trapping potential

$$V_{m_J}(\mathbf{r}) = -\mu_B g \hbar m_J |\mathbf{B}_0(\mathbf{r})| \approx \text{const.} + \frac{M}{2} \sum_{i=1,2,3} \Omega_i^2 (x_i - x_{t,i})^2 \tag{125}$$

We have assumed that $g m_J < 0$ so that a magnetic field minimum provides a potential well (weak-field seekers). This well is harmonic for the lowest states with (generally different) oscillation frequencies Ω_i; M is the atomic mass. In micrometer-sized magnetic traps, one can reach typical values $\Omega_i/(2\pi) \approx 1$ kHz . . . 1 MHz. The spatial extension of a sample of trapped cold atoms is typically in the micrometer range. For more details, see the review papers by Reichel et al. [10] and Folman et al. [11].

4.2. Dissipation and Decoherence

4.2.1. Spin Flip Rate in Broad Band Fields

When broad band magnetic field fluctuations couple to the trapped atom, they induce processes like spin relaxation and decoherence. The transitions induced between magnetic sublevels $|m_J\rangle$ [defined with respect to the static trapping field $\mathbf{B}_0(\mathbf{r})$] are called "spin flips." They lead to trap loss from a magnetic trap because typically only the weak-field-seeking sublevels with $g m_J < 0$ are trapped. Consider for simplicity a $J = 1/2$ spin. The state with quantum number $m_1 = -\frac{1}{2} \operatorname{sgn} g$ can be trapped in the adiabatic potential [Eq. (125)]. As soon as a transition to the state with $m_1 = +\frac{1}{2} \operatorname{sgn} g$ occurs, the adiabatic potential changes sign, and the atom is expelled from the trap center. We can identify the spin flip rate with a trap loss rate when the spin flip rate is small compared to the characteristic frequencies in the magnetic trap (i.e., the oscillation frequencies Ω_i); this has to be checked after the calculation.

The trap loss rate resulting from spin flips γ_{flip} is analogous to the two-level system decay rate γ introduced in Eq. (108). (Observe that the weak-field-seeking state has a higher energy.) If we denote by m_i, m_f the magnetic quantum numbers before and after the transition, one can show from Fermi's Golden Rule that the spin flip rate is given by an expression similar to Eq. (108)

$$\gamma_{\text{flip}} = \frac{1}{\hbar^2} \boldsymbol{\mu}_{fi}^* \cdot \mathcal{B}(\mathbf{r}_t, \mathbf{r}_t; -\omega_L) \cdot \boldsymbol{\mu}_{fi} \tag{126}$$

where $\boldsymbol{\mu}_{fi} = \mu_B g \hbar \langle m_f | \mathbf{J} | m_i \rangle$ is the magnetic transition moment, $\mathcal{B}$ is the magnetic fluctuation spectrum, and ω_L is the Larmor frequency at the trap center $\mathbf{r}_t$. Regarding the field, this rate is averaged over the initial states with the corresponding Boltzmann weight and summed over the final states (see Refs. [38] and [120]).

The atom is actually treated as a point particle in Eq. (126), ignoring the position spread of its wave function. If the atomic center-of-mass motion is quantized as well, one has to average the noise tensor over the position distribution of the trap eigenstates and has to incorporate the trap eigen energies in the transition frequency; see Eq. (134) below. Deviations from adiabaticity resulting from the motion of the atom in the trap can induce spin flips in a similar way. The corresponding flip rates can be made exponentially small by operating the trap in the adiabatic limit $|\omega_L| \gg \Omega_t$ [121–124]. In this regime, thermal field fluctuations and collisions between atoms become the dominant source of trap loss.

From the result [Eq. (101)] for the magnetic noise spectrum, we can estimate the spin flip rate above a metallic substrate [44]

$$\gamma_{\text{flip}} = \frac{(\mu_0 \mu_B g)^2 k_B T \sigma}{16\pi} |M_{if}|^2 \begin{cases} \dfrac{1}{z_t} & \text{for } z_t \ll \delta_L \\ \dfrac{\delta_L}{3 z_t^4} & \text{for } \delta_L \ll z_t \end{cases} \tag{127}$$

where σ is the substrate conductivity and $\delta_L \equiv \delta_{|\omega_L|}$ the skin depth in the metal at the Larmor frequency. The dimensionless matrix element M_{if} depends on the orientation of the field $\mathbf{B}_0(\mathbf{r}_t)$ at the trap center relative to the surface normal, and is given explicity in [44]. An expression that covers the intermediate range $z_t \sim \delta_L$ in the form of an integral for the magnetic Green tensor can also be found there.

To quote an estimate, for a trap distance z_t in the micron range and typical metallic conductivities, the loss rate is on the order of 1 s^{-1} or larger. This is much slower than the oscillatory motion in a microtrap, but still within reach of experiments with ultracold atom samples. Equation (127) and its intermediate-range generalization could indeed be quantitatively confirmed by the groups of Hinds [125], Cornell [126], and Vuletić [127].

4.2.2. Spin Dephasing

The simultaneous trapping of different spin states is very interesting for applications in quantum information processing. A quantum bit of information (qbit) can be encoded in two orthogonal spin states of an atom, and two qbits can be coupled by controlled collisions between the atoms [128, 129]. Any unitary transformation in the joint Hilbert space of two (or more) qbits can be generated by combining single-atom and atom-pair operations. Quantum information processing with long qbit registers is expected to outperform classical processors because, thanks to the superposition principle, calculations can be performed in a massively parallel way. For a review, see the textbook by Alber, Beth, and Horodecki [130].

It is clear that in this context, a basic requirement is to maintain the coherence of the superposition state of a qbit. For trapped atoms as physical implementations of the qbit, we estimate here the corresponding timescale resulting from the coupling to magnetic field fluctuations. The coherence of a qbit state can be quantified from the correlation function $\rho_{01} = \langle c_0 c_1^* \rangle$ of the probability amplitudes in a typical superposition state $|\psi\rangle = c_0|0\rangle + c_1|1\rangle$. The average is taken over the fluctuations of the magnetic field, or more generally over the environment with which the qbit interacts with. The quantity ρ_{01} is in fact an off-diagonal

element (called a coherence) of the 2×2 reduced-density matrix $\rho(t)$ of the two-state system. The expectation values of any observable $\mathcal{A}$ is encoded in the density matrix via the rule

$$\langle \mathcal{A} \rangle = \mathrm{tr}[\mathcal{A}\rho(t)] \tag{128}$$

The coupling to the environment leads to the following non-Hamiltonian dynamics for the density matrix, making the Born and Markov approximations [13, 80]

$$\left.\frac{\mathrm{d}\rho}{\mathrm{d}t}\right|_{\mathrm{diss}} = -\frac{1}{\hbar^2} \lim_{\Delta t \gg \tau_{\mathrm{corr}}} \int_0^{\Delta t} \mathrm{d}\tau \langle [H_{\mathrm{int}}(t), [H_{\mathrm{int}}(t-\tau), \rho(t)]] \rangle \tag{129}$$

where $H_{\mathrm{int}}(t)$ is the interaction Hamiltonian between the system and the environment. Its time-dependence corresponds to the free or uncoupled evolution (interaction picture). The average is taken in the equilibrium state for the environment. The correlation time τ_{corr} of the environment fluctuations is supposed to be smaller than the timescale Δt for the system's evolution (Markov approximation). Equation (129) generates what is called the master equation for the system density matrix. For a two-state atom coupled to the electric or magnetic field, it yields the rate Eqs. (104) and (105) and the Bloch Eq. (106), or alternatively the spin flip rate from Eq. (126). We apply Eq. (129) here to estimate the decoherence rate of a qbit superposition state.

In a magnetic trap with static trapping field $\mathbf{B}_0(\mathbf{r}_t)$, the qbit states are two eigenstates of the component $\mu_\parallel$ of the magnetic dipole operator that is parallel to the direction $\widehat{\mathbf{B}}_0(\mathbf{r}_t)$ of the trapping field. The interaction Hamiltonian with magnetic field fluctuations is of the form

$$H_{\mathrm{int}}(t) = -\boldsymbol{\mu}_\perp(t) \cdot \mathbf{B}_\perp(t) - \mu_\parallel B_\parallel(t) \tag{130}$$

where the subscripts $\perp$, $\parallel$ denote the components perpendicular and parallel to the static trapping field. All fields are evaluated here at the trap center for simplicity. The operator $\mu_\parallel$ commutes with the operator for the static trap potential, $-\mu_\parallel B_0$, and therefore, it does not evolve with time in the interaction picture.

The perpendicular field fluctuations in Eq. (130) lead to spin flips with the rate from Eq. (126). One half of this rate contributes also to the decay of the coherence ρ_{01}, as we have seen in the Bloch Eq. (106). The decay of ρ_{01} is accelerated by coupling to the parallel field fluctuations in Eq. (130). It is clear that these fluctuations do not induce any transitions between the qbit states. However, off-diagonal elements of the density matrix are affected because the fluctuations randomly shift the energy difference, and hence the relative phase, between the states $|0\rangle$ and $|1\rangle$. This leads to a process commonly called dephasing. The dephasing contribution to the master equation resulting from Eq. (129) reads

$$\left.\frac{\mathrm{d}\rho}{\mathrm{d}t}\right|_{\mathrm{deph}} = -\frac{\mathcal{B}_\parallel(\mathbf{r}_t; 0)}{2\hbar^2}\left(\mu_\parallel^2\rho + \rho\mu_\parallel^2 - 2\mu_\parallel\rho\mu_\parallel\right) \tag{131}$$

$$\mathcal{B}_\parallel(\mathbf{r}_t; 0) = \widehat{\mathbf{B}}_0(\mathbf{r}_t) \cdot \lim_{\omega \to 0} \mathcal{B}(\mathbf{r}_t, \mathbf{r}_t; \omega) \cdot \widehat{\mathbf{B}}_0(\mathbf{r}_t) \tag{132}$$

where the magnetic correlation spectrum projected onto the static trapping field appears, taken at zero frequency. The off-diagonal element of the density matrix thus decays like

$$\left.\frac{\mathrm{d}\rho_{01}}{\mathrm{d}t}\right|_{\mathrm{deph}} = -\frac{\mathcal{B}_\parallel(\mathbf{r}_t; 0)}{2\hbar^2}(\mu_{\parallel 0} - \mu_{\parallel 1})^2\rho_{01} \tag{133}$$

where $\mu_{\parallel 0}$ and $\mu_{\parallel 1}$ are the eigenvalues of $\mu_\parallel$ in the two qbit states. If these values differ, the coherence of the qbit superposition state thus decays with a rate proportional to the magnetic noise spectrum at zero frequency. We have seen in Eq. (101) that above a metallic substrate, the noise spectrum is approximately flat. In this case, the dephasing timescale is comparable to the lifetime with respect to spin flip processes [120].

A much longer coherence time is achieved with qbit states that have the same magnetic moment $\mu_{\parallel 0} = \mu_{\parallel 1}$. Such states exist to a good approximation in the hyperfine manifolds of

the alkali atoms, for example. The states have the same orientation of the electron spin, but opposite orientations of the nuclear spin, the magnetic moment of the latter being much smaller than μ_B. Long spin coherence times between these states have been observed in atom-chip experiments performed by Reichel and co-workers [131].

4.2.3. Vibrational Decoherence

Finally, we summarize the decoherence rates for the center of mass motion in a magnetic trap. The center of the mass coordinate is quantized, and the basic quantity characterizing the coherence of the atomic wave function is the correlation function $\langle \psi^*(\mathbf{x}, t)\psi(\mathbf{x}', t)\rangle$. It is the generalization of the density matrix to observables with a continuous spectrum. In a trap, the density matrix can be expanded onto a complete set of stationary wave functions $\{\varphi_n(\mathbf{x})\}$. Magnetic noise drives transitions between these states at a rate

$$\gamma_{n\to n'} = (\mu_B g)^2 \int d^3x\, d^3x' M_{nn'}(\mathbf{x}) \widehat{\mathbf{B}}_0(\mathbf{x}) \cdot \mathcal{B}(\mathbf{x}, \mathbf{x}'; -\omega_{nn'}) \cdot \widehat{\mathbf{B}}_0(\mathbf{x}') M^*_{nn'}(\mathbf{x}') \tag{134}$$

$$M_{nn'}(\mathbf{x}) = \varphi^*_n(\mathbf{x})\varphi_{n'}(\mathbf{x}) \tag{135}$$

where $\hbar\omega_{nn'} = E_n - E_{n'}$ is the Bohr transition frequency. We have allowed for a spatially changing direction of the static trapping field $\widehat{\mathbf{B}}_0(\mathbf{x})$. In tightly confined traps, we can expand the magnetic correlation function with respect to the deviations $\mathbf{x} - \mathbf{r}_t$ from the trap center. This expansion allows us to define the noise correlation length l_{corr} from

$$\mathcal{B}_{ij}(\mathbf{x}, \mathbf{x}'; \omega) \approx \mathcal{B}_{ij}(\mathbf{r}_t, \mathbf{r}_t; \omega)\left[1 - \frac{(\mathbf{x} - \mathbf{x}')^2}{l^2_{corr}}\right] \tag{136}$$

For $n \neq n'$, only the second term gives a nonzero contribution to the transition rate $\gamma_{n\to n'}$. This is intuitively clear because the magnetic field has to show some spatial gradient to couple different trap eigenstates.

In a harmonic trap, one finds, using the expansion (136), the selection rules for dipole transitions (exchange of a single vibrational quantum). The transition rate scales like $(a_i/l_{corr})^2$ compared to the spin flip rate, where $a_i = [\hbar/(2M\Omega_i)]^{1/2}$ is the size of the ground state along the x_i coordinate [11]. A short correlation length (i.e., a "rough" noise potential) excites the atomic motion more efficiently. If the atom is cooled to the ground state $|0\rangle$ of the trapping potential, a qbit can be encoded in $|0\rangle$ and the first excited vibrational state $|1\rangle$. The rate $\gamma_{0\to 1}$ gives the limit to the inverse lifetime of the ground state because of magnetic fluctuations. The energy per atom and unit time fed into the trap in the form of vibrational quanta is proportional to $\gamma_{0\to 1}$ as well. Finally, this rate also specifies the decoherence of a superposition of the lowest trap states $|0\rangle$ and $|1\rangle$, a figure of much interest for quantum information storage. More generally, the off-diagonal matrix element $\rho_{nn'}$ decays with a rate comparable to $\gamma_{n\to n'}$. Further details can be found in Refs. [120, 132] and in the review by Folman et al. [11].

We mentioned in Section 2.5.3.4 that the correlation length of thermal magnetic noise is on the order of the observation distance to the surface in the near-field regime (wavelength much larger than distance). This implies that $l_{corr} \sim z_t$ for atom-chip traps. The factor $(a_i/l_{corr})^2 \sim (a_i/z_t)^2$ in the heating and decoherence rates for the center of mass motion thus leads to a scaling with different power laws compared to the spin flip and spin dephasing rates. Because the ground-state size a_i in tightly confining traps is much smaller than the trap distance, heating and decoherence also happen more slowly.

5. MECHANICAL EFFECTS ON NANOPARTICLES

In this section, we briefly summarize recent developments regarding mechanical forces at nanometer distances from a solid surface, where electromagnetic field fluctuations have been invoked to play a role. Our selection is motivated by challenges that appear to be still open for theorists in the domain.

The best known mechanical effect is probably the Casimir force [133]: The attraction between two parallel, perfectly conducting plates that arises from the boundary conditions

imposed by the plates on the quantum fluctuations of the field. Casimir's paper triggered enormous theoretical activity (see the bibliographic resource paper by Lamoreaux [134] for an overview). Recent experiments have been able to test the theoretical predictions to better 10% [135–138], putting theorists on the road toward a precise evaluation of the effect of finite temperature, finite material conductivity, nonplanar shape, surface roughness, and so forth [22, 46, 139–148]. In the context of nanomechanical systems, the Casimir attraction is of particular interest because it can spoil the proper operation of the system by sticking together components separated by a few nanometers [149].

Electromagnetic fields have also been put forward to explain the friction force on a probe particle like a scanning tip that moves at nanometer distance above a solid surface [150–153]. One of the motivations is identifying a microscopic mechanism for the shear forces observed in scanning probe microscopy [154, 155]. It is still controversial whether electromagnetic fields can provide a sufficiently large effect. If the shear force derives from the electromagnetically induced friction between solids, however, then the system has to be at nonzero temperature [153, 156], in contrast to the Casimir force that persists even at zero temperature. The substrate below a metallic coating also appears to play a significant role [156]. The friction force to linear order in the velocity vanishes at zero temperature, but nonzero higher-order contributions have been found for atoms approaching a metallic surface [157].

Finally, we mention that radiation-induced forces provide a promising route for the nondestructive, controlled manipulation of nanoparticles and even biological objects like cells, bacteria, or cell subunits. One typical example is the optical tweezer in which a nanoparticle is attracted toward the maximum of the light intensity and can be dragged along with a focused light beam. A proper theoretical model is based on either the self-consistent evaluation of the Maxwell stress tensor at the particle surface or on an approximate multipole expansion for subwavelength particles. In addition to the laser-induced force, vacuum and thermal fluctuations of the field also give a contribution similar to the Casimir or the Van der Waals force. This contribution can, in particular, become significant at nanometer distances from a solid structure [136, 137, 158].

6. CONCLUSION

The near field of nanostructures reveals valuable information when it is detected with submicrometer probes. Optical properties like the local refractive index can be retrieved from single-molecule spectroscopy, using the fluorescence dynamics on electric dipole transitions. Spatial resolution on the nanometer scale can be achieved because the standard diffraction limit does not apply to the nonpropagating modes that dominate the electromagnetic near field. Magnetic field fluctuations on the micrometer scale can be detected with high sensitivity using cold atoms trapped in electromagnetic potentials above micro- or nanostructured substrates called atom chips. These fluctuations determine the timescales for stable and coherent trapping and establish practical limits for quantum information processing with atom chips.

One of the main recent results is that on the subwavelength scale, the near field behaves differently from blackbody radiation. It is characterized by an enhanced spectral density resulting from the contribution of nonpropagating modes. The material properties of the nearby nanostructure shape the features of the spectrum, with electromagnetic surface resonances giving significant peaks. The spatial behavior of the field is dictated by the geometry of the nanostructure and the detection setup, as retardation plays no role in the subwavelength domain. The same is true for the field's spatial coherence, as characterized by the two-point correlation function. The typical scales that emerge are related to the validity of local material susceptibilities, to the nanostructure corrugation, and to the attenuation of resonant surface modes, in addition to the distance of the molecular probe. The skin depth of the material defines a crossover distance for the electric and magnetic near-field energy densities: qualitatively, the nanostructure resembles a half-space (a layer) on distances much smaller (much larger) than the skin depth, respectively.

The thermal and quantum fluctuations of the field are characterized by the macroscopic Maxwell equations under quite general circumstances under which the fluctuation dissipation theorem applies. The electric and magnetic Green tensors encode the relevant information,

with the reflection or scattering from the solid structure providing the modification with respect to the far field. Nonequilibrium situations can be handled with the assumption that the nanostructure be in local thermodynamic equilibrium. Electromagnetic noise is then determined by the fluctuations of the material polarization and magnetization, whose strength is fixed by the local temperature and the absorption spectrum. At the quantum level (equivalently, for zero temperature), these material fluctuations are at the origin of the nonradiative decay of a molecule probing the near field.

ACKNOWLEDGMENTS

I thank numerous colleages with whom I learned a significant fraction of the physics summarized here. Apologizing in advance for omissions, I would like to mention the following persons: Alain Aspect, Rémi Carminati, Claudia Eberlein, Jens Eisert, Ron Folman, Jean-Jacques Greffet, Karl Joulain, Astrid Lambrecht, Klaus Mølmer, Serge Reynaud, Lavinia Rogobete, Vahid Sandoghdar, Jörg Schmiedmayer, and Martin Wilkens. Lukas Novotny, Stefan Scheel, and Marin Tomaš provided valuable feedback and suggestions to this contribution. This work has been supported by the Deutsche Forschungsgemeinschaft DFG, by the Deutscher Akademischer Austauschdienst DAAD, and by the European Commission through the Networks Atom Chip Quantum Processor (ACQP) and Field–Atom–Surface Interactions (FASTNET).

REFERENCES

1. J. D. Jackson, "Classical Electrodynamics," 2nd ed. Chap. 7. Wiley, New York, 1975.
2. L. Knöll, S. Scheel, and D.-G. Welsch, in "Coherence and Statistics of Photons and Atoms" (J. Perina, Ed.), Wiley, New York, 2001.
3. L. D. Landau and E. M. Lifshitz, "Statistical Physics I," 2nd edn, Pergamon, Oxford, 1984.
4. H. B. Callen and T. A. Welton, *Phys. Rev.* 83, 34 (1951).
5. O. Keller, *Phys. Rep.* 268, 85 (1996).
6. D. P. Craig and T. Thirunamachandran, "Molecular Quantum Electrodynamics," Academic, London, 1984.
7. R. C. Dunn, *Chem. Rev.* 99, 2891 (1999).
8. J. P. Dowling and J. Gea-Banacloche, *Adv. At. Mol. Opt. Phys.* 37(Suppl. 3), 1 (1997).
9. E. A. Hinds and I. G. Hughes, *J. Phys. D: Appl. Phys.* 32, R119 (1999).
10. J. Reichel, W. Hänsel, P. Hommelhoff, and T. W. Hänsch, *Appl. Phys. B* 72, 81 (2001).
11. R. Folman, P. Krüger, J. Schmiedmayer, J. H. Denschlag, and C. Henkel, *Adv. At. Mol. Opt. Phys.* 48, 263 (2002).
12. R. Loudon, "The Quantum Theory of Light," Oxford University Press, London, 1983.
13. P. Meystre and M. Sargent III, "Elements of Quantum Optics," 3rd ed. Springer, Berlin, 1999.
14. L. Mandel and E. Wolf, "Optical Coherence and Quantum Optics," Cambridge University Press, Cambridge, 1995.
15. R. R. Chance, A. Prock, and R. Silbey, in "Advances in Chemical Physics XXXVII" (I. Prigogine and S. A. Rice, Eds.), pp. 1–65. Wiley, New York, 1978.
16. H. Metiu, *Progr. Surf. Science* 17, 153 (1984).
17. P. Meystre, "Atom Optics," Springer, Berlin, 2001.
18. C. J. Bordé, *Adv. At. Mol. Opt. Phys.* 37(Suppl. 3), 257 (1997).
19. J. Schmiedmayer, *Eur. Phys. J. D* 4, 57 (1998).
20. G. Birkl, F. B. J. Buchkremer, R. Dumke, and W. Ertmer, *Opt. Commun.* 191, 67 (2001).
21. R. Frisch and E. Segrè, *Zeitschr. für Physik* 75, 610 (1933).
22. V. M. Mostepanenko and N. N. Trunov, "The Casimir Effect and Its Applications," Oxford Science Publications, Oxford, 1997.
23. D. Polder and M. V. Hove, *Phys. Rev. B* 4, 3303 (1971).
24. J.-P. Mulet, K. Joulain, R. Carminati, and J.-J. Greffet, *Appl. Phys. Lett.* 78, 2931 (2001).
25. J. A. Sidles, J. L. Garbini, W. M. Dougherty, and S.-H. Chao, Proc. *IEEE* 91, 799 (2003).
26. G. W. Ford and W. H. Weber, *Phys. Rep.* 113, 195 (1984).
27. L. Onsager, *Phys. Rev.* 37, 405 (1931).
28. M. Born and E. Wolf, "Principles of Optics," 6th ed. Pergamon Press, Oxford, 1959.
29. R. Sprik, B. A. van Tiggelen, and A. Lagendijk, *Europhys. Lett.* 35, 265 (1996).
30. K. Sakoda, "Optical Properties of Photonic Crystals," Springer, Berlin, 2001.
31. K. Joulain, R. Carminati, J.-P. Mulet, and J.-J. Greffet, *Phys. Rev. B* 68, 245405 (2003).
32. P. de Vries, D. V. van Coevorden, and A. Lagendijk, *Rev. Mod. Phys.* 70, 447 (1998).
33. M. S. Tomaš and Z. Lenac, *Phys. Rev. A* 56, 4197 (1997).
34. J.-J. Greffet and R. Carminati, *Progr. Surf. Sci.* 56, 133 (1997).

35. S. M. Rytov, Y. A. Kravtsov, and V. I. Tatarskii, "Elements of Random Fields," Principles of Statistical Radiophysics, Vol. 3, Springer, Berlin, 1989.
36. J. W. Goodman, "Statistical Optics," Wiley, New York, 1985.
37. C. W. Gardiner, "Handbook of Stochastic Methods," Springer, Berlin, 1983.
38. J. M. Wylie and J. E. Sipe, *Phys. Rev. A* 30, 1185 (1984).
39. C. H. Henry and R. F. Kazarinov, *Rev. Mod. Phys.* 68, 801 (1996).
40. G. S. Agarwal, *Phys. Rev. A* 11, 230 (1975).
41. M. Nieto-Vesperinas, "Scattering and Diffraction in Physical Optics," Wiley, New York, 1991.
42. P. Yeh, "Optical Waves in Layered Media," Wiley, New York, 1988.
43. M. S. Tomaš, *Phys. Rev. A* 51, 2545 (1995).
44. C. Henkel, S. Pötting, and M. Wilkens, *Appl. Phys. B* 69, 379 (1999).
45. T. Varpula and T. Poutanen, *J. Appl. Phys.* 55, 4015 (1984).
46. J. Schwinger, J. Lester L. DeRaad, and K. A. Milton, *Ann. Phys. (N.Y.)* 115, 1 (1978).
47. G. S. Agarwal, *Phys. Rev. A* 11, 243 (1975).
48. G. S. Agarwal, *Phys. Rev. A* 11, 253 (1975).
49. G. S. Agarwal, *Phys. Rev. A* 12, 1475 (1975).
50. J. M. Wylie and J. E. Sipe, *Phys. Rev. A* 32, 2030 (1985).
51. B. Huttner, J. J. Baumberg, and S. M. Barnett, *Europhys. Lett.* 16, 177 (1991).
52. B. Huttner and S. M. Barnett, *Europhys. Lett.* 18, 487 (1992).
53. B. Huttner and S. M. Barnett, *Phys. Rev. A* 46, 4306 (1992).
54. S. M. Barnett, in "Quantum Fluctuations" (S. Reynaud, E. Giacobino, and J. Zinn-Justin, Eds.), pp. 137–179, Elsevier, Amsterdam, 1997.
55. T. Gruner and D.-G. Welsch, *Phys. Rev. A* 51, 3246 (1995).
56. T. Gruner and D.-G. Welsch, *Phys. Rev. A* 53, 1818 (1996).
57. H. T. Dung, L. Knöll, and D.-G. Welsch, *Phys. Rev. A* 57, 3931 (1998).
58. S. Scheel, L. Knöll, and D.-G. Welsch, *Phys. Rev. A* 58, 700 (1998).
59. A. Tip, *Phys. Rev. A* 56, 5022 (1997).
60. A. Tip, *Phys. Rev. A* 57, 4818 (1998).
61. A. Tip, L. Knöll, S. Scheel, and D.-G. Welsch, *Phys. Rev. A* 63, 043806 (2001).
62. K. H. Drexhage, in "Progress in Optics XII" (E. Wolf, Ed.), pp. 163–232, North-Holland, Amsterdam, 1974.
63. K. G. Sullivan and D. G. Hall, *J. Opt. Soc. Am. B* 14, 1149 (1997).
64. A. A. Maradudin and D. L. Mills, *Phys. Rev. B* 11, 1392 (1975).
65. G. S. Agarwal, *Phys. Rev. B* 15, 2371 (1977).
66. H. Ogura and N. Takahashi, *J. Opt. Soc. Am. A* 2, 2208 (1985).
67. R. G. Llamas, *J. Opt. Soc. Am. B* 11, 618 (1994).
68. Z. L. Wang, H. Ogura, and N. Takahashi, *J. Opt. Soc. Am. A* 12, 1489 (1995).
69. B. N. J. Persson and N. D. Lang, *Phys. Rev. B* 26, 5409 (1982).
70. B. N. J. Persson and S. Andersson, *Phys. Rev. B* 29, 4382 (1984).
71. P. J. Feibelman, *Progr. Surf. Sci.* 12, 287 (1982).
72. H. Raether, "Surface Plasmons on Smooth and Rough Surfaces and on Gratings," Springer Tracts in Modern Physics, Vol. 111, Springer, Berlin, 1988.
73. R. Carminati and J.-J. Greffet, *Phys. Rev. Lett.* 82, 1660 (1999).
74. A. V. Shchegrov, K. Joulain, R. Carminati, and J.-J. Greffet, *Phys. Rev. Lett.* 85, 1548 (2000).
75. J.-Y. Courtois, J.-M. Courty, and J. C. Mertz, *Phys. Rev. A* 53, 1862 (1996).
76. W. L. Barnes, *J. Mod. Optics* 45, 661 (1998).
77. T. Setälä, M. Kaivola, and A. T. Friberg, *Phys. Rev. Lett.* 88, 123902 (2002).
78. T. Setälä, A. Shevchenko, M. Kaivola, and A. T. Friberg, *Phys. Rev. E* 66, 016615 (2002).
79. C. Henkel, K. Joulain, R. Carminati, and J.-J. Greffet, *Opt. Commun.* 186, 57 (2000).
80. N. G. van Kampen, "Stochastic Processes in Physics and Chemistry," revised ed, Elsevier, Amsterdam, 1992.
81. P. Lambropoulos, G. M. Nikolopoulos, T. R. Nielsen, and S. Bay, *Rep. Prog. Phys.* 63, 455 (2000).
82. J. Enderlein, *Phys. Chem. Chem. Phys.* 4, 2780 (2002).
83. C. Girard, O. J. F. Martin, and A. Dereux, *Phys. Rev. Lett.* 75, 3098 (1995).
84. L. Novotny, *Appl. Phys. Lett.* 69, 3806 (1996).
85. C. Henkel and V. Sandoghdar, *Opt. Commun.* 158, 250 (1998).
86. L. Rayleigh, *Proc. R. Soc. Lond. A* 79, 399 (1907).
87. M. Nieto-Vesperinas, *J. Opt. Soc. Am.* 72, 539 (1982).
88. J.-J. Greffet, *Phys. Rev. B* 37, 6436 (1988).
89. D. V. Labeke and D. Barchiesi, *J. Opt. Soc. Am. A* 10, 2193 (1993).
90. R. Carminati and J.-J. Greffet, *J. Opt. Soc. Am. A* 12, 2716 (1995).
91. G. Parent, D. Van Labeke, and D. Barchiesi, *J. Opt. Soc. Am. A* 16, 896 (1999).
92. G. Parent, D. Van Labeke, and D. Barchiesi, *J. Microsc.* 194, 281 (1999).
93. A. Rahmani, P. C. Chaumet, F. de Fornel, and C. Girard, *Phys. Rev. A* 56, 3245 (1997).
94. A. Rahmani, P. C. Chaumet, and F. de Fornel, *Phys. Rev. A* 63, 023819 (2001).
95. J.-J. Greffet, A. Sentenac, and R. Carminati, *Opt. Commun.* 116, 20 (1995).
96. R. Carminati and J.-J. Greffet, *Opt. Commun.* 116, 316 (1995).
97. J. Michaelis, C. Hettich, A. Zayats, B. Eiermann, J. Mlynek, and V. Sandoghdar, *Opt. Lett.* 24, 581 (1999).
98. J. Michaelis, C. Hettich, J. Mlynek, and V. Sandoghdar, *Nature* 405, 325 (2000).

99. V. V. Klimov and V. S. Letokhov, *Chem. Phys. Lett.* 285, 313 (1998).
100. V. V. Klimov, M. Ducloy, and V. S. Letokhov, *Eur. Phys. J. D* 20, 133 (2002).
101. V. V. Klimov, M. Ducloy, and V. S. Letokhov, *Quantum Electronics* 31, 569 (2001).
102. C. Girard, C. Joachim, and S. Gauthier, *Rep. Prog. Phys.* 63, 893 (2000).
103. A. Taflove and S. C. Hagness, "Computational Electrodynamics: The Finite-Difference Time-Domain Method," 2nd ed. Artech House, Norwood, MA, 2000.
104. C. Hafner, "The Generalized Multiple Multipole Technique for Computational Electromagnetics." Artech House, Boston, 1990.
105. B. Sick, B. Hecht, U. P. Wild, and L. Novotny, *J. Microsc.* 202, 365 (2001).
106. H. Gersen, M. F. García-Parajó, L. Novotny, J. A. Veerman, L. Kuipers, and N. F. van Hulst, *Phys. Rev. Lett.* 85, 5312 (2000).
107. H. Gersen, M. F. García-Parajó, L. Novotny, J. A. Veerman, L. Kuipers, and N. F. van Hulst, *J. Microsc.* 202, 374 (2001).
108. S. C. Kitson, W. L. Barnes, and J. R. Sambles, *Phys. Rev. B* 52, 11441 (1995).
109. S. C. Kitson, W. L. Barnes, and J. R. Sambles, *Phys. Rev. Lett.* 77, 2670 (1996).
110. S. Scheel, L. Knöll, and D.-G. Welsch, *Phys. Rev. A* 60, 4094 (1999).
111. S. Scheel, L. Knöll, D.-G. Welsch, and S. M. Barnett, *Phys. Rev. A* 60, 1590 (1999).
112. F. J. P. Schuurmans, P. de Vries, and A. Lagendijk, *Phys. Lett. A* 264, 472 (2000).
113. A. Rahmani and G. W. Bryant, *Phys. Rev. A* 65, 033817 (2002).
114. R. J. Glauber and M. Lewenstein, *Phys. Rev. A* 43, 467 (1991).
115. M. S. Tomaš, *Phys. Rev. A* 63, 053811 (2001).
116. H. Chew, *Phys. Rev. A* 38, 3410 (1988).
117. H. Schniepp and V. Sandoghdar, *Phys. Rev. Lett.* 89, 257403 (2002).
118. L. Rogobete, H. Schniepp, V. Sandoghdar, and C. Henkel, *Opt. Lett.* 28, 1736 (2003).
119. M. Thomas, R. Carminati, J. J. Greffet, and J. R. Arias-Gonzalez, *Appl. Phys. Lett.* (in press).
120. C. Henkel, P. Krüger, R. Folman, and J. Schmiedmayer, *Appl. Phys. B* 76, 173 (2003).
121. E. H. Bergeman, P. McNicholl, J. Kycia, H. Metcalf, and N. Balazs, *J. Opt. Soc. B* 6, 2249 (1989).
122. C. V. Sukumar and D. M. Brink, *Phys. Rev. A* 56, 2451 (1997).
123. S. Gov, S. Shtrikman, and H. Thomas, *J. Appl. Phys.* 87, 3989 (2000).
124. E. A. Hinds and C. Eberlein, *Phys. Rev. A* 61, 033614 (2000); Erratum: *Phys. Rev. A* 64 (2001).
125. M. P. A. Jones, C. J. Vale, D. Sahagun, B. V. Hall, and E. A. Hinds, *Phys. Rev. Lett.* 91, 080401 (2003).
126. D. M. Harber, J. M. McGuirk, J. M. Obrecht, and E. A. Cornell, *J. Low Temp. Phys.* 133, 229 (2003).
127. Y.-J. Lin, I. Teper, C. Chin, and V. Vuletić, *Phys. Rev. Lett.* 92, 050404 (2004).
128. T. Calarco, E. A. Hinds, D. Jaksch, J. Schmiedmayer, J. I. Cirac, and P. Zoller, *Phys. Rev. A* 61, 022304 (2000).
129. J. Schmiedmayer, R. Folman, and T. Calarco, *J. Mod. Optics* 49, 1375 (2002).
130. G. Alber, T. Beth, and M. Horodecki, "Quantum Information. An Introduction to Basic Theoretical Concepts and Experiments." Springer Tracts in Modern Physics, Vol. 173. Springer, Heidelberg, 2003.
131. P. Treutlein, P. Hommelhoff, T. Steinmetz, T. W. Hansch, and J. Reichel, *Phys. Rev. Lett.* 92, 203005 (2004).
132. C. Schroll, W. Belzig, and C. Bruder, *Phys. Rev. A* 68, 043618 (2003).
133. H. B. G. Casimir, *Proc. Kon. Ned. Akad. Wet.* 51, 793 (1948).
134. S. K. Lamoreaux, *Am. J. Phys.* 67, 850 (1999).
135. S. K. Lamoreaux, *Phys. Rev. Lett.* 78, 5 (1997); Erratum: 81, 5475 (1998).
136. U. Mohideen and A. Roy, *Phys. Rev. Lett.* 81, 4549 (1998).
137. H. B. Chan, V. A. Aksyuk, R. N. Kleiman, D. J. Bishop, and F. Capasso, *Science* 291, 1941 (2001).
138. G. Bressi, G. Carugno, R. Onofrio, and G. Ruoso, *Phys. Rev. Lett.* 88, 041804 (2002).
139. D. Kupiszewska, *Phys. Rev. A* 46, 2286 (1992).
140. I. Dorofeyev, H. Fuchs, B. Gotsmann, and G. Wenning, *Phys. Rev. B* 60, 9069 (1999).
141. G. L. Klimchitskaya, U. Mohideen, and V. M. Mostepanenko, *Phys. Rev. A* 61, 062107 (2000).
142. C. Genet, A. Lambrecht, and S. Reynaud, *Phys. Rev. A* 62, 012110 (2000).
143. A. Lambrecht and S. Reynaud, *Eur. Phys. J. D* 8, 309 (2000).
144. R. Matloob and H. Falinejad, *Phys. Rev. A* 64, 042102 (2001).
145. R. Esquivel-Sirvent, C. Villarreal, W. L. Mochan, and G. H. Cocoletzi, *Phys. Stat. Sol. (b)* 230, 409 (2002).
146. M. S. Tomaš, *Phys. Rev. A* 66, 052103 (2002).
147. C. Genet, A. Lambrecht, P. M. Neto, and S. Reynaud, *Europhys. Lett.* 62, 484 (2003).
148. C. Raabe, L. Knöll, and D. G. Welsch, *Phys. Rev. A* 68, 033810 (2003).
149. E. Buks and M. L. Roukes, *Phys. Rev. B* 63, 033402 (2001).
150. J. B. Pendry, *J. Phys. Cond. Matt.* 9, 10301 (1997).
151. I. Dorofeyev, H. Fuchs, G. Wenning, and B. Gotsmann, *Phys. Rev. Lett.* 83, 2402 (1999); comment by B. N. J. Persson and A. I. Volokitin, 84, 3504 (2000); reply p. 3505.
152. G. V. Dedkov and A. A. Kyasov, *Phys. Lett. A* 259, 38 (1999).
153. A. I. Volokitin and B. N. J. Persson, *Phys. Rev. B* 65, 115419 (2002).
154. K. Karrai and I. Tiemann, *Phys. Rev. B* 62, 13 174 (2000).
155. B. C. Stipe, H. J. Mamin, T. D. Stowe, T. W. Kenny, and D. Rugar, *Phys. Rev. Lett.* 87, 096801 (2001).
156. J. R. Zurita-Sánchez, J. J. Greffet, and L. Novotny, *Phys. Rev. A* 69, 022902 (2004).
157. F. Sols and F. Flores, *Sol. State Commun.* 42, 687 (1982).
158. D. Langbein, "Theory of Van der Waals Attraction" Springer Tracts in Modern Physics, Vol. 72. Springer, Berlin, 1974.

CHAPTER 9

Molecular Nanomagnets

Jens Kortus

Max-Planck-Institut für Festkörperforschung, Stuttgart, Germany, and Institut de Physique et Chemie des Matériaux de Strasbourg, France

Andrei V. Postnikov

Institute of Metal Physics, Russian Academy of Sciences, Yekaterinburg, Russia, and Osnabrück University, Department of Physics, Osnabrück, Germany

CONTENTS

ISBN: 1-58883-049-7

Handbook of Theoretical and Computational Nanotechnology
Edited by Michael Rieth and Wolfram Schommers
Volume 7: Pages (503–562)

1. INTRODUCTION

Simple magnetic devices, like loudspeakers or refrigerator magnets, that one uses every day are often based on magnetic alloys. Other applications like magnetic tapes make use of nonconductive magnetic oxides. Portable devices and cell phones demand an increasing amount of storage capacity for music, pictures, and movies, so that further structural miniaturization of memory devices will be required in the future. Moreover, the use of more sophisticated effects, for instance, combining the action of magnetism and light, that are already used in magnetooptical read/write devices may open a path to new applications. Magnetic materials play a very important role in the current information technology: The storage density of magnetic hard drives doubles every year and reached 100 GBit/in^2 in 2003. Such storage density corresponds to an area of 80×80 nm to store a single bit. The price to store 1 MByte of data on a hard drive is now lower than what it costs to write it down on paper [1].

One possible road toward increasing the capacity for information storage is decreasing the system size of the storage devices. This suggests the eventual need for a transition from bulk matter to nanoscale molecules and clusters. To achieve that goal, one needs the ability to assemble these structures with the required electronic and magnetic properties in a controlled way. To a large extent, this explains the enduring interest in molecular nanomagnets. The effects of quantum physics become pronounced at the nanometer scale. It seems to be fair to say that the current research in this field still deals almost exclusively with fundamental basic understanding, rather than with applications. The understanding of how these materials at the nanoscale react to external parameters such as temperature, light, or electromagnetic fields is a basic prerequisite for their possible application in the future. The path to success will depend on our detailed knowledge of the nature of the interactions between the molecules and the properties of the molecules itself.

Following the late Olivier Kahn, who is considered by many as one of the founders of this research topic, "Molecular magnetism deals with magnetic properties of isolated molecules and/or assemblies of molecules" [2]. This definition is quite general, and recently there has been more emphasis on the aspect of the rational design of molecular magnetic properties in the field [3]. Therefore, molecular magnetism is seen "as a discipline which conceives, realizes, studies, and uses new molecular materials bearing new but predictable magnetic (and other) physical property" [4]. At present, it conveniently joins many different activities involving methods of physical characterization of matter (optical, X-ray, Mössbauer and neutron spectroscopies, scanning microscopies) and physical models of different degrees of sophistication and abstraction. The progress in the field is clearly driven by advances in chemical synthesis of the materials and experiments. However, the combined efforts of physicists and materials scientists, and particularly theorists, inspire confidence that such efforts may one day become useful not only for explaining but also for guiding the synthesis of new, promising materials. As for now, the theory dealing with first-principles calculations tries to keep pace with experiment, still trying to reproduce rather than to lead.

Nevertheless, there has been a notable progress in the prediction of exchange interactions and magnetic anisotropy energies from density-functional theory during the last few years. In contrast to cases in which magnetic exchange interactions follow the famous Goodenough–Kanamori rules [5–7], we still have to await similarly clear general insights derived from theory in the case of magnetic anisotropy, which is another crucial property of molecular magnets, which could revolutionize the rational design of molecular nanomagnets.

In general, the properties of a nanoscale system of coupled spins depend directly on the strength of the exchange coupling and of the spin-orbit coupling. When the exchange

interactions are large enough, the lowest-energy magnetic excitations in a many-spin system occur, in fact, as a result of collective changes in the spin-orbit coupling energy. The involved energy scale is known as the magnetic anisotropy energy [8]. The ability to accurately predict and modify the magnetic anisotropy energy is a key property one has to understand, because it determines the temperature range up to which a system will retain its magnetic orientation.

This field of research is very attractive for first-principles microscopic simulations, because the crystal structure of new molecular magnets is well defined and reproducible and is made available rapidly. This is in sharp contrast to the situation in nanoparticle materials or in surface studies, where the structure data are usually more ambiguous. Synthetic chemists and theorists performing *ab initio* simulations despite different skills already speak a common language when discussing chemical bonding or magnetic interactions.

Whereas the actual execution of first-principles calculations did not require any special development of basically new numerical schemes, certain difficulties in performing calculations specific to this type of materials provide interesting challenges to computational methods. In general, the computational task involves a large number of atoms (often several hundreds) in the unit cell, a practical absence of useful crystalline symmetry, and a very inhomogeneous spatial distribution of charge density, with "very dense" and "almost empty" regions. The progress on the methodological side of atomistic first-principle calculations include more efficient basis sets and new order–N algorithms, along with general augmentation of computational power, which helps address molecular magnets at the atomistic level. This task would have been too complicated to perform a decade ago.

The number of materials brought into discussion as molecular magnets is considerable, and some systematics might be appropriate to define better the subject of our present discussion. In general, all materials containing organic building parts and spins associated with unpaired electrons fall into one or another category of molecular magnets. Possible classifications may depend on the origin of the unpaired electrons, the resulting character of the magnetic moment (localized or itinerant), the type of interaction between individual moments, or their spatial organization (weakly coupled molecular fragments; one-, two- or three- dimensional connected structures). A well-structured general overview of different classes of molecular magnetic materials can be found in a special issue of the Materials Research Society on molecule-based magnets [9].

Magnetic molecules contain one or more transition metal centers or rare-earth ions, or just organic radicals that are locked at their lattice sites by a careful chemistry of surrounding organic fragments. We will not discuss purely organic magnets, where spins are carried by free radicals, although such systems clearly belong to the topic of molecular magnetism and ferromagnetic ordering with T_c of 35.5 K [10] has been demonstrated in them. In the following text, we restrict ourselves to systems in which the spins reside in $3d$ shells of transition metal ions.

In particular, we will concentrate on the so-called single-molecule magnets (SMM) [11], which are often also called molecular nanomagnets. Such materials can often be crystallized, but interactions between the molecular entities remain weak, so that the magnetic behavior probed by experiments is often dominated by intramolecular effects. The discovery of a molecule containing 12 manganese ions $Mn_{12}O_{12}(CH_3COO)_{16}(H_2O)_4$ with a magnetic ground state of $S = 10$ showing a magnetic hysteresis [11] because of the properties of the single Mn_{12}- molecule has boosted interest in the field enormously. The observed hysteresis in molecular magnets is not caused by remagnetization of domains, as in conventional ferromagnets, but instead, it is an intrinsic molecular property that results from the magnetic anisotropy. This process can only be observed because the relaxation time is very large compared to the measurement time. The relaxation of the magnetization becomes very slow indeed at low temperatures (on the order of several months at 2 K). Therefore, by applying a magnetic field, the material can be magnetized and will keep this magnetization for a long time if it is at very low temperatures. The observed steps in the hysteresis curve reflect the "magnetization tunneling" [12] between quantum states of different m, $-S \leq m \leq S$, of the total spin S of the molecule, as the external magnetic field realigns the degeneracies of different states. The measurements of the relaxation time of the magnetization as a function of an external magnetic field reveal much faster relaxation at certain field values, at which

one also observes the steps in the hysteresis curves [13]. We will return in more detail to this exciting experimental observation, which has been a cornerstone in the research in the field of molecular magnetism.

A single molecule behaves like a single magnetic domain and is relatively independent of the magnetization of its neighboring molecules. The broadly used term "nanomagnet" may be somewhat misleading in this case, in which magnetism is not caused by extended colective interactions, like in bulk magnets, and the fundamentally different nature of hysteresis has been already mentioned above. To be correct, one should keep these differences in mind when the term "magnet" is used.

One of the biggest advantages of SMM materials for observing quantum effects is the well-defined, and identical, size of molecular building blocks, because any size distribution would smear out these effects. This makes SMM materials also very interesting for possible application in nanotechnology, because the molecules have all the same size, shape, and magnetic properties, and they will be nearly perfectly ordered in the molecular crystal. This circumvents technological problems like the need to control the size distribution or the distances between clusters in a nanomaterial.

What is the present state of art, "figures of merit," perspectives, and so forth in the field of molecular magnets? The field is becoming rapidly too large to cover all aspects in a compact introduction. There are several recent good reviews by Verdaguer [3], Barbara and Gunther [14], Gatteschi and Sessoli [15], along with special issues of journals and conference proceedings [4, 9], which help to access the situation. Here we single out several promising directions.

The first one is using single-molecule units as "bits" for magnetic storage. The size of the molecules of interest is about one order of magnitude smaller than that of presently accessible domains in magnetic layers, and further miniaturization of conventional domains will be prevented at some point by approaching the superparamagnetic limit. This problem does not arise for magnetic molecules, because the intramolecular magnetic order is set by the chemistry of a single molecule in question, and not a result of achieving a certain critical size or a certain amount of magnetic atoms. To become practicable, this application needs moecules with a net total magnetic moment. This implies ferromagnetic or ferrimagnetic intramolecular ordering of sufficient strength to achieve a high spin ground state S. The intramoecular magnetic interactions have to be strong enough to prevent decoupling of the spins within the molecule by thermal fluctuations, so the single molecule effectively behaves like an atom with a giant spin S. Moreover, a high magnetic anisotropy is required to prevent spontaneous reorientation of the magnetization of the molecular unit i.e., to increase its blocking temperature. The intrinsic magnetic anisotropy becomes the temperature-determining figure of merit. The weakness of interactions between adjacent molecules, a prerequisite for writing magnetic bits independently in each molecule, is usually taken for granted in the case of single-molecule magnets. Potential candidates for such applications, albeit with properties not yet sufficient enough for any real application, are represented by Mn_{12}-acetate and Fe_8 molecules, both with $S = 10$, to be discussed below.

The second one is using working units for quantum computation. This requires a scheme to populate and manipulate excited states of a molecular magnet in a controllable way. Leuenberger and Loss [16] and Leuenberger et al. [17] proposed a seemingly feasible scheme of imposing a prepared electron spin resonance impulse to write in, transform, and read out the information on a quantum state in the multilevel system. In particular, they considered such a system with the global spin $S = 10$, explicitly referring to the two above mentioned SMMs, Mn_{12}-acetate and Fe_8. A promising technique may make use of the above mentioned magnetization tunneling [12].

The third one is using room-temperature molecule-based permanent magnets, very different in some aspects (solubility in various solvents, biocompatibility) from "conventional" (e.g., intermetallic-compounds) magnets, and possessing the additional advantage of exhibiting interesting combined magnetooptical and electro-optical properties. Many such systems are based on Prussian blue analogues [18–20]. Curie temperatures as large as 42° [21], 53° [18], and 103°C [22] have been achieved with V, Cr-based Prussian blue anaogues. Although such systems can be investigated by first principles calculations without major

problems (see, e.g., Ref. [23]), and despite the fact that they are generally recognized as molecular magnets, we leave them beyond the current discussion, because they are formed of extended metallorganic patterns. Such three-dimensional connectivity is, of course, essential for obtaining substantial values of T_c.

Finally, we have systems that exhibit novel collective phenomena such as magnetism switching by light, temperature, pressure, or other physical interactions. Molecules that exhibit spin-crossover behavior would, for instance, fall into that class. Many such systems are among Fe-binuclear complexes [24, 25]. The spin-crossover effects (switching between high-spin and low-spin states in different combinations at two Fe centers) are usually discussed in terms of interplay between intramolecular and intermolecular magnetic interactions, with the latter being smaller but not negligible.

After this general introduction into the field of molecular magnets, we proceed with a more specific discussion of their most important representative species, single-molecule magnets incorporating transition metal ions, in Section 2. Section 3 discusses the interesting and unusual properties, such as the quantum tunneling of magnetization, that make research in the field of molecular magnetism so exciting. In Section 4, we overview traditional model approaches for the phenomenological description of such systems and outline some ways used to extract parameters of the corresponding models from experiment and from first-principle calculations. The term "first-principle calculation" here is used either for multideterminantal quantum chemical schemes or for those based on density functional theory. In particular, we discuss Heisenberg exchange parameters and magnetic anisotropy constants. As a practical example, an introduction into the problems and properties of some single-molecule magnets that gained a great deal attention within last years, namely, Mn_{12}-acetate, Fe_8, and V_{15} systems, is given. This introduction is followed in Section 5 by a critical comparison of calculation schemes, based on density-functional theory, that are particularly well suited for the study of molecular magnets. For the above systems, we select some benchmark results obtained by different methods. Finally, in Section 6, we outline recent progress in the study of other single-molecule magnets, including six-membered "ferric wheels," "ferric stars," and "Ni_4" molecules, which were studied with the use of first-principles methods.

2. EXAMPLES OF TYPICAL SINGLE-MOLECULE MAGNETS

To make first contact with the materials we are discussing, we will now review the basic structural properties of some examples of the most intensely investigated SMM. From here on we will use Mn_{12}-ac as a shorthand for the complete chemical formula $[Mn_{12}O_{12}(CH_3COO)_{16}(H_2O)_4 \cdot 2CH_3COOH \cdot 4H_2O]$. There are now several modifications of Mn_{12}-ac that are known, with different crystal structures, solvent molecules, and water coordination (see Gatteschi and Sessoli [15] for more information). Basically, the inner structure of the molecule is always the same, and we will uniformly refer to all these species as Mn_{12}-ac. The Mn_{12}-ac was the first SMM that showed the slow magnetization relaxation characteristic for a SMM. This compound is probably the most investigated SMM and, along with the oxo-nuclear iron compound Fe_8, has shown so far many manifestations of interesting magnetic behavior that keep the research in the field growing. We will discuss the structure and magnetic properties of these two magnetic molecules in more detail, because one can regard them as a kind of test cases that the theory was able to explain. For a concise and basic introduction into the field of SMMs, we refer the reader to Barbara and Gunther [14].

2.1. Mn_{12}-ac Magnet

Mn_{12}-ac was synthesized and reported in 1980 by Lis [26]. The molecular crystal has tetragonal symmetry with space group $I\bar{4}$, and a single Mn_{12}-ac cluster in the crystal possesses S_4 symmetry. Figure 1 shows a ball-and-stick model of the molecular structure, including only some organic ligands. No water of crystallization or acetic acid molecules are included, although they may play an important role—in particular for the process of the tunneling of the magnetization. The manganese atoms are sixfold coordinated but show significant

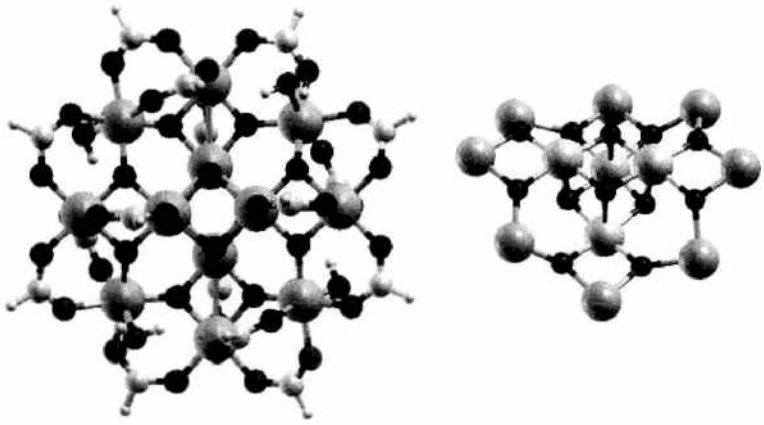


Figure 1. The ball and stick model of Mn_{12}-ac. Left panel: entire molecule, with methyl groups replaced for clarity by hydrogen atoms (small, light spheres). The large spheres represent Mn atoms and the dark ones oxygen. Right panel: magnetic core $Mn_{12}O_{12}$. The eight outer Mn ions have spins $s = 2$ ordered in parallel, and the four inner $s = 3/2$ are antiparallel to them, resulting in a ferrimagnetic structure with total spin $S = 10$.

Jahn–Teller-induced local O_h symmetry lowering because of partially filled e_g shells on the outer sites. The right panel of Fig. 1 shows only the magnetic core, for better clarity. The inner four Mn atoms that are in the charge state Mn^{4+} ($s = 3/2$) form, together with four O atoms, a (slightly distorted) cube. The eight outer Mn atoms are in the Mn^{3+} ($s = 2$) charge state. The inner Mn ions are coupled antiferromagnetically to the outer ones, yielding a ferrimagnetic groundstate with a total spin $S = 8 \times 2 - 4 \times 3/2 = 10$.

Evidence of the $S = 10$ ground state has been obtained from high field magnetization studies by Caneschi et al. [27] and was later confirmed by different experimental techniques, such as high field electron paramagnetic resonance (EPR) [28], high field magnetic torque measurements [29], or neutron scattering [30, 31].

The (outer) Mn^{3+} ions are distinguishable from manganese atoms in a different charge state by the elongated structure of the oxygen atom coordination octahedra or by the corresponding oxygen–manganese bond lengths that are typical for the Jahn–Teller distortions known in many Mn(III)-systems. This seems to be important for the magnetic anisotropy of the SMM [15].

A surprising feature of the Mn_{12} clusters is that they remain intact in solution. This has been demonstrated by nuclear magnetic resonance (NMR) measurements on several derivatives of the material [32]. This remarkable finding clearly indicates that the observed magnetic properties have indeed an intramolecular origin. This is further supported by specific heat measurements that found no evidence for long-range order in the material [33]. Each magnetic molecule in the crystal is well separated from its neighbors by water and acetic acid molecules; Barbara and Gunther [14] estimate the volume fraction of molecules in crystal to be merely 5%. The critical energy scale for the magnetic behavior is the magnetic anisotropy energy, which is on the order of 60 K. The dipole–dipole interaction between molecules is of about 0.03 meV, or 0.35 K, so that one can safely discard it for practical reasons and for setting up calculations.

2.2. Fe_8 Magnet

The octanuclear iron(III) molecular magnet of the chemical formula $[Fe_8O_2(OH)_{12}(tacn)_6]^{8+}$, with tacn = 1, 4, 7-triaza-cyclononane ($C_6N_3H_{15}$), is often referred to as the Fe_8-cluster. The structure of the Fe_8-molecular crystal, first synthesized by Wieghardt et al. [34], is shown in Fig. 2. It is acentric $P1$, with $a = 10.52$, $b = 14.05$, and $c = 15.00$ Å and $\alpha = 89.90$, $\beta = 109.65$, and $\gamma = 109.27$ degrees.

The approximate D_2 symmetry observed in the molecule [34] is formally broken by the presence of halide atoms and by waters of crystallization. The iron atoms form a structure that is often described as a butterfly. The central iron atoms are connected by oxo–hydroxo bridges to the four outer ones. The large spheres show the iron atoms, which are Fe(III) ions

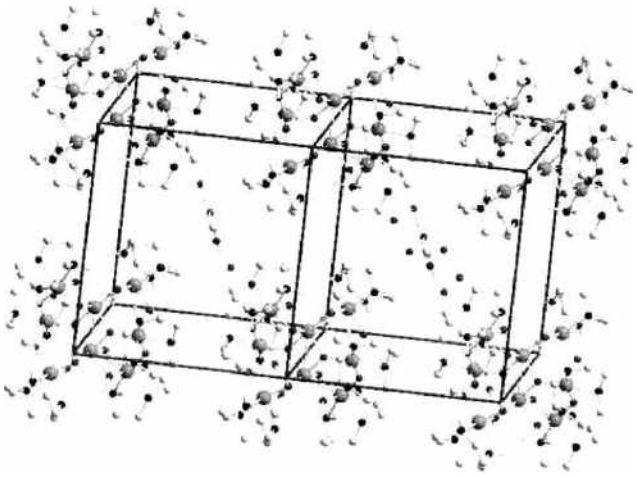


Figure 2. Repeated unit cells of the Fe_8 molecular crystal. Large balls represent Fe.

with a d^5 electron configuration. The two inner Fe(III) atoms are coordinated octahedrally to oxygen and the bridging hydroxy ligands. The outer iron atoms are also in octahedral coordination with the corresponding oxy and hydroxy ligands and nitrogen atoms of the tacn-rings. The organic tacn-rings are very important for stabilizing the magnetic core of the molecule because the three pairs of nitrogen dangling bonds complete a quasi-sixfold environment for the Fe atoms. Two of the Fe(III) atoms have antiparallel spin projections to the other six, so that the ferrimagnetic coupling of all eight results in the $S = 10$ spin ground state, as was directly proven by polarized neutron scattering measurements [35].

The tacn-rings separate the Fe_8-clusters in the crystal, resulting in negligible intermolecular dipole fields that are typically on the order of 0.05 T [36]. The resulting formal charge states are nominally Fe^{3+}, $(OH)^{-1}$ O^{-2}, and $tacn^0$, leading to a molecule with an overall formal charge state of +8, which must be compensated for by negatively charged halide ions. Because of its lower symmetry, as compared to Mn_{12}-ac, the Fe_8-cluster is allowed to have a transverse magnetic anisotropy, which is required to observe the quantum tunneling of the magnetization (QTM). This is because the transversal anisotropy is able to couple states with different m_s, which is a basic condition for "real" tunneling processes. In contrast, the tunneling in Mn_{12}-ac is often described as thermal, or phonon assisted, with dipolar and hyperfine interactions playing an important role [15].

One of the peculiar features of the Fe_8-cluster that make it particularly interesting is that its magnetic relaxation becomes temperature independent below 0.36 K, showing for the first time a pure QTM [36, 37]. Further, the topological quenching of the tunnel splitting predicted by Garg [38] has been observed in the form of a periodic dependence of the tunnel splitting on the magnetic field along the hard axis [37].

2.3. V_{15} Spin System

The $K_6[V_{15}As_6O_{42}(H_2O)] \cdot 8H_2O$ molecular crystal was first synthesized by Müller and Döring [39]. The V_{15} molecule comprises spins $s = 1/2$ at all vanadium atoms, which couple together to form a molecule with a total spin $S = 1/2$ ground state. The weakly anisotropic V_{15} demonstrates quantum behavior, such as tunneling splitting of low-lying spin states, and is an attractive model system for the study of the mesoscopic quantum coherence and the processes that destroy it. An understanding of such processes may be of interest for the field of quantum computing. V_{15} has a crystallographically imposed trigonal symmetry with three sets of inequivalent vanadium atoms [40]. They form two warped hexagons, separated by an

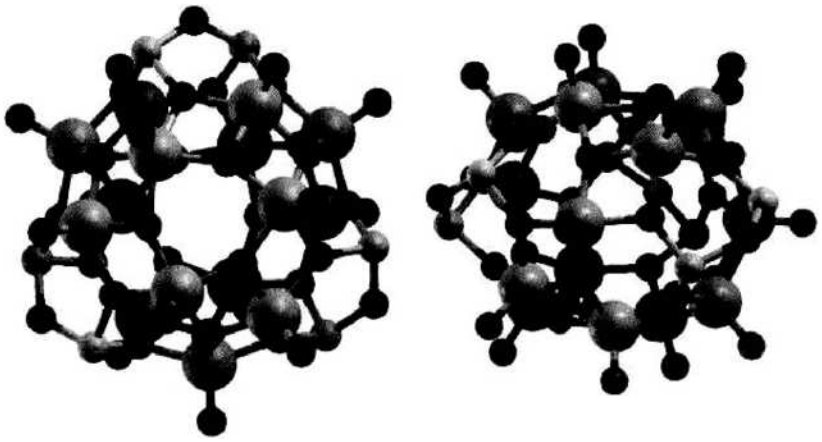


Figure 3. The top view (along the threefold axis, left panel) and the side view (right panel, note the sandwiched structure) of the ball-and-stick model of the V_{15} spin system. The large balls represent the 15 vanadium atoms. They all contain a single electron in the *d*-shell and couple in complicated ways to yield a total spin ground-state configuration of $S = 1/2$.

intermediate layer in the form of a large triangle. The vanadium atoms are hold in place by oxygens and arsenic atoms so that the complete cluster forms a ball- or cagelike structure (see Fig. 3). The empty space inside the cavity is often filled by a randomly oriented water molecule that, strictly speaking, would formally break the trigonal symmetry.

The unit cell contains two V_{15} clusters and is large enough so that dipolar interactions between them are negligible. Between 20 and 100 K, the effective paramagnetic moment is 3 μ_B, as for three independent spins, and below 0.5 K it changes to the $S = 1/2$ ground state. The experimental results were interpreted with antiferromagnetic interactions between all vanadium atoms [40]. To explain the magnetic behavior, a complicated spin Hamiltonian with at least five different exchange parameters J_{ij} is required [41]. Figure 4 shows only the V–O cage (left panel) and the V–V couplings, labeled according to notations of Kortus et al. [41].

Because of the layered structure and the trigonal axis, one expects that the V_{15} cluster will show interesting magnetic properties, such as a canted noncollinear magnetic ground state. Calculations on such correlated systems present a challenge to mean-field frameworks such

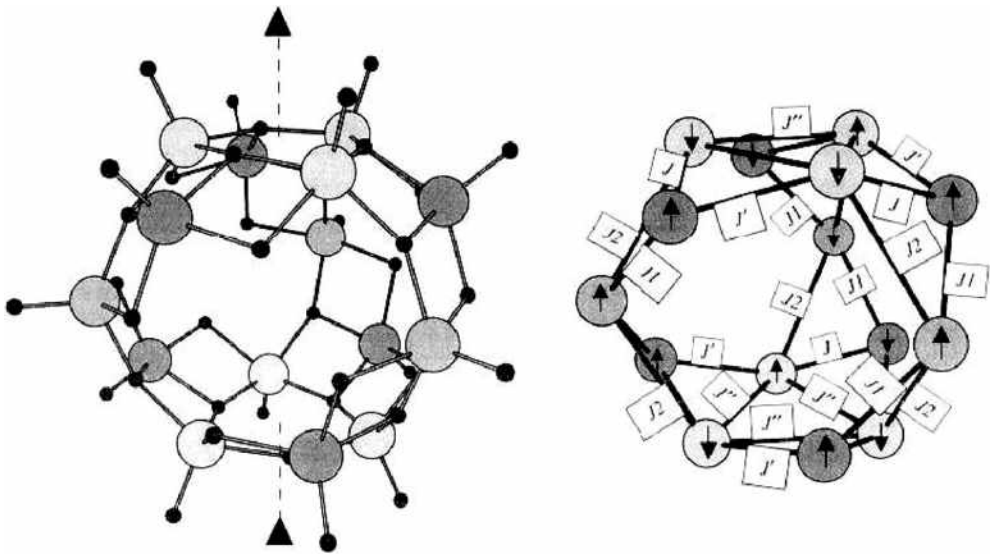


Figure 4. Left panel: V–O cage of the V_{15} molecule. Right panel: scheme of exchange couplings J, J', J'', $J1$, and $J2$ needed for the description of the magnetic behavior of this compound. The spin configuration shown has a total spin $S = 1/2$, corresponding to the lowest-energy density-functional theory Ising spin configuration. Different gray scale indicates structurally inequivalent atoms.

as density-functional theory because often it is not possible to construct a single collinear reference state that preserves the inherent symmetry of the system and that has the correct spin quantum numbers.

3. UNUSUAL AND INTERESTING PROPERTIES

Molecular magnets may potentially offer other, and different, applications, which would make use of their transparency, low density, and low conductivity. Small size and controllable structural perfection of "functional units" open the door to observe manifestations of quantum effects—so far demonstrated at extremely low temperatures only. The studies now underway serve two major goals: better basic understanding of these quantum effects, and learning how to manipulate microscopic properties of molecular entities with magnetic field, temperature, and light. In particular, the quantum phenomena resulting from the interplay between light and magnetism are responsible for interesting behavior in molecular magnets, which may be not less fascinating than the long-known Faraday and Kerr effects.

In this section, we outline several selected properties specific to SMM systems that make research in this field so interesting. In particular, the possibility of studying basic quantum effects on macroscopic samples is very exciting.

3.1. Quantum Tunneling of the Magnetization

The magnetization in Mn_{12}-ac at sufficiently low temperatures, below the so-called blocking temperature, exhibits a magnetic hysteresis—a property well known for bulk magnets. This effect was first reported by Sessoli et al. [42]. Figure 5 reproduces experimental data from a later work by Thomas et al. [13]. The magnetic field was applied along the easy magnetization axis. The precision in the sample orientation was checked by comparing the low-temperature remanent magnetization with the saturation magnetization, in which the former reached more than 99% of the latter. Before measurements, the sample was cooled from above the blocking temperature with an applied field of 5 T, so that all the magnetic moments were well aligned along the direction of the external field, reaching the saturation magnetization M_s. The field was then reduced in steps to the next measurement point over a time of 10 s. Then, after a break of 400 s, which was included to allow for stabilization, the magnetization was measured. The magnetization decreased in a series of steps. At lower temperatures, clear plateaus between the steps were visible, producing a "staircase" hysteresis loop. The relaxation times in the flat parts were much larger than the time window of about 600 s of the measurement.

What may be not immediately clear is that this hysteresis is not caused by long-range interactions and expansion of domains, as in classical magnets, but is of purely intramolecular

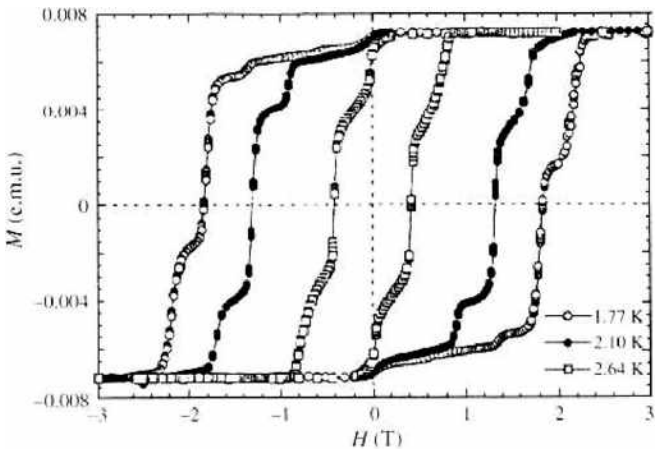


Figure 5. Magnetic hysteresis loop of a Mn_{12}-ac single crystal at different temperatures below the blocking temperature of about 3 K measured with a SQUID magnetometer. Reproduced with permission from [13], L. Thomas et al., *Nature* 383, 145 (1996). © 1996, Nature Publishing Group.

origin. This has been demonstrated by NMR measurements of single-molecule magnets in solution without long-range order on several derivatives of the material [32].

Because of their similarity to bulk magnets, these materials are often called molecular nanomagnets. It should be kept in mind, however, that only below the blocking temperature, which is about 3 K for Mn_{12}-ac, do these materials behave like magnets. Above this temperature, the magnetic moment fluctuates rapidly—an effect that is known as superparamagnetism.

Magnetic hysteresis appears if there is a barrier that the magnetization has to overcome. In a SMM, this role is played by a magnetic anisotropy barrier. The magnetic anisotropy energy (MAE) is defined as the difference between the ground state energies resulting from the rotation of the direction of magnetization. This barrier has its origin primarily in the spin-orbit coupling [8], which is a mechanism that explicitly couples the spin degrees of freedom with the spatial coordinate system defined by the molecular orbitals.

Strong spin-orbit coupling would try to align the spins of electrons in the magnetic d-shell with the magnetic field. In a chemical environment, the d-shell experiences a crystal field splitting because of the electric field of the neighboring atoms. The energy cost or gain of aligning the electron spins in an external magnetic field will then depend on the spatial arrangement of the d shell. In two special cases, however, the magnetic anisotropy can be suppressed. If the crystal field splitting (i.e., anisotropic interaction with the ligands) is small, the spins in the d shell may freely rotate and will follow the external magnetic field without much loss in energy. However, if the crystal field splitting is much larger than the spin-orbit coupling, the system (TM ion + ligands) prefers a configuration with zero (i.e., quenched) orbital moment and no magnetic anisotropy. In a semiclassical picture, this orbital quenching can be interpreted as the orbital momentum precessing in the crystal field, so that its magnitude remains unchanged but all its components average to zero. However, the spin-orbit coupling can not be completely ignored and will mix in states with nonzero orbital momentum. This is responsible for the deviation of the g-factor from the pure spin value of 2.

Therefore, the height of the MAE barrier emerges by a quite subtle interplay between the strength of the spin-orbit coupling and the interaction with the crystal field resulting from the ligands. The temperature up to which a magnetic device will retain its preferential magnetic orientation is therefore determined by the MAE; hence, high MAE is one of the major priorities in the design of useful SMM devices. Theoretical details concerning the calculation of the MAE within the DFT are discussed in Section 4.4, and numerical values for a number of SMM are presented in Section 6.3.

Another feature that makes the hysteresis curves in molecular magnets peculiar, other than the blocking temperature, is the presence of steps. This is essentially a quantum effect, a manifestation of QTM (see Thomas et al. [13], Wernsdorfer et al. [36], Wernsdorfer and Sessoli [37], Gunther and Barbara [43], Friedman et al. [44], and Barra et al. [45]). The observed steps correspond to sudden changes of magnetization as a function of magnetic field. QTM was theoretically discussed already by Gunther [46]. The observed transition from classical to quantum behavior is fascinating for fundamental research, but moreover, it is a challenge for applied nanotechnology, which tries to operate in that regime and to make use of these effects. Earlier experimental attempts to detect QTM were limited by difficulties in preparing sets of identical nanometer-sized magnets, because the tunneling amplitude varies exponentially with size, so that a large size distribution makes it impossible to detect the discrete character of quantum effects. A great advantage of SMMs like Mn_{12}-ac is the possibility of their growing single crystals of identical magnetic clusters, which are perfectly suited for measurements using highly sensitive magnetometers like superconducting quantum interference devices (SQUID).

The origin of steps in the hysteresis curve (Fig. 5) can be further clarified from Fig. 6, in which the relaxation time of the magnetization is plotted as a sequence of constant external magnetic field values. In each point of the hysteresis curve, the relaxation time τ has been obtained by repeated measurements of the magnetization over a timescale from 300 to 600 s, with subsequent fitting of the time dependence of the magnetization to a linearized exponential form: $M(t) = M_0(1 - t/\tau) - M_s$. The relaxation times τ have been carefully

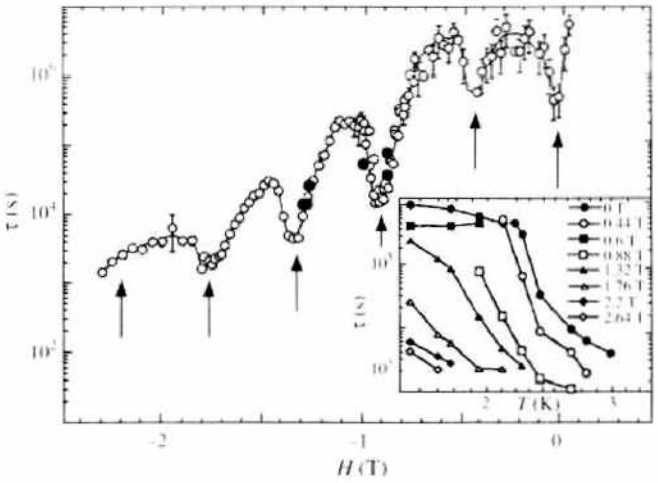


Figure 6. The magnetic relaxation time of Mn_{12} at temperature 2.1 K as a function of magnetic field obtained from repeated measurements for a given magnetic field and temperature on the hysteresis loop. The arrows show the fields at which jumps of the magnetization are expected as a result of the alignment of levels as explained in the text. Reproduced with permission from [13], L. Thomas et al., *Nature* 383, 145 (1996). © 1996, Nature Publishing Group.

verified by measurements over a longer timescale, 20–50 h, at the various temperatures and field strengths from which the filled symbols have been obtained. For certain magnitudes of the magnetic field, which are nearly equidistantly spaced, the relaxation of the magnetization becomes much faster. These are the magnetic field values at which the steps in the hysteresis curve are observed (i.e., at which sudden changes of the magnetization take place).

The simplest model to describe the observed magnetic behavior treats the resulting spin $S = 10$ of Mn_{12}-ac like a giant spin of a single atom, neglecting its composition from the spins of individual Mn ions. In a magnetic field, the $2S + 1 = 21$ degenerate m-levels would split. (A spin flip from $m = -10$ to the $m = 10$ state would microscopically mean a concerted spin flip on all magnetic ions in the Mn_{12}-ac molecule.) Using the model of a giant spin for the SMM, the simplest Hamiltonian including the magnetic anisotropy and a Zeeman term is

$$H = DS_z^2 + E\left(S_x^2 - S_y^2\right) + g\mu_B \mathbf{BS} \tag{1}$$

Differing from the later-given, more general form of Eq. (7), we introduce here two anisotropy constants, D and E. In uniaxial systems like the Mn_{12}-ac, $E = 0$ and the magnetic behavior is determined by the constant D. The term D has to be negative in a SMM, so that the magnetization will align along the easy axis. Without a magnetic field, the eigenvalues of Eq. (1) are $E_m = Dm^2$. As $D < 0$, the ground-state energy is degenerate and corresponds to $m = \pm S$, meaning that the magnetization will point along the easy axis up or down. Applying a magnetic field removes the degeneracy and shifts the energy levels. If the magnetic field $\mathbf{B}$ is aligned along the easy axis of SMM, and then the energy levels are shifted by $g\mu_B mB$ (i.e., some shift up and others down, depending on the sign of m). For certain values of B, the energies of two levels with different m may accidentally match. This is called a level crossing. It is easy to see that such level crossings are equally spaced in B and happen by steps $B_{\text{cross}} = D/g\mu_B$ (= 0.48 T for Mn_{12}-ac). Such equal spacing is indeed very close to what is experimentally observed; small deviations are likely to be caused by higher-order terms in the anisotropy, like the fourth-order terms in the spin operator S_z. The jumps in the relaxation times can now be interpreted as being caused by a largely enhanced transition rate between two levels at a level crossing, which is also accompanied by a jump in the hysteresis curve.

However, this picture cannot account for the full truth, because a prerequisite of tunneling is the mixing of states with different m. Without such mixing, all m-states are eigenstates of S_z and are hence orthogonal. Because of the tetragonal crystal symmetry, the lowest-order term in the spin operator that is allowed to mix will be proportional to $\sim S_x^4$ and $\sim S_y^4$.

This would couple only states $|m\rangle$ and $|m \pm 4\rangle$, in contrast to the experiment in which nearly all $\Delta m = \pm 1$ have been observed [13].

The tunneling in Mn_{12}-ac is believed to be a thermally assisted resonant tunneling [14, 47–50], although this determination was also debated [51]. The tunneling rate has a maximum for the $m = \pm 3$ levels without a magnetic field and is experimentally observable only for level crossings close to the top of the barrier. This indicates that thermal activation caused by phonons plays a role. The spin is climbing the Dm^2 ladder with the help of phonons, and close to the top it can take a shortcut by tunneling, allowing an enhanced relaxation of the magnetization. This general picture has been confirmed by EPR measurements, which detect transitions between the spin levels with high accuracy [28, 52–57] and high-frequency magnetic spectroscopy [58].

Applying a magnetic field along the hard axis (for systems with $E \neq 0$) will change the tunneling rates. This causes another interesting effect: the oscillations of the tunnel splitting as a function of the magnetic field along the hard axis. The period of oscillation has been given by Garg [38] as

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

The tunnel splittings for a model system with total $S = 10$ and magnetic anisotropy parameters $D = -0.53$ K and $E = 0.054$ K for three transitions $N \leftrightarrow -N$ as a function of magnetic field along the hard axis are displayed in Figure 7. One can see in this figure a clear period of about 0.4 T. It is noteworthy that fourth- and higher-order contributions to the spin Hamiltonian become very important, so that the tunnel splitting cannot be described by the simple formula in Eq. (2) anymore. This may have possible implications for applications in nanotechnology, because it allows us to control the tunneling process (relaxation of magnetization) by applying a magnetic field along the hard axis. This might be crucial, for instance, magnetic memory devices based on molecular nanomagnets.

The coupling of the magnetic cluster to the environment is a vast playground for theories explaining the mixing of the spin states via magnetic noise, incoherent phonons [59], dipolar [60–63] or hyperfine couplings to nuclear spins [64–68], dislocations in single crystals [69, 70], or coupling between the magnetic clusters (spin–spin relaxation, Wernsdorfer et al. [71, 72]). In the case of Mn_{12}-ac, angle-dependent single-crystal EPR data [73] gave clear evidence of

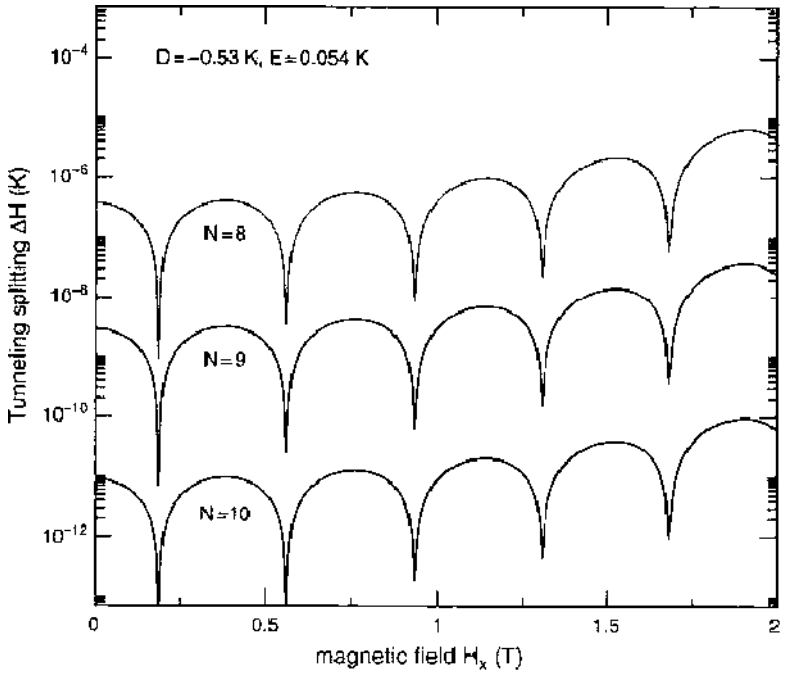


Figure 7. Tunnel splittings for three different transitions $M_s = -N \rightarrow N$ for a model system with $S = 10$ and the D and E as indicated in the figure. The magnetic field H_x is applied along the hard axis of the system.

a disorder resulting from crystallization of acetic acid [74], which induced a locally varying quadratic (rhombic) transverse anisotropy $E(S_x^2 - S_y^2)$.

Magnetic molecular clusters such as Mn_{12}-ac [26, 42, 75] or Fe_8 [34, 45] are probably the most widely studied materials for which QTM has been observed. The clarification of these fundamental details of the tunneling mechanism belongs to the hottest research topics in the field of molecular magnetism. A recent review on this broad area of research was given by Gatteschi and Sessoli [15].

Quite recently it has been theoretically shown that crystals of molecular nanomagnets can exhibit giant magnetic relaxation because of the Dicke super-radiance of electromagnetic waves [76]. Experimentally, it has been found that crystals of molecular nanomagnets exhibit enhanced magnetic relaxation when placed inside a resonant cavity. A strong dependence of the magnetization curve on the geometry of the cavity has been observed, providing indirect evidence of the coherent microwave radiation by the crystals [77].

3.2. Antiferromagnetic Rings

The molecular nanomagnets Mn_{12}-ac or Fe_8 show thermally activated QTM, as discussed before. This tunneling process is an incoherent tunneling of the magnetization. In contrast, antiferromagnetic molecular magnets such as ferric wheels, in which an even number of ions is coupled antiferromagnetically in a ring structure [78, 79], belong to the most promising systems for the observation of coherent quantum tunneling.

In coherent QTM, the spins tunnel back and forth between energetically degenerate states. The detection of the coherent tunneling will be challenging but possible with current experimental means.

An AFM coupling among nearest neighbors in an even-member ring with nearest-neighbor interaction results in an antiferromagnetic ground state ($S = 0$). Obviously there are two degenerate classical spin configurations for such a ring, related by an inversion of spins on all atoms, as schematically shown in Figure 8. One example of a hexa-nuclear ferric wheel is discussed in detail in Section 6.1.

Two degenerate ground states are, however, separated by an energy barrier because of the single-ion anisotropy. The coherent tunneling is a simultaneous tunneling of all N spins through this anisotropy barrier, back and forth. This process is also called tunneling of the Néel vector.

An additional interest in AFM ring structures comes from theoretical interest in finite Heisenberg systems, for which these wheels are experimental realizations. This allows us to test theoretical knowledge against experimental data. In particular, it is interesting to understand the low-lying excitations in these systems, which can be measured by torque magnetometry, for example [80].

The Heisenberg model has a long research history that is probably more focused on one-dimensional chains and on the thermodynamic limit with the number of centers going to infinity [81]. The effort of exact numerical diagonalization rapidly becomes prohibitively large with the increase of S and the number of centers. The molecular wheels, because of the

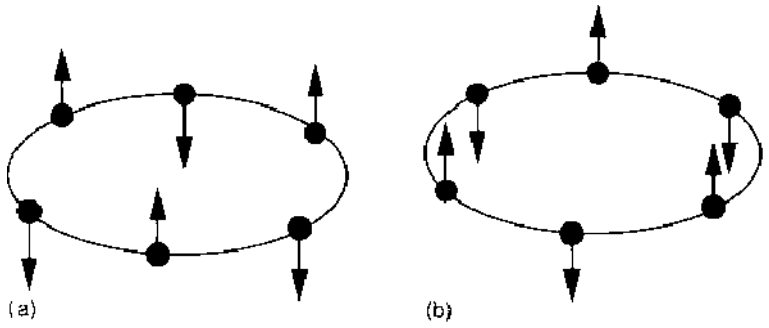


Figure 8. Schematic coupling of the magnetic centers in antiferromagnetic ring systems. The ground state is $S = 0$ with two degenerate classical spin configurations as shown in (a) and (b).

limited number of spins and magnetic couplings, often allow for an exact numerical solution. The expectation values of physical observables can then be calculated from the spectrum of eigenvalues as indicated, for example, in Eqs. (9) and (11). This area of research evolves rapidly, to a large extent because of progress in our computational means, which allow for the exact numerical solution, but also because of progress in the theory for finding exact bounds for the solutions.

The effect of a small anisotropy, as already included in Eq. (7), became crucial for the qualitative and quantitative understanding of the experimental data obtained from the antiferromagnetic wheel structures. Schnack and Luban [82] presented numerical evidence for the existence of a rotational band in systems of quantum spins interacting with nearest-neighbor antiferromagnetic Heisenberg exchange. In contrast to earlier results, which noted this behavior for rings with an even number of spin sites, the authors find that it also applies to rings with an odd number of sites, as well as for a number of other configurations (tetrahedron, cube, octahedron, icosahedron, triangular prism, and axially truncated icosahedron). Further, they demonstrated how to predict in many cases the rotational band levels, using the underlying sublattice structure of the spin array. The characteristics of the rotational band can provide estimates for the low-temperature magnetic susceptibility.

Waldmann [83] numerically investigated the spin pair-correlation function of finite bipartite antiferromagnetic Heisenberg rings by means of exact diagonalization techniques. The spin pair correlation function consists of few characteristic peaks at low temperatures and contains a broad featureless signal at high temperatures for large spins. This arises as the energy spectrum exhibits a set of parallel rotational bands at low energies emerging into a quasi-continuum of states, with transitions from the lowest rotational band to the quasi-continuum being highly suppressed. The energies of the rotational bands can be accurately described by a generalized dispersion relation that depends on the total spin quantum number and on the shift quantum number. These regularities become more pronounced for larger spins and smaller ring sizes. All these features are associated with the underlying sublattice structure of the spin rings, and it is argued that they are valid for a more general class of finite Heisenberg systems than rings.

Meier and Loss [84] theoretically studied the thermodynamic properties and spin dynamics of modified AFM rings, in which one of the Fe (III) ions has been replaced by a dopant ion to create an excess spin. Using a coherent-state spin path integral formalism, they derived an effective action for the system in the presence of a magnetic field. Further, they calculated the functional dependence of the magnetization and tunnel splitting on the magnetic field and showed that the parameters of the spin Hamiltonian can be inferred from the magnetization curve. From the spin dynamics in these systems, the authors infer that quantum tunneling of the Néel vector also results in tunneling of the total magnetization. Hence, the spin correlation function shows a signature of Néel vector tunneling, and electron spin resonance techniques or ac susceptibility measurements should be able to measure both the tunneling and the decoherence rate.

The molecular six-membered ferric wheel discussed below (Section 6.1) shows an interesting cooling behavior under the influence of a magnetic field. The process of adiabatic cooling, which means cooling by demagnetization, is well known. Waldmann et al. [85] have demonstrated the opposite effect—cooling the molecules by increasing the strength of the applied field—the possibility of which has been discussed earlier for other materials [86–88].

3.3. Quantum Computing with Molecular Magnets

This will not be a review on quantum computing, and we refer the interested reader to numerous accounts at http://arxiv.org, to review papers on that topic [89] or to recent books [90, 91] for more information.

Our primary interest is that SMM systems are often mentioned as possible brick stones for the realization of quantum computers; therefore, we will give a very short introduction to explain the underlying physics of this suggestion. In this, we will mostly follow the excellent review by Steane [89].

The broad field of quantum informatics merges ideas from classical information theory, computer science, and quantum physics. The idea of information as a basic concept in physics

is used here and is supported by a strict mathematical treatment of information and information processing. Among several interesting directions of research in this field, we would like to mention:

- quantum states used for secure transmission of classical information (quantum cryptography)
- quantum entanglement used for reliable transmission of quantum states (teleportation)
- preservation of quantum coherence in the presence of irreversible noise (quantum error correction)
- use of the inherent parallelism in quantum entanglement as a computational resource (quantum computing).

The elementary unit of quantum information is the *qubit*, which can be represented in form of a two-state system (e.g., a spin 1/2, or any 2-level system). A quantum system contains n qubits if its Hilbert space has 2^n dimensions so that one has 2^n mutually orthogonal quantum states available. If one writes two orthogonal states of a qubit as $\{|0\rangle, |1\rangle\}$, then two qubits result in four states $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$. Simple unitary operation on qubits correspond to quantum logical gates. The NOT operation would simply be expressed as

$$X \equiv |0\rangle\langle 1| + |1\rangle\langle 0|$$

Other logical operations can be realized in a similar manner, although their realization may require further qubits. For example, the AND operation can be realized using three qubits. Combinations of such gates can be used to perform elementary arithmetical operations such as binary addition and multiplication.

So far we have not demonstrated anything special that would make the construction of a quantum computer interesting. Obviously, a classical computer can calculate the behavior of a quantum system, and as long as we can write down our theories in the form of equations that can be coded on a computer, there are no problems a classical computer cannot solve—in principle. Most likely, quantum computers will not extend the set of problems that are solvable in principle, although there may be tasks for which classical computers are too slow, but that may become solvable by quantum computers.

The most straightforward application of a quantum computer is a simulation of some other quantum system. To store a quantum mechanical state in a 2^n-dimensional Hilbert space, a classical computer needs 2^n complex numbers, whereas a quantum computer requires only n qubits. To be fair, it is worth noting that both types of computers will be inefficient in simulating the evolution of the quantum state. For the matrix vector operations, a classical computer must manipulate matrices of the order $2^n \times 2^n$, involving many multiplications and additions, which requires a number of operations exponentially large in n. The quantum computer must build unitary transformations in the 2^n-dimensional Hilbert space requiring an exponentially large number of elementary quantum logic gates. There is no guarantee that a quantum computer is able to simulate every physical system efficiently. However, there is a large class of systems for which no efficient classical algorithms exist (e.g., many-body systems with local interactions) but that can be simulated by a quantum computer.

One of the problems for which a quantum computer can hopefully show its full strength is the search of the period τ of a function $f(x) = f(x + \tau)$. This problem is closely related to prime factorization, which is very important in cryptography. Assume that there is no analytic possibility to deduce the period and that all we know is that $N/2 < \tau < N$ for some N. Using a classical computer, we have to calculate $f(x)$ for about $N/2$ values of x to find when the function repeats itself. The task can be solved efficiently on a quantum computer by a method coming from Shor [92]. The two main components of the algorithm are modular exponentiation (computation of $a^x \bmod N$) and the inverse quantum Fourier transform. The advantage of the algorithm comes from calculating the value of $f(x)$ for 2^n values of x in parallel (imagine having 2^n classical computers in parallel). For details of the algorithm see Steane [89] and Shor [92].

To build a quantum computer that can be programmed universally, like today's desktop computers, is well beyond the abilities of current technology. However, the development

of algorithms for quantum computation is already progressing, and several principles have already been demonstrated to work. As an example, a "quantum computer" realized on seven spin-1/2 nuclei in a molecule, hence 7 qubits, and using nuclear magnetic resonance techniques to manipulate them successfully factorized the number 15 using Shor's algorithm [93].

Another problem of high practical interest involves the search of an unstructured list for an item (e.g., finding a name corresponding to a known telephone number from a telephone directory). A classical algorithm can only make a plain search through the list. For a list of N items, it will require an average of $N/2$ steps to find the item. The task will still be computationally demanding on a quantum computer, but Grover [94] presented an algorithm that requires only $\sqrt{N}$ steps. Bennett et al. [95] proved an interesting point: that Grover's algorithm is optimal, and no quantum algorithm can do better. In the following text, we will try to give a brief outline of Grover's algorithm because Leuenberger and Loss [16] proposed an implementation using molecular magnets.

Assume that each item of the list has a label i and that we need a unitary operator U_j that is able to test whether a given item i is equal to the entry j we are looking for. This operator could have the form $U_j|i\rangle = |i\rangle$ for $i \neq j$, or $U_j|j\rangle = -|j\rangle$ for the special entry j. The method begins by preparing an equally weighted superposition of all states, which means that each state has the same amplitude $(1/\sqrt{N})$. Next, one has to execute $\sim \sqrt{N}$ times the following loop: operate with U_j on the state, and then apply Grover's diffusion operator.

The diffusion operator corresponds to a reflection of all the amplitudes about their mean. As a result, the amplitude of the state in question increases (decreases) by the same amount it was below (or above) the average over all amplitudes. If the number of states N is large, the average amplitude is close to $1/\sqrt{N}$, so that only the state j that had negative amplitude will change. Because the sign of the amplitude we want has been reversed, this operation enhances this amplitude at the expense of the others. In Figure 9, we illustrate how this works for the case of $N = 6$. If one finally carries out a classical measurement, one will measure with high probability the state $|j\rangle$, which has the largest probability amplitude.

Grover [96] then showed in a following paper that the algorithm can be used to search an arbitrarily large database by a single query if it were possible to query the quantum computer about multiple items. As shown by Leuenberger and Loss [16], this can be implemented in terms of a unitary transformation applied to the single spin of a molecular magnet like Mn_{12}-ac or Fe_8-cluster. The required superposition of states could eventually be prepared through the use of multifrequency coherent magnetic radiation in the microwave and radiofrequency range. As shown theoretically, in principle it should be possible to coherently populate and manipulate many spin states simultaneously by using advanced electron spin resonance techniques. A single pulse of a magnetic field can then create a nonlinear response of the SMM through multiphoton absorption processes involving particular sequences of σ and π photons that allow the encoding and decoding of states. The final measurement of the decoded

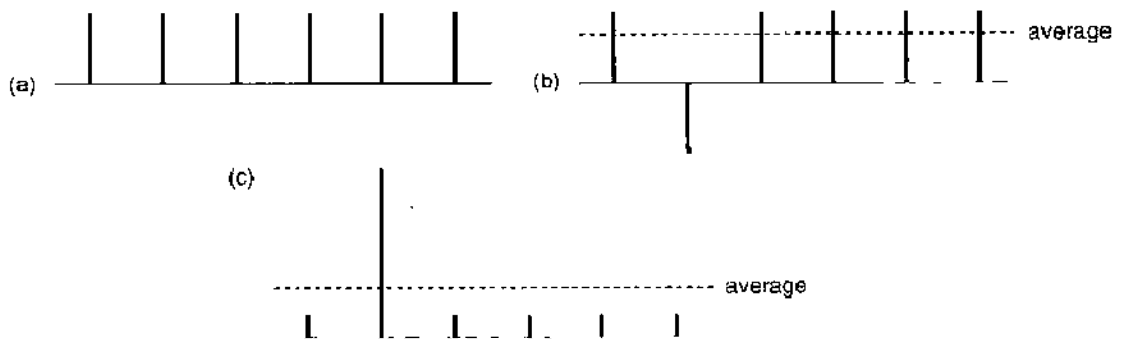


Figure 9. Schematic representation of the amplitudes and operators involved in Grover's algorithm for $N = 6$ bit. The second bit should be the special item looked for. The staring state (a) is a superposition of all states with equal amplitudes. The steps (b) and (c) will be iterated in a loop $\sqrt{N}$ times, so that the amplitude of that special state becomes large and a final classical measurement will deliver that one with high probability. This is achieved by changing the phase of the special state by π by an unitary operator in (b). The amplitude of that state is enhanced at cost of the other amplitudes by an inversion about average.

state by pulsed ESR techniques is based on the nonequidistance of the energy levels typical for molecular magnets.

4. MAGNETIC INTERACTIONS

A rather complicated and subtle subject that often appears in the discussion of experiments and theory, and not only of molecular magnets, is that of interatomic magnetic interactions and related exchange parameters. We begin by introducing a consistent conceptual framework for the following discussion, after which we overview some basic knowledge related to the physics of magnetic interactions and introduce parameterized spin Hamiltonian models. Later on, we discuss how the parameters of these models can be accessed from experiment and first-principles calculations. Section 6 will contain numerical estimations of model parameters for certain SMMs.

It should be noted that the only truly magnetic interaction is the magnetic dipolar interaction (e.g. between two remote spins), which is mediated by the magnetic field and is effective also in empty space; strictly speaking, higher-order magnetic interactions like quadrupole interaction are possible, but because of their weakness and short-range character, they will be neglected in the following discussion. The dipolar interaction is weak and decreases as the inverse of the cube of the distance. In chemical systems, it may play a noticeable role in special cases; for instance, in governing the arrangement of spins of different magnetic molecules, which are well separated in a molecular crystal. Much stronger and more important are interactions through chemical bonds, which are, however, electrostatic in their nature. The mutual arrangement of spins occurs as a result of an interplay between Coulomb interaction and Pauli exclusion principle, in what is referred to as exchange interaction. It is crucial for understanding the short-range (and also long-range) magnetic order.

The attraction by atomic nuclei and repulsion by other electrons governs the arrangement of electronic states in molecular orbitals, whereby the Pauli principle demands that only two electrons with opposite spin may share the same orbital. The simple case of two electrons helps us to understand why this interaction is called exchange. It also demonstrates that quantum mechanics is the foundation of magnetic ordering in solids. Given two noninteracting electrons at $\mathbf{r}_1$ and $\mathbf{r}_2$ with their respective wave functions $\varphi_1(\mathbf{r}_1)$ and $\varphi_2(\mathbf{r}_2)$, we can try to construct the two-electron wave function from products of the latter. As the electrons are indistinguishable, the probability density should remain unaffected if we exchange them. This leaves us, as allowed possibilities for the normalized two-electron wave function, the symmetric (S) and the antisymmetric (AS) combinations of the products of two single-electron wave functions

$$\psi_S(\mathbf{r}_1, \mathbf{r}_2) = \frac{1}{\sqrt{2}}[\varphi_1(\mathbf{r}_1)\varphi_2(\mathbf{r}_2) + \varphi_1(\mathbf{r}_2)\varphi_2(\mathbf{r}_1)]$$

$$\psi_{AS}(\mathbf{r}_1, \mathbf{r}_2) = \frac{1}{\sqrt{2}}[\varphi_1(\mathbf{r}_1)\varphi_2(\mathbf{r}_2) - \varphi_1(\mathbf{r}_2)\varphi_2(\mathbf{r}_1)]$$

So far we have considered only the spatial component of the wave function, but the electrons are also characterized by a spin. As the complete wave function for electrons $\psi(\mathbf{r}_1, \sigma_1; \mathbf{r}_2, \sigma_2)$, including spatial and spin components, must be antisymmetric, we have two possibilities: an antisymmetric spin singlet state S ($S = 0$) together with a spatial symmetric state, or symmetric spin triplet T ($S = 1$) with an antisymmetric spatial part

$$\psi(\mathbf{r}_1, \sigma_1; \mathbf{r}_2, \sigma_2) = \begin{cases} \chi_S(\sigma_1, \sigma_2)\psi_S(\mathbf{r}_1, \mathbf{r}_2) \\ \chi_T(\sigma_1, \sigma_2)\psi_{AS}(\mathbf{r}_1, \mathbf{r}_2) \end{cases} \tag{3}$$

The energy difference between the singlet E_S and the triplet E_T states allows to define the exchange constant J between the spins

$$J = \frac{E_S - E_T}{2} \tag{4}$$

A positive value of $J(E_S > E_T)$ favors the triplet state with $S = 1$ and a negative J—the singlet one.

Where does the difference between the energies of singlet and triplet states come from? Obviously it is determined only by the spatial shape of ψ_S and ψ_{AS}. Some general ideas about the sign of J, with the use of kinetic energy and Coulomb repulsion arguments, have been given by Anderson [97]. Let us assume that there are two electrons on the same atom. The spatially antisymmetric wave function minimizes the Coulomb repulsion, because the two electrons are spatially separated. Hence the spin triplet state is lower in energy, J is positive, and the resulting ferromagnetic-like interaction between the spins is consistent with Hund's first rule. Another enlightening example is represented by two electrons on neighboring atoms, so that they can form bonds. The corresponding molecular orbitals can be spatially symmetric (bonding) or antisymmetric (antibonding), as outlined above. The antibonding orbital has larger kinetic energy (larger curvature, on the average), which implies that it is energetically more expensive. This favors the spin-antisymmetric singlet state with the spatially symmetric bonding molecular orbital, and hence the exchange constant is negative, resulting in an antiferromagnetic-like interaction.

Although this concept is straightforward for two electrons, the extension to many electrons becomes complicated. Still, it provides a useful starting point for further qualitative discussion of exchange in many-electron systems.

As a further refinement, when discussing exchange interactions through chemical bonds, one has to distinguish between direct exchange (resulting from an immediate overlap of atomic states of magnetic atoms) or indirect exchange, occurring via an intermediate atom or group of atoms (diamagnetic groups), which form an exchange path of coupled chemical bonds. The direct exchange is typical for transition metals (TMS) and is responsible for ferromagnetism in Fe, Co, and Ni, or for the more complex magnetic structures of Cr and Mn. The indirect exchange again splits into several variations, but the variety of its possible chemical realizations might make classification by chemical bonds ambiguous. Nevertheless, a number of fundamental works coined the definition of several types of indirect exchange, which are by now well established and routinely referred to.

Zener [98] introduced the concept of double exchange, having in mind particularly perovskite-type manganites with mixed valence of Mn, $(La_{1-x}A_x)MnO_3$, with A = Ca, Sr, or Ba. These compounds are both conducting and ferromagnetic at intermediate doping $0.2 < x < 0.4$, being neither at $x = 0$ and 1. The double exchange enforces FM (ferromagnetic) orientation of spins at two TM ions, having different nominal valencies (Mn^{3+} and Mn^{4+} in Zener's original work). The exchange is realized by borrowing and refunding an electron from or to the closed O^{2-} shell of an intermediate ligand. The back-and-forth process flips the charge states of the involved Mn ions and accounts for the observed conductivity, which is blocked if all Mn valencies become uniform. The physics of double exchange were further elaborated by Anderson and Hasegawa [99], who proposed the appropriate (semiclassical) model Hamiltonian. The studies on manganites were essential to put foundations of the Goodenough–Kanamori rules [5–7], at which we will immediately arrive. The interplay between magnetic order and conductivity in manganites stimulated persistent interest in these systems up to this time, which lead to the discovery of "colossal magnetoresistence" and a remarkable number of publications see, e.g., Coey et al. [100] for a review.

Anderson [101] qualitatively solved the problem of indirect exchange, or "superexchange," in insulating TM oxides, showing how the "kinetic" exchange resulting from the virtual transfer of electrons may become dominant in magnetic insulators and how it leads, in most cases, to a net AFM coupling. The starting point of Anderson's analysis is the introduction of localized magnetic orbitals, related to a certain TM ion but including also some electron states of a diamagnetic ligand. Different magnetic orbitals may therefore share a common ligand and either experience an overlap there or be orthogonal. The simple argument of Anderson is "that antiparallel electrons can gain energy by spreading into nonorthogonal overlapping orbitals, where parallel electrons cannot" (Ref. [101], p. 2). This is generalized in the form of the Goodenough–Kanamori (–Anderson) rules, stating that if two electrons are in orbitals that directly overlap, their exchange (i.e., the 180 degree exchange)

is strong and of an AFM type, whereas for the (interacting) orbitals that are orthogonal, the (90 degree) if exchange is weak and of the FM type. The real world sometimes deviates from these rules. Weihe and Güdel [102] critically assessed the order-of-magnitude estimations of Anderson [101] and demonstrated that in some materials, the FM contribution in the kinetic exchange may dominate the AFM one. In general, however, one must note that in cases in which magnetic interactions are mediated by diamagnetic ligands, as is exactly the case in typical SMMs, the AFM arrangement seems to be the rule and the FM one a relatively rare exception. This is why Kollmar and Kahn [103] considered molecular ferromagnetism as a challenge and outlined several strategies for enforcing the dominance of the FM coupling in special chemical environments.

The formalization of the above-cited physical mechanisms, and hence the estimates of whether FM or AFM coupling ultimately wins in different systems, proceeds in terms of basic parameters like the Coulomb integral, the overlap of different orbitals, and transfer probabilities. Their order of magnitude can be reasonably guessed, but these parameters are difficult to evaluate precisely because of ambiguities in their definition—which wave functions to use, how to account for screening effects, and so forth. The direct numerical evaluation of such "internal" parameters of models proved to be relatively worthless, because subsequent predictions of actual exchange couplings (e.g., energy differences between FM and AFM configurations) come out usually far from reality. More promising is the strategy that focuses not on the "internal" parameters of one model or another but on those models that allow a "low-level" parameterization of the interaction energy in terms of few basic observables. A good example is the Heisenberg model, which casts the interaction of two (presumably well-defined) spins into a simple analytical form, incorporating all underlying physics in a single isotropic interaction parameter. An advantage of introducing such parameters (of interaction between nominal spins) is that they are better accessible in experiment; for example, the Heisenberg exchange parameter straightforwardly recovers the coefficients in the Curie–Weiss behavior of magnetic susceptibility. Therefore, a reasonable mutual verification of calculations (at different levels of microscopic complexity) and experiment becomes possible. The disadvantage is that the physics remains somehow hidden. This disadvantage is not so strong with respect to the *ab initio* calculations, because the microscopic interactions leading to the final result are, of course, accounted for in the calculation, and they therefore can be analyzed in detail. For example, the issues of hybridization, localization, and charge transfer are typical ingredients of a quantum chemical analysis, and normally they do give insight into the origin of a particular magnetic ordering. If we understand them, we may hope to find possibilities to "design," or at least to influence, the magnetic properties.

For the subsequent quantitative discussion of magnetic interactions, we must therefore always refer to a particular physical model that, in general, does not explain but, rather, describes mathematical relations between observables. It is convenient to keep such models simple and the number of required parameters small. However, the deviations of "reality" (either experiments or first-principles calculations) from the predictions of a simple model may demand the introduction of sophistications that bring in additional parameters. The values of the model parameters can be extracted from both experiment and *ab initio* calculation. Observables like temperature or field dependency of magnetization or magnetic susceptibility may be fitted more or less satisfactorily to the predictions of a certain model, yielding the values of interaction parameters (in the sense of this particular model only). The first-principles calculations provide the spectrum of eigenvalues or compare total energy in different magnetic configurations and derive estimates of interaction parameters from fitting these data, again, to a particular physical model. Therefore, one should be careful in comparing "measured" and "calculated" interaction parameters: They are accessed indirectly and from different starting points; hence, their agreement may be accidental and the source of disagreement not immediately obvious.

4.1. Spin Hamiltonians

The model Spin Hamiltonians neglect the true chemical environment and bonding and reduce all interactions to just a few model parameters. One is dependent on experimental

input to estimate these parameters, and the accuracy of quantitative predictions depends on the parameters chosen. Even more problematic, it is not *a priori* clear which interactions are important and should be included in the model and which are negligible. Because the interactions (e.g., in the Heisenberg model) are reduced to effective interaction between spins only, all microscopic information about different exchange paths is lost, which is a limitation if one wants to optimize magnetic properties by changes of structure or bonding.

Even considering these limitations, the theoretical techniques based on Heisenberg or Ising Spin Hamiltonions belong to the most powerful approaches to simulate and understand the magnetic behavior of materials. The parameterization of a magnetic interaction normally includes, as the presumably leading term, the Heisenberg Hamiltonian

$$\mathcal{H} = -2\sum_{i>j} J_{ij}\mathbf{S}_i\mathbf{S}_j \tag{5}$$

with the summation indicating that each pair of spins $\mathbf{S}_i$, $\mathbf{S}_j$ is counted only once (though the definition of sign and prefactor may vary between publications). As only the relative orientation of both spins matters, this interaction is isotropic.

The dependence on absolute spin orientation (i.e., with respect to the crystal lattice) can be brought in via a modification of the Heisenberg model, taking anisotropy into account:

$$\mathcal{H} = -2\sum_{i>j} J_{ij}[\mathbf{S}_i^x\mathbf{S}_j^x + \gamma(\mathbf{S}_i^y\mathbf{S}_j^y + \mathbf{S}_i^z\mathbf{S}_j^z)] \tag{6}$$

This form of interaction recovers the conventional Heisenberg model in the case of $\gamma = 1$ and reduces to the Ising model for $\gamma = 0$, or to the two-dimensional interaction for $\gamma \gg 1$. Surprisingly, even in the case of the seemingly simple Ising model, so far analytic solutions are known only for one-dimensional and two-dimensional lattices [104].

Further on, the single-spin anisotropy can be included and the Zeeman term added, yielding

$$\mathcal{H} = -2\sum_{i>j} J_{ij}(\mathbf{S}_i\mathbf{S}_j) + D\sum_i(\mathbf{e}_i\mathbf{S}_i)^2 + g\mu_B\sum_i \mathbf{B}\mathbf{S}_i \tag{7}$$

with the Landé g-factor being normally close to 2, and μ_B being the Bohr magneton. The single-spin anisotropy term may lack some of the true physics. It is scaled with its corresponding constant D and depends on the orientation of each spin $\mathbf{S}_i$ relative to a reasonably chosen fixed direction in space $\mathbf{e}_i$; the Zeeman term scales with the external magnetic field B for the chosen value of the g factor.

Such model spin Hamiltonian can be further sophisticated by introducing additional parameters; that is, distinguishing between random (varying from site to site) and constant (global) magnetic anisotropy, yielding the appearance of distinct D parameters in Eq. (7). Moreover, higher-order terms in isotropic interaction (biquadratic exchange, etc.), as well as from antisymmetric Dzyaloshinsky–Moriya spin exchange [105, 106],

$$\mathcal{H}_{DM} = \sum_{i>j} \mathbf{D}_{ij}\cdot[\mathbf{S}_i \times \mathbf{S}_j] \tag{8}$$

can be introduced. This might be necessary to grasp essential physics, but it makes the extraction of parameters, usually from a limited set of experimental data, more ambiguous, leading to a problem of overparameterization.

It should be noted that the definition of the Heisenberg Hamiltonian in different publications differs sometimes in the sign and in the presence of prefactor 2, which must be taken into account when comparing different sets of extracted parameters. The notation as above corresponds to $J > 0$ for the FM ferromagnetic coupling.

Although these Hamiltonians look very simple, they describe an enormously rich magnetic behavior, particularly in what regards phase transitions and excitations. Once the J_{ij} are known (or chosen), one needs "only" to diagonalize the Hamiltonian to get the partition function and to calculate all desired thermodynamic properties (see Section 4.2). The problem is often a technical one, because the dimension of the Heisenberg Hamiltonian

grows very fast with the number of involved spins, N, and their S value, namely $(2S+1)^N \times (2S+1)^N$. For example, Fe_8 (eight Fe ions with $S = 5/2$) gives $(2S+1)^8 = 1679616$, so the allocation of 4 bytes for each matrix element of the Hamiltonian would require 10509 Gbyte of memory just for the storage of the Hamilton matrix. However, most of the matrix elements would be zero and, making use of existing symmetries and special diagonalization techniques, such as the Lanczos algorithm, it could still be possible to deal with systems of this size. Clearly, one reaches very fast the limits of current (and eventually future) computer technology.

In the following text, we discuss the relations of model (spin) Hamiltonians to experiments and to first-principles calculations. The introduction of microscopic concepts, like orbitals, permit us to link the exchange constants with the real chemical structure and bonding. Depending on the character of the states involved in the exchange mechanism, one can distinguish between them and quantify the above-mentioned concepts like direct exchange, superexchange (indirect exchange) in insulators, itinerant exchange (RKKY interaction) in metals, and double exchange in some oxides or anisotropic (Dzyaloshinsky–Moriya) exchange. More details were given by Anderson [97] and Blundell [107].

The advantage of *ab initio* approaches in the extraction of interaction parameters is that certain mechanisms of interaction can be switched on and off in a fully controllable way. Thus, all anisotropy terms may only have an effect if the spin-orbit interaction is explicitly present in the calculation. The noncollinear orientation of individual spins can sometimes be arbitrarily chosen, or at least different settings of "up" and "down" configurations of spins with respect to a global quantization axis are available in a calculation scheme, so that angles between spins become, in one way or another, directly accessible.

The experiment does not allow such grade of control on the microscopic level: The magnetic field and temperature are eventually the only tunable parameters, and the availability of good-quality oriented monocrystalline samples is not a rule. Experimental microscopic techniques include only a few spectroscopic studies and (exclusively for Fe-based magnets) Mössbauer effect measurements, which are able, to some extent, to probe charge and spin state of an ion in question.

4.2. Relation to Experiment

Probably the most commonly available characteristics of magnetic materials are their bulk (molar) magnetization and magnetic susceptibility, as functions of the magnetic field and temperature. These properties can be directly derived from the spin Hamiltonian, so that the general applicability of the underlying model can be verified and the numerical values of the involved parameters can be tuned. Other, microscopic techniques like nuclear magnetic resonance, EPR, Mössbauer effect, or optical or X-ray spectroscopy allow to probe certain parameters of a model Hamiltonian selectively.

Once the spin Hamiltonian is agreed on, it can be, at least in principle, diagonalized, and its eigenvalues E_n can determine the partition function and all thermodynamic properties with their dependency on magnetic field B and temperature. Specifically, the molar magnetization is

$$M_{\mathrm{mol}} = -N_A \frac{\sum_n \partial E_n/\partial B \exp(-E_n/kT)}{\sum_n \exp(-E_n/kT)} \tag{9}$$

where N_A is the Avogadro number. The zero-field molar magnetic susceptibility, taking into account the dependence of the eigenvalues E_n on the homogeneous magnetic field B_z up to the second order

$$E_n = W_n^{(0)} + B_z W_n^{(1)} + B_z^2 W_n^{(2)} + \cdots \tag{10}$$

yields

$$\chi_{\mathrm{mol}} = \mu_0 N_A \frac{\sum_n \left[\left(W_n^{(1)}\right)^2/kT - 2W_n^{(2)}\right]\exp(-W_n^{(0)}/kT)}{\sum_n \exp(-W_n^{(0)}/kT)} \tag{11}$$

The evaluation of the values of the parameters in the spin Hamiltonian proceeds by fitting the calculated temperature (or magnetic field) dependencies to the measured data. The practical difficulty lies in the diagonalization of the Hamiltonian, whose dimension grows very rapidly with the number of spins and their S values.

A common conceptual difficulty is the necessity to choose between several sets of parameters that yield an equally reasonable fit. An example of such ambiguity is given by Katsnelson et al. [108] in fitting a model eight-spin Hamiltonian for Mn_{12}-ac to the neutron scattering data.

4.2.1. Mössbauer Spectroscopy

Mössbauer spectroscopy, based on the effect discovered by Mössbauer [109, 110], uses in most cases iron in its application and requires the enrichment of samples with the ^{57}Fe isotope. It probes the transition between the excited $I = 3/2$ and $I = 1/2$ states of the ^{57}Fe nuclei. The $I = 1/2$ ground state is reached by emission of a gamma quant with 14.4 keV. This gamma quant is able to excite transitions in the investigated sample so that the sample can be probed by its resonant absorption properties. Foundations of the Mössbauer spectroscopy are described, for example, in the textbooks by Greenwood and Gibb [111] and Vértes et al. [112]. A broad review of applications is given in Gonser [113], including a chapter on applications in chemistry that are closely related to our subject (see Ref. [114]).

Basic pieces of information from the Mössbauer spectra are the isomer shift, quadrupole splitting, and magnetic hyperfine splitting. The transition energy between nuclear energy levels of the sample (probed by a resonant absorption of γ-quanta) may differ from the reference one (recoil-free emission of the γ-quanta by the Mössbauer source) as a result of different electric fields at the probed nucleus and the emitting nucleus in the source. Such an isomer shift δ reflects the contact electron density (at the nucleus, because of s-electrons), in which the $4s$ contribution is screened by the more localized $3d$ states. This allows us to discriminate the charge/spin state of iron (with some ambiguities) by the value of the observed isomer shift

$$\delta = \frac{4\pi}{5} Ze^2 \left(\frac{\delta R}{R}\right) [|\psi_s(0)_{\text{source}}|^2 - |\psi_s(0)_{\text{sample}}|^2] \tag{12}$$

where $\delta R = R_{I=3/2} - R_{I=1/2}$ is the difference in the ^{57}Fe nuclear radii in the excited and the ground state, and R is their average value. Both constants in Eq. (12), $\delta R/R$ and $\psi_s(0)_{\text{source}}$, are fundamental properties of the Mössbauer source, albeit being dependent on temperature.

The quadrupole splitting E_Q arises because of the "nonsphericities" of both the nucleus and the ambient electric field. The $I = 3/2$ state possesses an electric quadrupole moment

$$eQ = \int d^3r\, \rho(\mathbf{r}) r^2 (3\cos^2\theta - 1) \tag{13}$$

where $\rho(\mathbf{r})$ is the spatial density distribution in the nucleus ($Q > 0$ corresponds to a prolate nucleus and $Q < 0$ to an oblate one), and hence the $I = 3/2$ energy level splits by an external electric quadrupole field into $m_I = \pm 3/2$ and $m_I = \pm 1/2$ levels, yielding a doublet in the detected spectrum. The effect of an external electric field E can be comprehended by its electric field gradient (EFG) tensor at the nucleus site

$$\nabla E_{\alpha\beta} = -V_{\alpha\beta} = -\frac{\partial^2 V}{\partial x_\alpha \partial x_\beta}$$

where x_α, x_β are Cartesian coordinates, which can be transformed to principal axes chosen such that $|V_{zz}| \geq |V_{yy}| \geq |V_{xx}|$. Moreover, as the Laplace equation requires this tensor to be traceless, it is fully specified by just two parameters, V_{zz} and $\eta = (V_{xx} - V_{yy})/V_{zz}$, $0 \leq \eta \leq 1$. The EFG tensor reflects the asymmetry of the near environment, because of the shell structure of the Fe atom, and to the ligand field. The Hamiltonian of the quadrupole interaction—see Vértes et al. [112] or Gütlich [114] for details—allows an exact solution

for $I = 3/2$, according to which this nuclear state splits by

$$E_Q = eQ\frac{V_{zz}}{2}\sqrt{1+\frac{\eta^2}{3}}$$

The EFG can be directly calculated by high-precision *ab initio* methods, as demonstrated by Jena [115] and then by Blaha et al. [116]. More details on practical calculation of EFG, and more results for different systems, were given by Petrilli et al. [117]. Such studies provide good benchmarks for testing the accuracy of calculations, but they also help to decipher experimental results, as the value of Q cannot be in practice calculated according to Eq. (13), and earlier literature (see, e.g., Ref. [111]) reflects a certain controversy on this subject. A first-principles calculation of EFG in a number of binary Fe compounds [118] proposed to set the hitherto disputable Q value for ^{57}Fe at 0.16 barn. Later, first-principles calculations using this number resulted in good agreement between calculated quadrupole splittings and measured values for Fe-based molecular magnets [119].

Finally, the degeneracy of both $I = 1/2$ and $I = 3/2$ states in m_I may be lifted in an external magnetic field B as a result of the nuclear Zeeman effect

$$E_M = g\mu_N m_I B \tag{14}$$

where g is the nuclear Landé factor and $\mu_N = eh/(2Mc)$ the nuclear Bohr magneton; that is, the magnetic hyperfine splitting. Taken together with the selection rule $\Delta m_I = 0, \pm 1$, this leads to a sextet in the Mössbauer spectrum.

Mössbauer spectroscopy (with varying temperature or magnetic field) allows us to probe electronic properties; that is, essentially the parameters of spin Hamiltonians resulting from the relative independency of nucleus-related parameters (like Q) in the above formulae on the chemical environment. A recent review of applications to biomolecules was given by Schünemann and Winkler [120]. In the domain of molecular magnets, Mössbauer spectroscopy became a standard tool for a primary characterization of newly synthesized Fe-based compounds. Taft et al. [121] reported (among other properties) the isomer shift and hyperfine field parameters of the circular "ferric wheel" molecule, $[Fe(OMe)_2(O_2CCH_2Cl)]_{10}$. Gorun et al. [122] performed a study on the $Fe_{11}O_6(OH)_6(O_2CPh)_{15}$. For these two systems, the isomer shift and quadrupole splitting were later calculated from first principles by Zeng et al. [123, 124], correspondingly. Bottyan et al. [125] studied two-dimensional molecular ferrimagnets with a Fe(II)–Fe(III) crossover. Coronado et al. [126] reported a series of hybrid layered molecular magnets in which Fe enters different crystallographic positions and appears in different spin configurations. In dinuclear spin-crossover Fe(II) compounds, the changes of Mössbauer spectra with temperature [127] and magnetic field, combined with light irradiation [128], have been reported.

4.2.2. Magnetic Resonance Techniques for Electron and Nuclear Spins

A reorientable magnetic moment **M** put in an external magnetic field $\mathbf{B} = (0, 0, B_z)$ at a general angle starts, according to classical electrodynamics, a precession with a frequency $\omega \sim MB_z$. The precession occurs at no energy cost, but feeding in some energy to the rotating moment [i.e. via applying a small alternating magnetic field in the (x, y) plane] leads to a resonance behavior: As the frequency of the (small) alternating field approaches ω, the system starts to absorb, and the rotating moment sets itself at a different (higher energetic) angle to **B**. In quantum systems, instead of precession, one speaks of splitting of energy levels in the external magnetic field, accounted for by the Zeeman term in the Hamiltonian—see Eq. (7) for spin systems—that lifts the degeneracy in m_S:

$$E_S = g\mu_B m_S B_z \tag{15}$$

If the rotating magnetic moment is that of a nucleus, the energy splitting is given by Eq. (14). The alternating field at frequency $\omega' = 2\pi\nu$ may then induce transitions between the split levels, either within a spin doublet $m_S = \pm\frac{1}{2}$, or—in case of nuclear spin I—between

$m_I = (-I, \ldots, I)$ levels, compatible with the selection rules $\Delta m_I = \pm 1$ for the perturbing matrix elements $\langle m_I | I_+ | m_I' \rangle$. The resonance condition is

$$h\nu = g\mu_{(\mathrm{B\,or\,N})} B_z \tag{16}$$

for μ_B or μ_N are Bohr magneton or nuclear magneton, and g is corresponding (spin or nuclear) Landé factor.

Although the physical foundation is the same (see, e.g., Ref. [129] for an introduction to the subject), the relation between the applied magnetic field B_z and the resonance frequency is very different for cases in which either electron spin or nuclear spin experience transitions between Zeeman levels. This gives rise to two different experimental techniques: electron spin resonance (ESR) and NMR. The resonance line can be scanned either by varying B at fixed probing frequency (which is a more common practice), or by tuning the frequency in the fixed field. Current NMR spectrometers have magnets with 12–15 T and operate in a radiofrequency range of 500–650 MHz; for ESR, because of smaller electron mass, and hence $\mu_B \gg \mu_N$, the characteristic frequency/field relation $\mu_B/h = 13.99624$ GHz/T is larger (e.g., typical spectrometer parameters could be 0.34 T at 9.5 GHz; there are different "bands" for microwave generation, which use different element base and cover, with gaps, the frequency range ~ 1 to ~ 35 GHz). This is, therefore, the microwave range, which requires quite different element base than the radiofrequency domain of NMR. The sample is placed in a cavity with a very high quality factor that is designed to enhance the microwave absorption. The latter requirement is one reason why it is much more convenient to sweep the external magnetic field and to keep the microwave radiation fixed. However, the frequency-domain magnetic resonance spectroscopy has its advantages, as is discussed below.

The ESR is also called EPR (see, e.g., 130), because its realization needs the presence of unpaired electron spins in partially occupied shells (i.e., paramagnetic centers in the sample). These spins can be TM ions or purely organic free radicals. TM ions may possess higher spin than $S = 1/2$; correspondingly, the structure of levels in magnetic field, and of the transitions between them, may become more complicated, as is illustrated by Fig. 10.

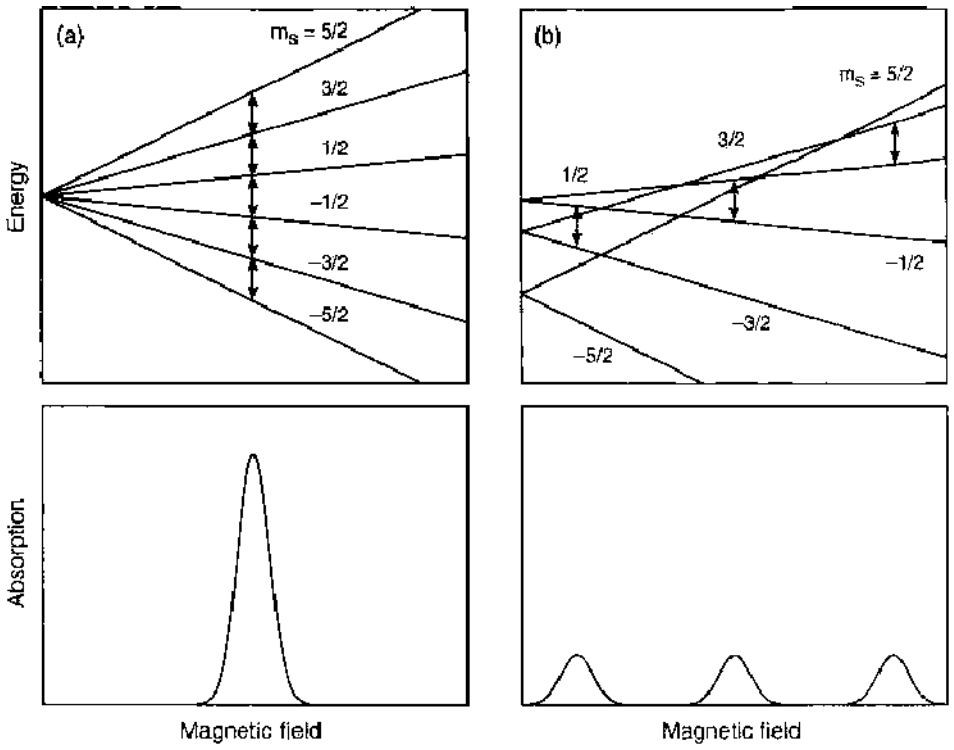


Figure 10. Schematic illustration of ESR absorption spectra for $S = 5/2$ without zero-field splitting (a) and with zero-field splitting (b) together with corresponding energy levels as function of magnetic field. The zero-field splitting is introduced by adding a term $D_0 m_S^2$ with $D_0 < 0$ to the spin Hamiltonian that results in multiple ESR absorption lines.

The case of $S = 5/2$ chosen there corresponds to, for example the Mn^{2+} ion. Figure 10a corresponds to the Zeeman splitting and five allowed equidistant transitions at any B value. Figure 10b is complicated by adding zero-field splitting—the term in the right side of of Eq. (7)—which lifts the degeneracy of levels in m_S and scales as m_S^2. In place of a single resonance value of B_z, one gets multiple ESR lines. (Note that not all allowed transitions are shown in Fig. 10b). One can expect that the intensities of these lines will change as the temperature decreases, so that the transitions involving the energetically lowest state will ultimately dominate. This can be either the low-field or the high-field line, depending on the sign of D.

The NMR (see, e.g., Ref. [131, 132], for an introduction to the subject) can be detected on nuclei that possess a spin moment; there are many isotopes that fulfill this requirement, but as the concentration of isotope must be sufficiently high, there are essentially ^1H, $^2H(\equiv {}^2D)$, and ^{13}C, which come into discussion for the study of molecular magnets (other possible candidates are ^{15}N and ^{31}F). Goto et al. [133] reported, however, a ^{55}Mn NMR study on Mn_{12}-ac and discussed the results for the transverse relaxation rate in terms of thermal fluctuations of the $S = 10$ quantum state.

The importance of magnetic resonance techniques for chemical studies is a result of the fact that the exact position of the resonance line, its shape, and possible fine structure depends on the chemical environment of the probing spin. It can be both an advantage and a disadvantage for extracting the information from the spectra that the detected resonance signal is accumulated over many individual spins, which may experience different environments.

Probably the most important characteristic for chemical characterization by means of NMR spectroscopy is the chemical shift. It means that as a result of shielding effects by orbital electron motion, and also because of variations in the spatial spin density, the local magnetic field at various crystallographic positions of chemically identical probing nuclei may differ, and these tiny variations on top of the large external field B_z (on the relative order of 10^{-6}) can still be resolved in NMR because of its high sensitivity (fractions of Hz, in the frequency domain). The amount of shielding for a given chemical environment is often known empirically, or it can be calculated from first principles (see, e.g., Sebastiani [134]). The chemical shift must in principle be present in the ESR spectra, but it is usually discussed in terms of effective g parameter. It incorporates the variations of the effective magnetic field at the position of the sample, some of which scale with the external field (because of the magnetic polarization of the medium), and others remain constant, as they are caused by dipole fields of the electron system and nuclei.

Another issue responsible for additional features in spectra is the hyperfine interaction. From the point of view of ESR, this is the splitting of m_S spin levels resulting from interaction with nuclear spins in the neighborhood.

Typical for the ESR practice is the incorporation of essentially all details related to the fulfillment of the resonance condition (i.e., the relation between $\mathbf{B}$ and $\mathbf{S}$) into the g factor. The observed resonance frequency, and hence the Zeeman splitting, can be anisotropic and is not well described by Eq. (15). The origin of this anisotropy is the contribution of the electron's orbital moment to the interaction with the magnetic field. The orbital moment is, of course, coupled to the local coordinate system at the paramagnetic center and depends on the chemical environment. Whereas one can formally consider a Zeeman effect for the orbital moment and the spin-orbit coupling, experimentally the orbital moment remains inaccessible to ESR. This is the reason why the spin Hamiltonian in ESR is formulated in terms of controllable effective vector variables, $\mathbf{B}$ and $\mathbf{S}$, with effective parameters in tensor form. Specifically, the single site part of Eq. (7) could be typically modified as

$$\mathcal{H} = \mathbf{S}\,\hat{\mathbf{D}}\cdot\mathbf{S} + \mu_B\mathbf{B}\cdot\hat{\mathbf{g}}\cdot\mathbf{S} - \mu_N g_N\mathbf{B}\cdot\mathbf{I}_N - \mathbf{S}\cdot\hat{\mathbf{A}}_N\cdot\mathbf{I}_N \tag{17}$$

including the hyperfine interaction with the nucleus and the nuclear Zeeman effect. The single-site anisotropy is also represented by a tensor, the main components of which, as with those of $\hat{g}$, can be obtained from fitting the experimental spectrum to the spin Hamiltonian.

Applications of ESR to molecular magnets by now have become quite numerous; apparently the majority of them deal with Mn_{12}-ac. Hill et al. [52] performed measurements on a single crystal at different orientations and temperature values. Among recent publications, Rakvin et al. [135] probed the magnetic fields and magnetic field gradients near the surface of crystalline Mn_{12}-ac. Bill et al. [136] illustrated the complementarity of ESR with Mössbauer spectroscopy for the study of a dimeric complex containing Fe(III) in high-spin state.

As mentioned above, the scanning of the ESR resonance signal is typically done by varying the magnetic field at fixed frequency. However, the technically more demanding frequency domain studies promise certain big advantages. For one thing, they allow to probe transitions between states with different m_S not only split by the magnetic field but also subject to zero-field splitting (as indicated in Fig. 10, right panel). Either by doing measurements at zero field or by observing the variations of resonant frequencies with B, one can effectively eliminate the Zeeman term from the spin Hamiltonian, thus reducing the number of parameters to be determined from fitting the data. The removal of all ambiguities related to the g-tensor allows us to concentrate on extracting the magnetic anisotropy parameters D, E from zero-field splittings and to estimate them with high precision. The ESR is indeed one of the most powerful experimental tools for accessing the zero-field splitting parameters, which are directly related to the magnetic anisotropy barrier of the SMM.

Working at low magnetic fields (including zero field) minimizes distortions of the line shape and allows us to extract additional information from the latter. An example of calculating the ESR linewidth on the basis of assumed values of spin Hamiltonian parameters (for Fe_8 and Mn_{12}-ac molecular magnets) has been given by Park et al. [137, 138].

Another group of allowed ESR transitions in molecular magnets, related neither to the Zeeman effect nor to zero-field anisotropy, are transitions between excited states m_S of the total spin S of a molecule. Taking for example Mn_{12}-ac, these are transitions between $m_S = \pm 10$ (ground state) and $m_S = \pm 9$, but as the temperature grows (from about 3–15 K; see Ref. [139]), the higher states with smaller $|m_S|$ are populated, and more transitions become observable. The energy difference between $m_S = \pm 10$ and $m_S = \pm 9$ is about 10 cm^{-1} in Mn_{12}-ac, which corresponds in terms of frequency to 300 GHz. Such transitions were indeed detected by Parks et al. [139], using time-domain terahertz spectroscopy.

A review on the frequency-domain ESR techniques, including many applications to molecular magnets, has been recently provided by van Slageren et al. [140].

4.3. Relation to First-Principles Calculations

By first-principles calculations, we mean those that attempt to solve the most general quantum-mechanical equations; i.e., either Schrödinger or Dirac equations, for a system of either fixed or movable point-charge nuclei, accommodating many electrons. The exact analytic solution for the distribution of electrons is not possible for cases more complex than a hydrogen-like (one-electron) ion. The necessary approximations can be divided into conceptual and technical ones; but important for both types of approximations is that the approximation should be of general type and not *ad hoc* for the particular system under investigation. In other words, we speak about methods that do not include adjustable parameters specific to the system of our interest and that would need tuning to agree with external (experimental) information, in contrast to empirical or semiempirical methods. The mere chemical composition and, at most, the assumption of the crystal structure should be fully sufficient to perform the simulation and to calculate the electronic and magnetic properties. From the ability to calculate total energies and forces, it becomes further possible to investigate the relative stability of crystal structures and to predict elastic, vibrational, or transport properties.

The conceptual approximations are those that make the problem technically solvable. The approximation with the largest effect historically is the mean field approximation, which reduces the many-body problem to a single-particle one. This approximation laid the foundations of the Hartree–Fock (HF) method and was one way that lead to density-functional theory. The basic assumption in the HF method is that the N-electron wavefunction

$\psi(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N)$ is represented by a single Slater determinant

$$\Phi(\mathbf{r}_1, \ldots, \mathbf{r}_N) \equiv \frac{1}{\sqrt{N!}} \begin{vmatrix} \varphi_1(\mathbf{r}_1) & \cdots & \varphi_1(\mathbf{r}_N) \\ \vdots & & \vdots \\ \varphi_N(\mathbf{r}_1) & \cdots & \varphi_N(\mathbf{r}_N) \end{vmatrix} \tag{18}$$

constructed from one-electron wavefunctions $\varphi_\alpha(\mathbf{r})$. Such functions can be obtained variationally from solving a set of coupled integro-differential HF equations in a fixed external potential (e.g., Coulomb field of fixed nuclei) $U(\mathbf{r})$

$$\left[-\frac{\hbar^2}{2m}\nabla^2 + U(\mathbf{r}) + \sum_j \int \frac{\varphi_j^*(\mathbf{r}')\varphi_j(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\right]\varphi_\alpha(\mathbf{r}) - \sum_j \int \frac{\varphi_j^*(\mathbf{r}')\varphi_j(\mathbf{r})\varphi_\alpha(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d\mathbf{r}' \equiv \varepsilon_\alpha \varphi_\alpha(\mathbf{r}) \tag{19}$$

The HF method as such does not provide chemical accuracy: that is, a quantitative satisfactory attribution of total energies to ingredients and products of a typical chemical reaction. However, a systematic improvement can be obtained by allowing additional contributions of the type of Eq. (18), or different electronic configurations, to the many-body wave function. The construction of such multideterminantal wave functions ultimately provides sufficient variational freedom to adjust the expectation values of the total energy in different electronic configurations of the chemical constituents to reach chemical accuracy, making the method interesting for practical predictions. First-principles calculations with the use of multideterminantal wave functions constitute the domain of quantum chemistry (QC). Apart from details of implementation [how to mix different configurations, how to express one-electron functions $\varphi_\alpha(\mathbf{r})$ in Slater determinants numerically], the HF method and its multiconfigurational generalization do not have any parameters whatsoever. They can therefore often serve as clear-cut benchmarks, and the notion of Hartree–Fock limit in a certain calculation makes sense as an ultimate result, if the errors of the technical (numerical) realization are sufficiently suppressed.

The situation is somehow different when we consider methods based on density-functional theory. Leaving aside the variety of ways one has introduced, or justified, the theory (see Dreizler and Gross [141] and Eschrig [142] for reviews), its central message is the removal of the many-electron wave function from the picture entirely, putting in its place the one-electron density

$$\rho(\mathbf{r}) = \int \psi^*(\mathbf{r}, \mathbf{r}_2, \ldots, \mathbf{r}_N)\psi^*(\mathbf{r}, \mathbf{r}_2, \ldots, \mathbf{r}_N) d\mathbf{r}_2 \ldots d\mathbf{r}_N$$

as subject to variation in the search for the total energy of the ground state. This is sufficient for applications that just aim at the density $\rho(\mathbf{r})$ and the total energy, but not at the structure of the wave function, and such applications indeed comprise a large class of relevant numerical studies for practical purposes, whereas the applicability to the ground state only is a certain limitation. In practice, it is important that the variational search occurs near a well-isolated local minimum on a total energy hypersurface, with respect to variation parameters. For example, different magnetic configurations might be typically treated without problem, and their respective "ground-state energies" reasonably compared.

The cornerstone of DFT is the Hohenberg–Kohn theorem [143] that the total energy E in the ground state is a functional of the charge density $\rho(\mathbf{r})$ only. As this functional $E[\rho(\mathbf{r})]$ remains (yet) unknown, however, several working approximate schemes have been developed. Following Kohn and Sham [144], the total energy can be split into several clearly defined parts (kinetic energy of noninteracting quasiparticles, electrostatic interaction of electron density with itself and with external fields) and a parameterized rest (the exchange–correlation energy, $E_{XC}[\rho]$). An important difference to semiempirical methods is that this parameterization is not supposed to vary *ad hoc* from system to system but, rather, attempts to reflect certain basic and universal features of the underlying electronic system one deals with. In particular, one hopes that it describes exchange and correlation effects of the inhomogeneous electron liquid adequately. Using perturbation expansions within the many-body

theory, or high-precision numerical simulations of the model system with the quantum Monte Carlo technique [145, 146], it has been possible to obtain some estimates of the exchange and correlation energy for various densities. This information has been used to derive a parameterized description of the exchange-correlation energy. This allows us to include the effect of electronic correlations (i.e., mutual avoidance of electrons beyond the simplest mean-field approximation) that remains out of the HF scheme by its construction. However, the extent to which the correlations are included remains unclear, and different prescriptions to treat exchange-correlation have been developed over the last decades to give rise to numerous "families" of exchange-correlation potentials beyond the earliest and simplest one, the local density approximation (LDA). In particular, emphasizing different aspects of the inhomogeneous electron density resulted in gradient corrected approximations [147, 148], weighted density approximation [149, 150], self-interaction corrected schemes [151, 152], and special schemes for treating strongly correlated systems, like LDA + U [153]. Such flexibility considerably improved the numerical precision, meeting the chemists' aspirations of chemical accuracy, and paved the way to broad use of density-functional methods in computational chemistry as a reliable and computationally less expensive alternative to quantum-chemical methods [154–156]. On the other side, as some recipes tend to work well for certain systems and not for others, one has to accept the not so pleasant reality that essentially all prescriptions for exchange-correlation contain explicit or hidden assumptions and parameterizations, the domain of validity of which is not always obvious.

Even as the concept of wave function is pushed to the background in the density-functional theory formalism, it often reemerges for the sake of chemical transparency and analysis of results. It is difficult to discuss bonding and antibonding states having access only to the charge density. Although a density-functional theory calculation for a N-electron system can in principle be realized by referring to the density $\rho(\mathbf{r})$ only, it turns out to be very convenient, as was proposed by Kohn and Sham [144], to express the density $\rho(\mathbf{r})$ (searched for) via fictitious functions $\psi_i(\mathbf{r})$, which are postulated to be wavefunctions of noninteracting quasiparticles without apparent physical meaning, but possessing the same density as the true physical system:

$$\rho(\mathbf{r}) = \sum_{\alpha=1}^{N} |\psi_\alpha(\mathbf{r})|^2 \tag{20}$$

The primary advantage of this construction is the simplicity in calculating the kinetic energy, which otherwise has no obvious form as a functional of $\rho(\mathbf{r})$, and the possibility of formulating a set of (Kohn–Sham) equations for $\psi_i(\mathbf{r})$ that are conceptually similar to the HF equations in Eq. 19

$$\left[-\frac{\hbar^2}{2m}\nabla^2 + U(\mathbf{r}) + \int \frac{\rho(\mathbf{r}')d\mathbf{r}'}{|\mathbf{r}-\mathbf{r}'|} + \frac{\delta E_{\mathrm{XC}}[\rho]}{\delta\rho(\mathbf{r})}\right]\psi_\alpha(\mathbf{r}) \equiv \varepsilon_\alpha\psi_\alpha(\mathbf{r}) \tag{21}$$

This brings in a convenient quantum-chemical environment in the sense that one may discuss hybride orbitals, positions of energy levels, and density of states distribution, even as all this has no solid foundation and, strictly speaking, no physical meaning in the framework of the Kohn–Sham formalism. Only the total energy and the electron density can be meaningfully discussed. Even as some amount of interelectronic correlation may be taken into account by the choice of exchange-correlation functional, the resulting charge density still has a one-determinantal character because its construction in the Kohn–Sham method. This imposes certain limitations on the interpretation of DFT results, which must be treated differently from their QC counterparts for the extraction of magnetic interaction parameters from calculation, as discussed below.

The reintroduction of the many-body wave function beyond its single-determinantal representation, however, faces no principal obstacles in the framework of density-functional theory. Capelle [157] proposed a way in which to perform the variational search over several determinants of Kohn–Sham functions that is similar but not identical to allowing for configuration mixing in quantum chemical schemes. This should result in a systematic lowering of

the estimated ground-state energies and improve the accuracy in representing charge density distributions in more complex systems.

After discussing the conceptual limitations, we note briefly that any first-principles scheme also contains approximations of technical, or numerical, character. They deal with organizing the solution of either HF or Kohn–Sham equations, the choice of appropriate basis set, performing spatial integrations, and so forth. Some discussion on this subject, in view of the applications to SMM, is given in Section 5.

Turning now to the problem of evaluating the parameters of spin Hamiltonians from first principles, we note that in executing the calculation, one has the freedom to impose certain constraints (fix the magnitude or orientation of magnetization, modify the potential felt by certain electronic states, switch the relativistic effects on or off) and inspect the effect of these constraints on the total energy. Moreover, one-electron eigenvalues and corresponding (Kohn–Sham or HF) eigenfunctions are also available from a self-consistent calculation. There are certain subtleties related to the extraction of exchange parameters from QC and DFT calculations that one should be aware of.

In QC, one deals with a multiconfigurational scheme that allows us to mix different spin configurations and to classify energy eigenvalues according to different total spin values. For two interacting spins $\mathbf{S}_1$, $\mathbf{S}_2$ summing up to $\mathbf{S}' = \mathbf{S}_1 + \mathbf{S}_2$, one gets

$$2\mathbf{S}_1\mathbf{S}_2 = \mathbf{S}'^2 - \mathbf{S}_1^2 - \mathbf{S}_2^2$$

with eigenvalues $[S'(S'+1) - S_1(S_1+1) - S_2(S_2+1)]$. For a textbook example $S_1 = \frac{1}{2}$, $S_2 = \frac{1}{2}$ this yields singlet ($S' = 0$) and triplet ($S' = 1$) states. The corresponding eigenvalues of the Heisenberg Hamiltonian must then be $3/2J$ and $-1/2J$, correspondingly.

Indeed, the basis functions in an *ab initio* calculation are normally pure spin states. On the basis of spin functions $|m_{S_1} m_{S_2}\rangle$, for the case $S_1 = \frac{1}{2}$, $S_2 = \frac{1}{2}$, the Heisenberg Hamiltonian takes the form:

$m_{S_1} m_{S_2}$	$\left\vert\frac{1}{2}\frac{1}{2}\right\rangle$	$\left\vert-\frac{1}{2}\frac{1}{2}\right\rangle$	$\left\vert\frac{1}{2}-\frac{1}{2}\right\rangle$	$\left\vert-\frac{1}{2}-\frac{1}{2}\right\rangle$
$\left\vert\frac{1}{2}\frac{1}{2}\right\rangle$	$-\frac{J}{2}$			
$\left\vert-\frac{1}{2}\frac{1}{2}\right\rangle$		$\frac{J}{2}$	$-J$	
$\left\vert\frac{1}{2}-\frac{1}{2}\right\rangle$		$-J$	$\frac{J}{2}$	
$\left\vert-\frac{1}{2}-\frac{1}{2}\right\rangle$				$-\frac{J}{2}$

(22)

The diagonalization of Eq. (22) is achieved by a basis transformation that mixes different m_S values:

$$\begin{aligned}
&\frac{1}{\sqrt{2}}\left(\left\vert\frac{1}{2}-\frac{1}{2}\right\rangle - \left\vert-\frac{1}{2}\frac{1}{2}\right\rangle\right) && \text{(singlet)}\ \ S=0\ \ E=\frac{3}{2}J \\
&\left\vert\frac{1}{2}\frac{1}{2}\right\rangle \\
&\frac{1}{\sqrt{2}}\left(\left\vert\frac{1}{2}-\frac{1}{2}\right\rangle + \left\vert-\frac{1}{2}\frac{1}{2}\right\rangle\right) && \text{(triplet)}\ \ S=1\ \ E=-\frac{1}{2}J \\
&\left\vert-\frac{1}{2}-\frac{1}{2}\right\rangle
\end{aligned} \tag{23}$$

In a QC (multideterminantal) calculation, the eigenvalues of singlet and triplet states, E_S and E_T, are immediately accessible. This allows the (formal yet unambiguous) mapping of a first-principles result onto the Heisenberg model:

$$E_S - E_T = 2J \tag{24}$$

compared to the model case of two electrons in Eq. (4). The case $S_{1,2} = \frac{1}{2}$ corresponds to, for example, two interacting Cu^{2+} ions. Other ions from the $3d$ row yield richer systems of eigenvalues; for instance, $S_{1,2} = 1$ (two Ni^{2+} ions) produces a quintet level E_Q beyond singlet and triplet, with the energy separation

$$E_S - E_Q = 6J \tag{25}$$

Whether both Eqs. (24) and (25) can be satisfied by the same J is a measure of validity of the Heisenberg model.

In density-functional theory, the search for the "true" wave function is avoided (see Capelle [157] for an attempt to address this problem) and is substituted for the variational search for the charge density and total energy. As the eigenvalues of multideterminantal states are not available, one must rely either on the Kohn–Sham eigenvectors or on total energies in specially prepared symmetry-breaking metastable states, subject to different constraints with respect to the spin states of a system. In practice, one can try FM or antiferromagnetic (AFM) configurations of two spins, or impose the fixed spin moment (FSM) scheme, first introduced by Schwarz and Mohn [158]. The total energy in different spin configurations does not relate to the eigenvectors but to diagonal elements of, for example, the Hamiltonian $\mathscr{H}$ of Eq. (22)

$$\begin{aligned} E_{FM} &= \left\langle \frac{1}{2}\frac{1}{2} \middle| \mathscr{H} \middle| \frac{1}{2}\frac{1}{2} \right\rangle = -\frac{1}{2}J \\ E_{AFM} &= \left\langle -\frac{1}{2}\frac{1}{2} \middle| \mathscr{H} \middle| -\frac{1}{2}\frac{1}{2} \right\rangle = -\frac{1}{2}J \end{aligned} \tag{26}$$

and hence

$$E_{AFM} - E_{FM} = J \tag{27}$$

for the above case of $S_{1,2} = \frac{1}{2}$. This is a valid representation for J provided the Heisenberg model itself remains valid throughout the path from FM to AFM state. The latter formula can be approximated using the concept of a magnetic transition state [159]. In general, according to Slater, the shift in the density-functional theory total energy ΔE resulting from a whatever change Δn_i in the occupation of certain orbitals is

$$\Delta E = \sum_i \Delta n_i \varepsilon_i^* + \mathcal{O}(\Delta n^3) \tag{28}$$

where ε_i^* are Kohn–Sham eigenvalues obtained self-consistently with occupation numbers midway between initial and final states. For the flip from FM to AFM configuration,

$$E_{FM} - E_{AFM} \simeq \sum_i (n_{i\uparrow}^A - n_{i\downarrow}^A)(\varepsilon_{i\downarrow}^* - \varepsilon_{i\uparrow}^*) \tag{29}$$

where $(n_{i\uparrow}^A - n_{i\downarrow}^A)$ is the magnetic moment (which gets inverted) in the orbital i. The latter bracket is the spin splitting (in energy) of the same orbital, calculated in the configuration with zero spin on atom A (transition state); that is, induced fully via the interaction with the second spin. Although being approximative, the magnetic transition state scheme might have a certain advantage of numerical stability over explicit comparison of large total energy values. Moreover, the result is available from a single calculation and offers a microscopical insight into how different orbitals are affected by magnetic interaction—information that remains hidden in the total energy numbers. As it has been of use a number of times in the

past (primarily for magnetic oxides), the method has recently been applied to the analysis of exchange parameters in Mn_{12}-ac [124].

The validity of either the "finite difference" scheme of Eq. (27), or "differential" procedure of Eq. (29) presumes that the mapping onto the Heisenberg model makes sense in the first place. However, with just two interacting spins, we have no immediate criterion about whether this is true. The applicability of the Heisenberg model would mean that the functional part of the interaction comes from the scalar product of two spin operators, with the parameter J_{ij} being independent on $\mathbf{S}_i$ and $\mathbf{S}_j$. The mapping on the Heisenberg model may be less ambiguous if it is done as a limiting case of small deviations from a certain stationary state. The meaning of such deviations in the density-functional theory might be some admixture to pure spin states (in the sense of local spin density functional); that is, the nondiagonal (in the spin space) form of density matrices. It allows a transparent quasi-classical interpretation in terms of noncollinear magnetic density varying from point to point in space (see Sandratskii [160] for a review). If a pair of local magnetic moments can be reasonably identified in the calculation, and their small variations from the global magnetization axis are allowed, the counterparts in the Heisenberg model will be deviations of local exchange fields at two corresponding sites. Matching the leading terms in the angular dependence of interaction energy in the density-functional theory and in the Heisenberg model leads to the desired mapping. This line of arguing goes back to at least Oguchi et al. [161], who extracted interaction parameters in simple $3d$ oxides from DFT calculations. Liechtenstein et al. [162, 163] and Antropov et al. [164] worked out closed expression for J_{ij} in a form consistent with spin-fluctuation theories (with magnitude of spin varying and attributed to S) in terms of the elements of the Green's function [Note, therefore, the difference in the definition of J_{ij} between these papers and that in Eq. (5)]. When using the final formulae, one should be careful to check whether it was not implicitly assumed that $S = 1/2$, and also examine the prefactor and sign of the exchange parameter that may be introduced differently. The following line of argument leads to a formula that has been used in a number of calculations. If the total interaction energy of two quasi-classical spins is

$$E = J_{ij}\mathbf{S}_i\mathbf{S}_j \tag{30}$$

its variation caused by the change of the angles of the spins $\delta\varphi_i$, $\delta\varphi_j$ reads

$$\delta^2 E = J_{ij}S^2\delta\varphi_i\delta\varphi_j \tag{31}$$

In the attempt to cast a variation of density-functional theory total energy in a comparable form, one can profit from Andersen's local force theorem, which works here because we are interested in infinitesimal deviations from the ground state. An explicit derivation of the local force theorem in the desired form is given in an Appendix of the paper by Liechtenstein et al. [163]. In terms of the Green's function G and Kohn–Sham Hamiltonian $\mathcal{H}$, the first variation of the total energy reads

$$\delta E = -\frac{1}{\pi}\int^{\epsilon_f} d\epsilon \ \mathrm{Im\ Tr}\ (\delta\mathcal{H}G) \tag{32}$$

(which can be shown to be zero), and the second variation is

$$\delta^2 E = -\frac{1}{\pi}\int^{\epsilon_f} d\epsilon \ \mathrm{Im\ Tr}\ (\delta^2\mathcal{H}G + \delta\mathcal{H}G\,\delta\mathcal{H}G) \tag{33}$$

The variation of the Kohn–Sham Hamiltonian can be explicitly related to rotations in spin space as

$$\delta\mathcal{H} = \frac{i}{2}\delta\varphi_i[\mathcal{H}, \sigma] \tag{34}$$

with the Hamiltonian composed of a spin-dependent part at the site i, with $\Delta_i = V_i^{\uparrow} - V_i^{\downarrow}$ [a potential, in general, nondiagonal in (l, m)] and the rest $\mathcal{H}_0$:

$$\mathcal{H} = \frac{\Delta_i}{2}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} + \mathcal{H}_0\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{35}$$

This yields for the variation of $\mathcal{H}$

$$\delta\mathcal{H} = \frac{i}{2}\delta\varphi_x\Delta_i\begin{pmatrix}0 & 1\\ -1 & 0\end{pmatrix} + \frac{i}{2}\delta\varphi_y\Delta_i\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix} \tag{36}$$

Extracting from Eq. (33) the terms bilinear in $\delta\varphi_i$, recovering site and spin indexes in the elements of the Green's function G_{σ}^{ij} and implying the summation in (l, m) yields

$$J_{ij} = -\frac{1}{2\pi}\mathrm{Im}\int^{\epsilon_F} d\epsilon\left(\Delta_i G_{\uparrow}^{ij}\Delta_j G_{\downarrow}^{ji} + \Delta_i G_{\downarrow}^{ij}\Delta_j G_{\uparrow}^{ji}\right) \tag{37}$$

This is the final formula for the interaction between isolated spins in an otherwise infinite and unperturbed environment. If one is interested in the interaction between two sublattices of periodically repeated atom types i and j, the Green function follows explicitly in terms of Kohn–Sham eigenfunctions $\psi_{n\mathbf{k}\sigma}^{ilm}$ and eigenvalues $\epsilon_{n\mathbf{k}\sigma}$

$$G_{lm,l'm'}^{ij}(\epsilon) = \sum_{n\mathbf{k}}\frac{\psi_{n\mathbf{k}}^{*ilm}\psi_{n\mathbf{k}}^{jl'm'}}{\epsilon - \epsilon_{n\mathbf{k}}} \tag{38}$$

Using the following relation for the product of Green's functions

$$\frac{1}{(\epsilon-\epsilon_n)(\epsilon-\epsilon_{n'})} = \frac{1}{\epsilon_n-\epsilon_{n'}}\left(\frac{1}{\epsilon-\epsilon_n} - \frac{1}{\epsilon-\epsilon_{n'}}\right) \tag{39}$$

the integration in energy over occupied states yields

$$J_{ij} = \sum_{\{m\}}\Delta_{mm'}^{i}\chi_{mnm''m'''}^{ij}\Delta_{m''m'''}^{j} \tag{40}$$

in terms of a nonlocal susceptibility, which depends on the Kohn–Sham occupation numbers $n_{n\mathbf{k}\sigma}$,

$$\chi_{mnm''m'''}^{ij} = \sum_{\mathbf{k}nn'}\frac{n_{n\mathbf{k}\uparrow} - n_{n'\mathbf{k}\downarrow}}{\epsilon_{n\mathbf{k}\uparrow} - \epsilon_{n'\mathbf{k}\downarrow}}\psi_{n\mathbf{k}\uparrow}^{*ilm}\psi_{n\mathbf{k}\uparrow}^{jlm''}\psi_{n'\mathbf{k}\downarrow}^{jln'}\psi_{n'\mathbf{k}\downarrow}^{*ilm'''} \tag{41}$$

which is a formula probably first given by Liechtenstein et al. [165] and used in a number of publications, notably by Boukhvalov et al. [166] for Mn_{12}-ac. It should be understood that this formula describes the interaction between two sublattices rather than two spins, and hence may give numbers very different from those. The above derivation relates to $S = 1/2$, and therefore the values reported for J_{ij} had to be rescaled according to the values of the actual interacting spins (e.g., Mn atoms), a fact not always clearly stated in publications. Such scaling in Ref. [166] has been performed using the LDA (fractional) calculated values of magnetic moments, rather than nominal (integer) values (D. Boukhvalov, private communication).

4.4. Spin-Orbit Coupling and Magnetic Anisotropy Energy

Even for cubic bulk crystals like Fe, Co, or Ni, it has been found that along certain crystallographic directions it is easy to magnetize the crystal, and along others it is harder (see, e.g., Blundell [107]). In a general case, there will be a magnetic easy, medium, and hard axis. The anisotropy leads to an additional energy dependence on the direction of magnetization that is caused by spin-orbit coupling and partial quenching of the orbital momentum by the crystal field.

A special case is uniaxial anisotropy in which the energy depends only on the angle with respect to a single axis. Two of the axes have the same energy, and one axis is different (this is fully analog to the definition of uniaxial crystals). If the lowest energy corresponds to the single axis, that means that the magnetization would prefer to align along that axis, and the system is called "easy axis". In the case in which the magnetization would be favored in the plane spanned by the other two axes, the system is called "easy plane."

The magnetic anisotropy energy is quite small—usually between several meV to μeV per magnetic atom. Please note that in cubic systems, the second-order contributions to the anisotropy energy vanish by symmetry, so that like in bulk Fe, Co, or Ni, there will be only fourth-order contributions that are very small. The molecular nanomagnets we are interested in are mostly interesting because of their large magnetic anisotropy. All the single-molecule magnets that generated interest in the past have a relatively low symmetry, so that the second-order contributions, which are orders of magnitude larger than the fourth-order contributions in bulk crystals, do not vanish. If one wants to design a molecular nanomagnet, one has to stay away from cubic symmetry but should try to make it an easy axis system.

An additional but completely different form of anisotropy is shape anisotropy, which is caused by the demagnetization energy associated with the sample shape. In thin films, for example, the shape anisotropy may lead to a preferential in plane magnetization. We will not deal with this kind of anisotropy here, although one should remember that experimental results may have to be corrected for that contribution. In the remainder of the text, we will refer to the magnetocrystalline anisotropy as the magnetic anisotropy only, neglecting the shape anisotropy completely.

As early as 1937, van Vleck pointed out that the MAE arises mainly because of spin-orbit coupling and other relativistic terms in the Hamiltonian. Calculations of MAE in solids, layered structures, and films have been carried out for many years using density-functional theory [167–171]. Several problems associated with the accurate density-functional-based determination of MAE in the solid state have been identified. For example, the role of incomplete orbital polarization has been shown to be one issue related to inaccuracies in the solid; others may be related to correlation effects beyond the mean-field treatment of correlations in the DFT.

Recently, Pederson and Khanna [172, 173] have developed a method for accounting for second-order anisotropy energies. This method relies on a simple albeit exact method for spin-orbit coupling and a second-order perturbative treatment of the spin Hamiltonian to determine the dependence of the total energy on spin projection. It makes use of the Cartesian representation of the spin-orbit term, which is exact and is also more adaptable for multicenter systems

$$U(\mathbf{r}, \mathbf{p}, \mathbf{S}) = -\frac{1}{2c^2}\mathbf{S} \cdot \mathbf{p} \times \nabla\Phi(\mathbf{r}) \tag{42}$$

Using single-particle wavefunctions expressed in terms of a basis set

$$\psi_i(\mathbf{r}) = \sum_{j,\sigma} C_{j\sigma}^{i\nu}\phi_j(\mathbf{r})\chi_\sigma \tag{43}$$

where the $\phi_j(\mathbf{r})$ are the spatial functions and χ are spin functions, the matrix elements can be expressed as

$$U_{j,\sigma,k\sigma'} = \langle\phi_j\chi_\sigma|U(\mathbf{r}, \mathbf{p}, \mathbf{S})|\phi_k\chi_{\sigma'}\rangle \tag{44}$$

$$= -i\langle\phi_j|V_x|\phi_k\rangle\langle\chi_\sigma|S_x|\chi_{\sigma'}\rangle \tag{45}$$

where the operator V_x is defined as

$$\langle\phi_j|V_x|\phi_k\rangle = \frac{1}{2c^2}\left(\left\langle\frac{d\phi_j}{dz}\middle|\Phi\middle|\frac{d\phi_k}{dy}\right\rangle - \left\langle\frac{d\phi_j}{dy}\middle|\Phi\middle|\frac{d\phi_k}{dz}\right\rangle\right) \tag{46}$$

In the above, $\Phi(\mathbf{r})$ is the Coulomb potential. Thus, this treatment uses matrix elements of the Coulomb potential with partial derivatives of the basis functions, thereby avoiding the time-consuming task of calculating the gradient of the Coulomb potential directly.

Here we generalize some of the derivations from uniaxial symmetry to an arbitrary one. The same definitions and lettering of the symbols is used as by Pederson and Khanna [173]. In the absence of a magnetic field, the second-order perturbative change to the total energy

of a system with arbitrary symmetry can be expressed as

$$\Delta_2 = \sum_{\sigma\sigma'} \sum_{ij} M_{ij}^{\sigma\sigma'} S_i^{\sigma\sigma'} S_j^{\sigma'\sigma} \tag{47}$$

which is the generalization of Eq. (19) of Pederson and Khanna [173]. In the above expression, σ sums over the spin degrees of freedom and i, j sums over all the coordinate labels, x, y, z, respectively. The matrix elements $S_i^{\sigma\sigma'} = \langle \chi^{\sigma} | S_i | \chi^{\sigma'} \rangle$ implicitly depend on the axis of quantization. The matrix elements $M_{ij}^{\sigma\sigma'}$ are given by

$$M_{ij}^{\sigma\sigma'} = -\sum_{kl} \frac{\langle \phi_{l\sigma} | V_i | \phi_{k\sigma'} \rangle \langle \phi_{k\sigma'} | V_i | \phi_{l\sigma} \rangle}{\varepsilon_{l\sigma} - \varepsilon_{k\sigma'}} \tag{48}$$

where $\phi_{l\sigma}$ are occupied and $\phi_{k\sigma'}$ are unoccupied states and ε is the energy of the corresponding states.

The above equation can be rewritten in a part diagonal in the spin index plus the nondiagonal remainder according to:

$$\Delta^2 = \sum_{ij} \sum_{\sigma} M_{ij}^{\sigma\sigma} S_i^{\sigma\sigma} S_j^{\sigma\sigma} + \sum_{ij} \sum_{\sigma \neq \sigma'} M_{ij}^{\sigma\sigma'} S_i^{\sigma\sigma'} S_j^{\sigma'\sigma} \tag{49}$$

Using the following relation for the expectation value of a spin operator in a closed shell molecule with excess majority spin electrons ΔN

$$\langle 1 | S_i | 1 \rangle = -\langle 2 | S_i | 2 \rangle = \frac{\langle S_i \rangle}{\Delta N} \tag{50}$$

the first term of Eq. (49) can be expressed as

$$\sum_{ij} \left(M_{ij}^{11} + M_{ij}^{22} \right) \frac{\langle S_i \rangle \langle S_j \rangle}{(\Delta N)^2} \tag{51}$$

With the help of

$$\begin{aligned} \langle 1 | S_i | 2 \rangle \langle 2 | S_j | 1 \rangle &= \langle 1 | S_i S_j | 1 \rangle - \langle 1 | S_i | 1 \rangle \langle 1 | S_j | 1 \rangle \\ &= \langle 1 | S_i S_j | 1 \rangle - \frac{\langle S_i \rangle \langle S_j \rangle}{(\Delta N)^2} \end{aligned} \tag{52}$$

and similar relation for $\langle 2 | S_i | 1 \rangle \langle 1 | S_j | 2 \rangle$; with a bit of algebra, the second term of Eq. (49) becomes

$$\sum_{ij} -\left(M_{ij}^{12} + M_{ij}^{21} \right) \frac{\langle S_i \rangle \langle S_j \rangle}{(\Delta N)^2} + \frac{1}{4} \sum_{i} \left(M_{ii}^{12} + M_{ii}^{21} \right) \tag{53}$$

Therefore, the total second order shift Δ_2 together from Eq. (51), and Eq. (53) becomes

$$\Delta_2 = \frac{1}{4} \sum_{i} \left(M_{ii}^{12} + M_{ii}^{21} \right) + \sum_{ij} \left(M_{ij}^{11} + M_{ij}^{22} - M_{ij}^{22} - M_{ij}^{21} \right) \frac{\langle S_i \rangle \langle S_j \rangle}{(\Delta N)^2} \tag{54}$$

As it can be easily verified, the last equation gives the same result for uniaxial symmetry as Eq. (21) of Pederson and Khanna [173], where the Cartesian off-diagonal M_{ij} matrices vanish and $M_{xx}^{\sigma\sigma'} = M_{yy}^{\sigma\sigma'}$. For the derivation of the above expression of Δ_2, we did not assume any particular symmetry: therefore, the resulting expression is general.

In the following, we overview the record of first-principles calculations on some SMM, outline a few typical problems, and discuss the achieved results and remaining difficulties.

5. OVERVIEW OF DENSITY-FUNCTIONAL THEORY COMPUTATIONAL SCHEMES

Beyond the conceptual approximations adopted for solving the any-electron problem, as discussed in Section 4.3 (e.g., HF approximation, QC multideterminantal approach, or Kohn–Sham equations), one has to choose means by which to solve the corresponding equations numerically. This involves additional approximations, which are of purely technical character but demand a fair amount of physical insight and programming sophistication to combine accuracy with feasibility of the calculations. In virtually all cases, one has to decide first on an appropriate set of basis functions $\chi_\mu(\mathbf{r})$ used to expand the sought-for one-electron orbitals $\psi_\alpha(\mathbf{r})$, which enter the Kohn–Sham equations Eq. (21):

$$\psi_\alpha(\mathbf{r}) = \sum_{\mu=1}^{Q} C_{\alpha\mu}\chi_\mu(\mathbf{r}) \tag{55}$$

This expansion is always finite, but the dimension of the basis Q must be reasonably larger than the number of occupied electronic states N for providing sufficient flexibility in the variational search for the solution of either Kohn–Sham, or HF equations. The two most common choices of $\chi_\mu(\mathbf{r})$ are plane waves and atom-centered localized functions. The former are defined as

$$\chi_G(\mathbf{r}) = \frac{1}{\sqrt{\Omega}} e^{i\mathbf{G}\mathbf{r}} \tag{56}$$

(i.e., labeled by vectors $\mathbf{G}$ of the reciprocal lattice, which corresponds to a periodic unit cell of volume Ω). (We skip the $\mathbf{k}$-dependence in our discussion, as the energy dispersion in the Brillouin zone is almost negligible for the weakly interacting SMM.) Such a periodic cell (simulation box) must always be introduced for calculations with planewave basis sets, even if the simulated system is in reality not periodic (e.g., a single molecule). The number of planewave basis functions needed for sufficient accuracy grows rapidly with the size of the simulation box, independent of the actual number of atoms contained in the box. The planewave basis (see, e.g., Ref. [174], for more details) has the advantage of becoming ultimately complete under the variation of a single cutoff parameter, $G_{\max}$, as it includes all planewaves with $|G| \leq G_{\max}$. The planewave cutoff energy can be kept reasonably low by using, instead of true (deep near the nuclei) Coulomb potentials, screened pseudopotentials and correspondingly smoothed pseudofunctions for electrons in valence shells, thus excluding the core states from the calculation.

The atom-centered functions, in contrast are better suited for describing strong spatial fluctuations of the one-electron functions within atoms, and hence allow much smaller (and even almost minimal, $Q \gtrsim N$) basis sizes. However, they face problems, or at least ambiguities, in a consistent generation of efficient basis sets and in performing spatial integrations. The atom-centered functions can be further divided into numerical and analytical ones, energy dependent or not, fixed or adjustable in the course of iterating the Kohn–Sham equations to self-consistency. A common workable choice among fixed analytical basis functions are Gaussian-type orbitals (see e.g., Ref. [175], for a review).

Recently, there has been a notable increase in the number of calculations that solve the underlying equations on a real-space grid [176, 177], with finite differences or finite elements technique [178, 179]. Yet this is equivalent to the use of piecewise linear or polynomial functions, localized at grid points, as a basis. Attempts to combine the advantages of planewave and localized-basis techniques resulted in a number of high-precision calculation schemes, like the full-potential linearized augmented plane wave (FLAPW [174, 180]) or the projected augmented-wave [181, 182] methods. In essentially all calculation schemes, the introduction of the basis expansion Eq. (55) reduces the system of coupled integro-differential Eq. (19) or (21) to a generalized eigenvalue problem

$$\sum_{\mu} C_{\alpha\mu}\left[\int \chi_\eta^*(x)\, \hat{h}\, \chi_\mu(x)\, dx - \varepsilon_\alpha \int \chi_\eta^*(x)\chi_\mu(x)\, dx\right] = 0 \tag{57}$$

where $\mathcal{H}$ is the operator acting at the function $\varphi_\alpha(\mathbf{r})$ or $\psi_\alpha(\mathbf{r})$ on the left side of Eq. (19) or (21), correspondingly. With small technical variations (e.g., resulting from the energy dependence of basis functions in linearized schemes [183, 184]) the remaining technical problem reduces to the evaluation of matrix elements of the Hamiltonian $\mathcal{H}$ and of the overlap and the diagonalization. The problem cannot be reduced to a single diagonalization if basis functions depend on energy, as, for example, in the Korringa–Kohn–Rostoker method. After solving the matrix equations, one can calculate the electron density, as in Eq. (20), and the total energy, which are the basic characteristics of the ground state. A number of other properties (spin density, forces on atoms, and vibrational frequencies) may be calculated as well.

The physical questions that are of interest in the study of molecular magnets are not intrinsically different from those encountered in the study of magnetism and electronic structure of, say, bulk solids, surfaces, of clusters from first principles in the density-functional theory. One is interested in a description of the ground-state electronic structure and, as far as possible, of the lowest excitations, in terms of Kohn–Sham eigenvalues and the corresponding charge and spin density. It is advantageous to have access to sufficiently accurate total energies for comparing different competing charge or spin configurations; moreover, forces could be needed to perform conjugate-gradient structure optimization or simulation of vibrations. These requirements are quite common in the practice of DFT calculations. The simulation of molecular magnets presents, however, certain technical difficulties that are not necessarily typical for all DFT applications and impose limitations both on the choice of the computational code for an efficient use and on the number of systems addressed so far in a first-principle simulations. These difficulties are

1. Large number of atoms, up to several hundreds of atoms per repeated structural unit
2. Low space group (or, point group) symmetry—or none at all—that does not allow methods that use (l, m) expansions in spherical harmonics (e.g., FLAPW) to profit from efficient block diagonalization
3. Typically, a large size of a simulation box and, on the average, low density of atoms, which makes planewave methods with a global basis set cutoff inefficient
4. The presence of transition metal, or even rare earth, atoms with deep core states and sometimes important semicore states, together with the rest of predominantly light organic atoms. This may create difficulties for the use of norm-conserving pseudopotentials
5. In tight-binding methods with fixed basis sets, specific problems may arise as a result of the need to tune and optimize the basis, as charge configurations and spatial distribution of density in molecular magnets may differ from those with which one is acquainted in crystalline compounds
6. The lack of energy dispersion (resulting from very week coupling between molecular units) and quite commonly a dense spectrum of nearly degenerate discrete states in the vicinity of HOMO-LUMO gap, which makes the self-consistency slowly convergent or even unstable.

Retrospectively, it seems understandable that a large number of calculations done so far employed one or another scheme using flexible tight-binding bases. Pseudopotential planewave calculations are not much represented, although one may expect an increase of their fraction, particularly with the use of ultrasoft pseudopotentials, in the future. Other all-electron methods (FLAPW) were used only for benchmark calculations on simplified systems. One can also anticipate a certain impact of basis-free, purely numerical approaches in the future.

In the following text, we critically compare several families of methods that played, or that are expected to play, an important role in DFT calculations on molecular magnets, and emphasize several representative results.

5.1. Tight-Binding Linear Muffin-Tin Orbital Methods

The tight-binding linear muffin-tin orbitals (TBLMTO) [184–186] method has been used by the Ekaterinburg group for calculations of electronic structure and interaction parameters

of Mn_{12}-ac [166] and V_{15} [187]. The computational method used in these works was indeed TB-LMTO, and not the LMTO method in the less accurate "orthogonal approximation" [188], as erroneously claimed in these publications (D. Boukhvalov, private communication). This calculation used the real crystal structure of molecular crystal and periodic boundary conditions. The interatomic exchange parameters J were estimated along Eqs. (40) and (41).

Having the advantage of a compact and flexible (numerical and adjustable in the course of calculation) basis set, the LMTO method faces difficulties in the treatment of loosely packed structures, as it employs space filling by atomic spheres or "empty spheres," which in crystals with large and low-symmetric cavities is a cumbersome and ambiguous procedure. There are further drawbacks of LMTO for the treatment of molecular magnets. First, the method always employs periodic boundary conditions, so that molecular units must be posed either in their true (and very diffuse) crystalline arrangement or, to simulate them as isolated entities, with substantially enlarged lattice parameters. Second, the method has limitation of only one principal quantum number per l value in the basis set (i.e., $3p$ and $4p$ states cannot be simultaneously present in the valence band). These deficiencies are known to degrade the delicate results of calculation, such as the placement of some bands, or their dispersion in solids. Therefore, one should access the quantitative results of these LMTO calculations with care. Possible indications of inferior numerical accuracy are the total magnetic moment of the Mn_{12}-ac system, which is $19\mu_B$ in the LDA, at variance with experiment and other calculations [124, 172, 173, 189], yielding $20\mu_B$, as well as the absence of HOMO-LUMO gap in both Mn_{12}-ac and V_{15}, again in variance with calculations by different methods. One should note, however, that the overall shape of local DOS is consistent with results of other calculations.

Boukhvalov et al. [166, 187] emphasize the importance of intraatomic correlation in the description of magnetic interactions and excitation spectra of Mn_{12}-ac and V_{15}. This might well make sense, as evidenced by rich experience on this subject from manganites and vanadates, where the local coordination of transition metal ions and electronic structure is somehow similar to that in molecular magnets. The intraatomic correlation may be brought into the calculation by means of the LDA + U approach [153], depending on the *ad hoc* choice of an average Coulomb parameter U. There are certain arguments for the choice of this parameter in the papers given, $U = 4$ eV for V_{15} and $U = 8$ eV for Mn_{12}-ac. Not less important than the actual results with these parameter values are the trends with varying U, which have been reported for Mn_{12}-ac. One finds that as U changes from 4 to 6 to 8 eV, the exchange interaction parameters between the four inner Mn atoms of the cubane core vary from 37 to 33 to 30 K, respectively (other Mn–Mn interaction constants, on the same order of magnitude, change in a similar manner). Moreover, the local magnetic moments on all Mn atoms become slightly enhanced, and the band gap increases from 1.35 to 1.78 to 2.01 eV. These trends follow from the qualitatively transparent fact that higher U values shift the occupied $3d$ states down in energy and lift up the unoccupied ones, thus increasing the band gap. As spin-flip excitations across the gap become more difficult, and they contribute to the nonlocal susceptibilities [the denominator in Eq. (41) increases], this has the effect of reducing the interatomic exchange interaction. This mechanism will further be discussed in Section 6.

5.2. Gaussian-Type Orbital Methods: Naval Research Laboratory Molecular Orbital Library

Methods based on the linear combination of atomic orbitals with a basis of Gaussian-type orbitals, have often been used, and there are several "flavors" that are known.

This approach is a "full-potential" one in the sense that no muffin-tin or atomic sphere geometry is imposed, and the spatial form of the potential is general. In particular, the version implemented in the Naval Research Laboratory Molecular Orbital Library (NRL-MOL) code [190–192] has been frequently used in calculations on molecular magnets.

The NRLMOL program package developed by Pederson, Jackson, and Porezag is an all-electron Gaussian-type orbital implementation of DFT [190, 191, 193–200], and it has been applied successfully to calculate the electronic and magnetic properties of several molecular

nanomagnets [41, 172, 189, 201–212]. By including the spin-orbit coupling, it is possible to calculate the magnetic anisotropy energy, which is a crucial parameter for understanding the magnetic behavior of SMMs. The agreement between experiment and the result from the first-principles calculation is in many cases surprisingly good. Therefore, it seems to be suitable to give some details on this particular numerical implementation.

The molecular orbitals are expanded as linear combinations of Gaussian functions centered at the atomic sites; multicenter integrals are evaluated numerically on a specially generated variational integration mesh (see Pederson and Jackson [190] for details). An efficient parallelization [200] makes all-electron calculations with more than hundred atoms feasible in affordable time, a prerequisite for useful applications in the domain of SMMs. The problem of basis optimization, a severe and common problem in all methods employing localized and fixed basis functions, is solved in NRLMOL by tuning to the solutions of self-consistent isolated atoms [199].

The self-consistent potentials, obtained numerically, are least-square fitted to the sum of bare spherical Gaussians or Gaussian-screened $1/r$ potentials to facilitate multicenter integrations.

Given the basis sets and the Gaussian-representation of the atomic potentials, it is possible to obtain very good insight into the class of multicenter integrands that need to be integrated, and this information is used to generate a numerical variational integration mesh [190] that allows to precisely determine integrals required for calculation of secular matrices, total energies, and derivatives according to

$$I = \int d\mathbf{r} Q(\mathbf{r}) = \sum_i Q(\mathbf{r}_i)\Omega_i \tag{58}$$

where Ω_i is the volume associated with point $\mathbf{r}_i$. Errors arising from the numerical integration can easily be checked and controlled by adjusting a few parameters that control the mesh construction. It should be emphasized that the Gaussian-screened potentials are only used to optimize the numerical quadrature schemes used for mesh generation.

Once self-consistency is achieved, the forces acting on each atom are determined from the Hellmann–Feynman–Pulay theorem [213–215]. After obtaining all the forces acting on all the atoms, a conjugate-gradient method, or other force-based algorithms, can be used to carry out geometry optimizations. Once the equilibrium geometry and the Kohn–Sham wavefunctions are obtained, the properties available for the analysis include (beyond the standard set provided by any DFT package) polarizabilities, vibrational frequencies, infrared and Raman spectra, and magnetic anisotropy energies.

For the $[Mn_4O_3Cl_4(O_2CCH_2CH_3)_3(NC_5H_5)_3]$ system, containing as its core a Mn_3^{3+} Mn^{4+} pyramid and possessing a magnetic moment of $9\mu_B$ per unit (Mn^{3+} spins are antiferromagnetically coupled to Mn^{4+}), Park, Pederson, and Hellberg [216] calculated the properties related to dimerization. The Mn_4 units were presumed to couple antiferromagnetically, based on their unusual quantum tunneling properties [71], which was now confirmed in a calculation by NRLMOL [217]. A fit to the Ising model yields intraatomic exchange parameters of 44 K (ferromagnetic; Mn_3^{3+}–Mn^{3+}) and -152 K (antiferromagnetic; Mn_3^{3+}–Mn^{4+})—both of which are overestimated by roughly a factor of two in comparison with experiment-derived values. The intermolecular coupling of merely -0.24 K is also twice as large as the experimental fit value. In addition to structure relaxation, Park et al. [217] performed a calculation of vibrational spectra with infrared and Raman intensities—the data not yet available from experiment but extremely important for identification and further characterization of this molecular magnet.

5.3. Numerical Atom-Centered Basis Functions: Siesta

The Siesta method and computational code [218, 219] also uses compact atom-centered basis functions, but (differently to NRLMOL) it employs numerical ones with strict spatial confinement as the most frequent choice. (Gaussian-type orbitals, or other fixed functions at the user's convenience, are equally available for the basis set). Because of the strict confinement of the basis functions, the program can make a clean distinction between cases of an isolated

fragment (molecule or cluster), "chain," "slab," or "crystal" (with periodic boundary conditions in one, two, or three dimensions, correspondingly), and correctly construct Madelung terms according to each case. Keeping track of the local neighborhood in the calculation of matrix elements, in combination with order-N facilities (see, e.g., [220, 221]), makes Siesta a great method for treating large, low-coordination, low-symmetry structures, as molecular magnets are. In contrast to NRLMOL, which determines the Coulomb and exchange-correlation potentials analytically from the Gaussian representation of the wavefunctions, Siesta employs fast Fourier transform of the residual charge density (after subtraction the dominant atom-centered contributions) for the solution of the Poisson equation, which also yields high (and controllable) accuracy, needed especially in the calculation of forces and optimization of structure. Moreover, for periodic systems (as molecular magnets generally are, in a crystalline state) the components of the stress tensor are calculated and can be used for simultaneous optimization of lattice parameters and internal coordinates subject to target pressure. Particularly important for magnetic systems is the option of treating the noncollinear (i.e., not diagonal in the spin space) density matrix, that allows us to simulate deviations of local magnetic moments from the global magnetic axis (for a recent application, see Postnikov et al. [222]). Different from the two above-discussed methods, Siesta is not an all-electron method but employs norm-conserving pseudopotentials (Ref. [223], among other choices) and allows us to apply the core correction after Louie et al. [224]. As the basis set consists of localized functions and not planewaves, the use of hard pseudopotentials, such as those for transition metals (also "small core," with semicore states attributed to the valence band) or oxygen, is not problematic. Siesta was designed in view of large distorted systems and dynamical simulations therein, so that properties of space group (or point group) symmetry are essentially lost. Therefore, no special treatment of symmetrized molecular orbitals is provided.

As with other pseudopotential methods, Siesta in its present version requires some care in choosing and testing pseudopotentials before calculation and moreover, in choosing basis orbitals. A certain freedom in the tuning of the latter is more a matter of experience than of consistent control in the variational procedure (as it is the case with planewave cutoff). Although being, as a rule, reasonably workable, such settings are difficult to improve consistently. More insight into the problem of basis sets was provided by Sánchez-Portal et al. [225] and Junquera et al. [226].

The application of Siesta to molecular magnets is relatively new. We outline some recent results below.

5.4. Discrete Variational Method

The discrete variational method (DVM) [227, 228], one of the earliest density-functional theory schemes to find applications in chemistry, seems to be potentially very well suited for the studies of molecular magnets. The method is an all-electron one, it uses basis of numerical atomic orbitals, and the three-dimensional integration over the space outside the spheres circumscribing core regions of each atom is done on a pseudorandom numerical grid. DVM was used quite early for first-principles calculations of the electronic structure of a large molecular magnet such as the 10-member "ferric wheel" albeit in a simplified form, excluding H and Cl atoms; see Ref. [123]. The method was also applied in one of the first *ab initio* studies of the Mn_{12}-acetate [124]. Apart from discussing charge states, magnetic moments, and local DOS of the three distinct groups of Mn and O atoms in the molecule, which largely remained uncontested by subsequent calculations, Zeng et al. [124] estimated Heisenberg exchange parameters in the magnetic transition state scheme [159], an extension of Slater's original transition state *unsatz*, through a procedure outlined in Section 4. Flipping the spin at one atom and detecting the shift of the $3d$ energy level at another one as result of induced magnetic polarization helps us to arrive at a system of equations in which different interatomic exchange parameters are coupled. For the sake of simplicity and the clearness of analysis, only collective (non-symmetry-breaking) spin flips on all atoms belonging to each set of Mn atoms—Mn(1) in the inner cube and, Mn(2) and Mn(3) in the peripheral region (see Fig. 1)—were allowed in the analysis of Zeng et al. [124] This means that four spins within

Table 1. Electronic structure parameters (magnetic moments and Heisenberg exchange parameters) of Mn_{12}-ac from first-principles calculations.

	Magnetic moments (μ_B)			Exchange parameters (K)		
Method	Mn(1)	Mn(2)	Mn(3)	J_{12}	J_{13}	J_{23}
DVM[a]	3.056	−3.889	−4.039	−136	−72	−102
NRLMOL[b]	2.57	−3.63	−3.58	−57	−41	−8
LMTO[c], U=4 eV	2.72	−3.44	−3.65	−53	−47	−19
LMTO[c], U=8 eV	2.92	−3.52	−3.84	−47	−26	−7

[a] [124]; LDA
[b] [173]; GGA; moments within a sphere of 2.5 Bohr, J values by Park et al. [216].
[c] [166]; LDA + U; moments within spheres of 2.7/2.8 Bohr (inner/outer Mn atoms)

each group always remained rigidly ferromagnetically coupled. This resulted in a system of three equations, from which the values of J_{12}, J_{23}, and J_{13} could have been determined. The density-functional theory results were explicitly fitted to the Heisenberg Hamiltonian of the form in Eq. (5). However, the parameters J_{11} and so on, representing the coupling within each group, did not appear in the fit because the spin excitations necessary to probe them, which would break the symmetry of the molecule, were not allowed. Their inclusion in an otherwise executed calculation could result in a renormalization of exchange parameters.

The values of J_{12}, J_{23}, and J_{13} are given in Table 1: they are all negative (i.e., they indicate an AFM coupling, as could be expected because of a more-than-90 degree superexchange pathway through bridging oxygens), and hence cause frustration when accommodating the three spin subsets.

5.5. Planwave Methods

The use of planewave basis for calculation on molecules is, as already mentioned, computationally inefficient, but technically feasible and, with sufficiently high cutoff, also ultimately accurate. Massobrio and Ruiz [229] compared recently straightforward (from the total energy difference in low-spin and high-spin configuration) estimates of Heisenberg-model exchange parameters J for several Cu-based binuclear molecules: $Cu_2(CH_3COO)_4$, $[Cu_2(\mu\text{-}OH)_2(\text{bipyrimidine})_2](NO_3)_2 \cdot 4H_2O$, and $[(dpt)Cu\ (\mu\text{-}Cl)_2Cu(dpt)]Cl_2$($dpt$ = dipropylenetriamine), using identical norm-conserving pseudopotentials and an exchange-correlation scheme (among other, differing options) with Gaussian-type basis functions. The largest system consisted of 62 atoms, a moderate number by the standards of a calculation with localized basis functions. For the computational load with the planewave basis, however, it is the size of the simulation box that primarily matters. Here its linear size of 18.5 Å resulted in about 2.4×10^6 plane waves for the expansion of the charge density and demanded hours of CPU on parallel computers. The small values of J obtained in the planewave calculation (-518, -95, and $+61$ cm^{-1}, correspondingly) were of correct sign and order of magnitude in all cases, although deviations in absolute value, from experimental estimates and between different exchange-correlation flavors, were up to 50%.

6. DISCUSSION OF SELECTED SYSTEMS OF CURRENT INTEREST

In the following text we outline some recent results on relatively "new" molecular magnets (i.e., systems that have only become available during the last few years). For the study of their electronic characteristics, several questions arose that our calculations attempted to clarify. "Ferric wheels" gained interest, and not in the last place, because of their "esthetically rewarding" [230] shape. Two examples discussed below have an AFM ground state; as a consequence, they might find applications related to quantum tunneling and quantum computing. Other structurally similar (although chemically different) examples include $3d$ ions (notably Mn) at larger distances, with magnetic interactions mediated by organic radical groups that lead to strong antiferromagnetic couplings of Mn ions. An example

is the molecule $[Mn(hfac)_2(NITPh)]_6$(hfac=hexafluoroacetylacetonate, NITPh=2-phenyl-4,4,5,5-tetramethyl-4,5-dihydro-1*H*-imidazol-1-oxy-3-oxide) (see Gatteschi and Pardi [230]), with a net spin of $S = 12$. "Ferric stars" include a central 3*d* ion to which peripheric ions couple antiferromagnetically, resulting in a net spin value in the ground state. Because of their non-negligible magnetic anisotropy, these systems look like possible prototypes for magnetic storage. The molecule Ni_4 is a seemingly simple magnetic molecule for which a fit of experimental data of magnetization versus magnetic field to the Heisenberg model fails quite dramatically, and possible reasons for deviation have been studied with the help of first-principles calculations. Finally, we consider a two-nuclei model system with the aim of studying the effect of intraatomic correlation ("Hubbard U") on the electronic structure and interatomic magnetic interactions in a more numerically accurate calculation than has yet been accomplished (for Mn_{12} by TB-LMTO, see Boukhvalov et al. [166]). In the most recent case, the calculations have been performed with the FLAPW method [231], for other systems, by methods using atom-centered localized basis functions, they have been performed with either Siesta or NRLMOL.

6.1. Ferric Wheels

Hexanuclear ferric wheels $MFe_6[N(CH_2CH_2O)_3]_6Cl$ (M = Li, Na; see Fig. 11), the systems to be discussed below, were synthesized at the Institut für Organische Chemie in Erlangen [78] and labeled as substances 4 and 3 in later publication. There exist a large family of ferric wheels with a different even number (N = 6, 8, 10, 12, 18) of iron atoms [78–80, 121, 232–238]. In addition to the ferric ones, there have been reports on wheels with other transition metal ions such as an eight-membered Cr(III) wheel [239], a Cu(II) [240, 241], a Co(II) [242], a Mn(II) [243], and a 24-membered Ni(II) wheel [244]. The latter structure contains the largest number of transition metal ions in a wheel-like structure so far. The synthesis of odd-numbered magnetic wheels or necklaces appears to be a nontrivial task.

Fe atoms in these compounds are connected by oxo-bridges that are reminiscent of the 90 degree coupling of magnetic atoms in TM oxides. The nearest coordination of the Fe atom is octahedral—two pairs of O ions form bridges to the neighboring Fe atoms on both sides—and the fifth oxygen (referred to below as "apical") and the nitrogen ion are connected by the C_2H_4 group. The octahedra are slightly distorted to accommodate the stiffness of oxo-bridges with the curvature of the molecular backbone. Although the Fe–O–Fe angles differ slightly in the Li-centered and Na-centered wheels (101.1 and 103.3 degrees, respectively), the structure of the two molecules is almost identical.

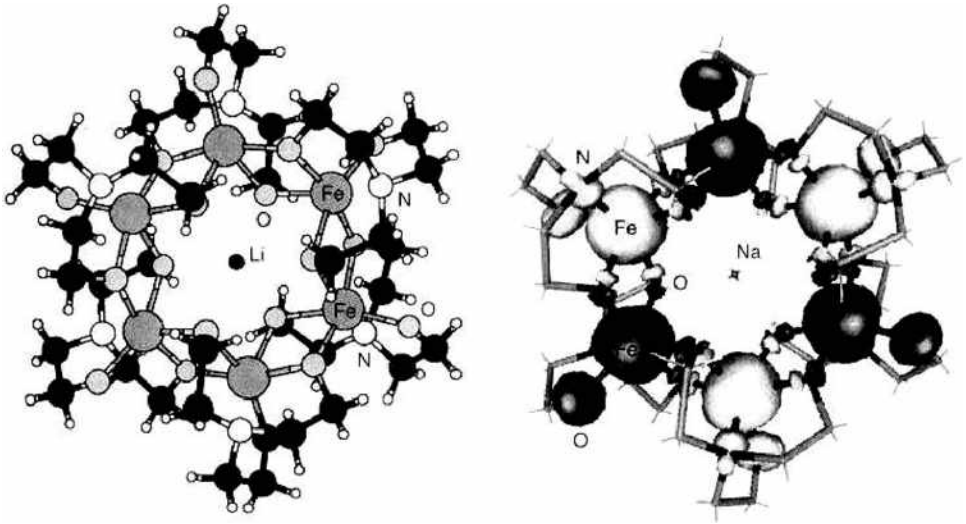


Figure 11. Structure and spin density distribution in ferric wheel molecules. Left panel: two views of the Li-centered molecule. The Li ion is in the middle of the ring; the distant Cl ion included in the simulation is not shown; the rest of (electrically neutral) solvent is neglected. Right panel: iso-surfaces correspond to $\pm 0.01 e/Å^3$, according to NRLMOL calculation [245]. Although most of the magnetic moment is localized at the Fe atoms, there is still some spin polarization on O and N.

According to magnetization and torque measurements by Waldmann et al. [79], these systems are characterized by $S = 5/2$ on the Fe site, thus implying a highly ionized Fe(III) state. Moreover, a fit to the spin Hamiltonian of the Heisenberg model in Eq. (5) yields the J values of -18 to -20 K for the Li-wheel (depending on sample and method) and -22.5 to -25 K for the Na-wheel, thus implying an AFM ground state [79]. X-ray photoelectron and x-ray emission spectroscopy studies [245] allowed for probing of the electronic structure in the valence band and on the Fe site, albeit without resolution in spin. Whereas the magnetic measurements data are by now well established, the spatially resolved distribution of magnetization was not yet accessed before the present calculation. Specifically, we compare the results of electronic structure calculations by two different methods within the DFT, Siesta, and NRLMOL (see the discussion on the methods in Section 5). In both cases, we used the generalized gradient approximation after Perdew, Burke, and Ernzerhof [148]. We emphasize that the most important difference between the two methods, with regard to this study, is that Siesta uses norm-conserving pseudopotentials, whereas NRLMOL implements an all-electron method. For an *ab initio* pseudopotential code such as Siesta, benchmark calculations, based on the very accurate NRLMOL suite of codes, help to assess the accuracy of pseudo-potential-based methods in some critical cases or for new systems.

We outline below the results obtained by Siesta for the Li-centered molecule and by NRLMOL for the Na-centered one, as presented in more detail by Postnikov et al. [246]. The NRLMOL treatment was restricted to the ground-state AFM configuration (alternating orientations of Fe magnetic moments over the ring); the Siesta calculation addressed, in addition, different magnetic configurations, which allowed for the extraction of density-functional theory-based exchange parameters.

Figure 12 displays the partial densities of states (DOS) on Fe and its several neighbors in the AFM configuration, as calculated by both methods. The discrete levels of the energy spectra are weighted (with the charge density integrated over atom-centered spheres in NRLMOL, or according to Mulliken population analysis in Siesta) and broadened for presentational purposes with the broadening parameter of 0.15 eV (Siesta) and 0.14 eV (NRLMOL). The local moments corresponding to integrating such partial DOS over occupied states are given in Table 2. Both calculations give a consistent description of state densities at Fe and O sites, even though this property is rather loosely defined (and its calculation differently implemented in Siesta and NRLMOL).

Notably, both methods find the local magnetic moments on Fe sites very close to 4 μ_B and not to 5 μ_B, as is generally assumed, based on the above-mentioned magnetization data. The maximal magnetization $S = 5/2$ of the Fe atom corresponds to a Fe(III)-ion with in $3d^5_\uparrow d^0_\downarrow$ configuration. Our first-principles calculations indicate a somewhat different picture: the minority-spin DOS has a nonzero occupation because of the hybridization (chemical bonding) of Fe3d with O2p states. However, the magnetic polarization caused by Fe in the organic ligand that provides the octahedral coordination for the iron atoms is substantial, with the most pronounced effect being on the apical oxygen atom (which is not participating in the bonding to the next Fe neighbor). Taken together with the (smaller) polarization of the bridging oxygen atoms and magnetization at the nitrogen site, the distributed magnetic moment per Fe atom yields 5 μ_B, recovering the agreement with the magnetization results.

A clear visualization of the above-discussed delocalized (or, rather, distributed) magnetic moment associated with the Fe atom comes from the map of spin density, obtained from the NRLMOL calculation (Fig. 11, right panel). One should take into account that the volume enclosed by the iso-surfaces is not directly correlated to the total moment at the site. One sees, moreover, an absence of magnetization on carbon and hydrogen sites. The fact that the magnetization is noticeable and changes its sign when passing through bridge oxygen atoms emphasizes the failure of methods depending the spherical averaging of atom-centered potentials.

An important consequence is that the charge state of iron is not Fe(III) but closer to Fe(II), according to our calculations. Moreover, the distributed magnetic moment behaves like a rigid one, in the sense that it can be inverted, following a spin flip on a Fe site. This is illustrated by the analysis of other magnetic configurations, done with Siesta [246]. The local DOS does not change considerably when switching from AFM to FM configuration—only

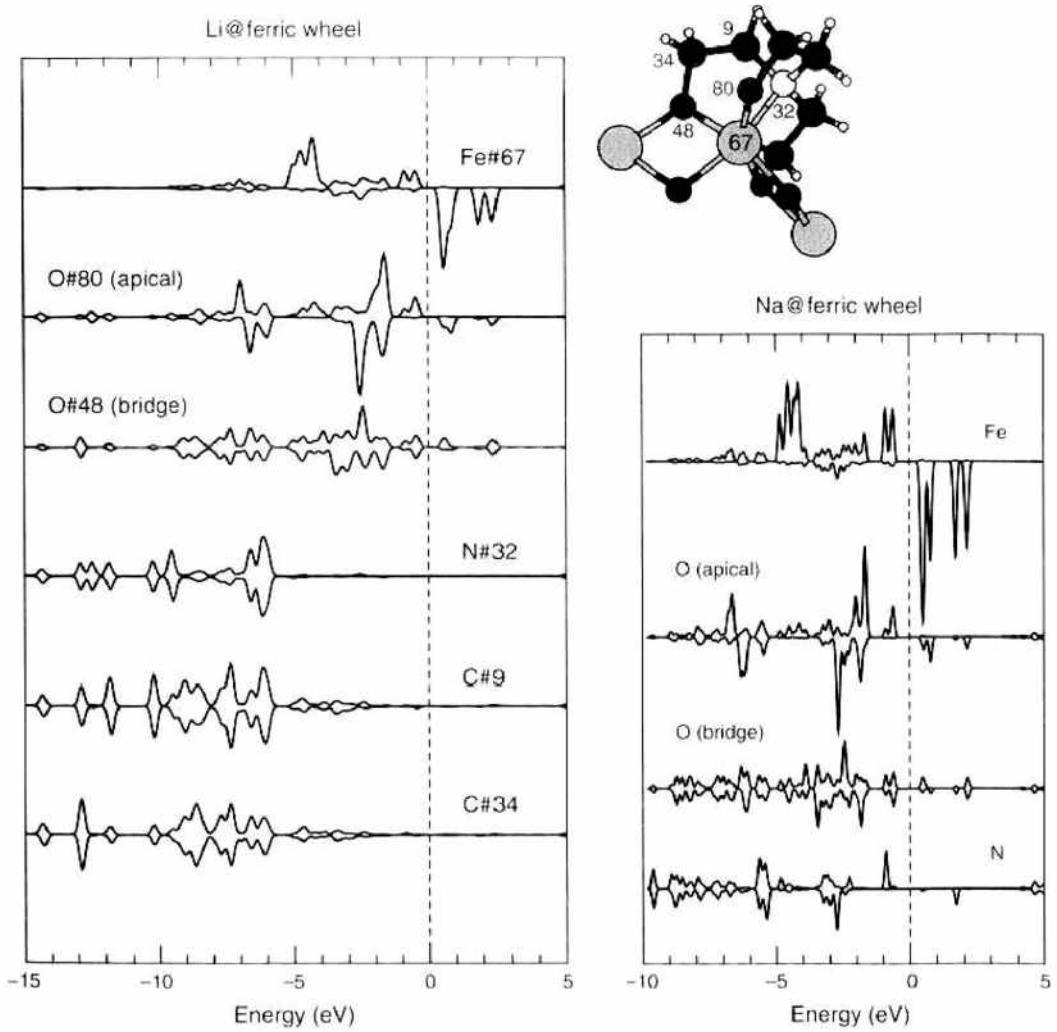


Figure 12. Atom- and spin-resolved partial densities of states as calculated for Li-centered molecule by Siesta (left panel) and for Na-centered molecule by NRLMOL (right panel). The DOS at the Fe site is scaled down by a factor of two relative to other constituents. The numbering of atoms that are neighbors to the Fe atom is shown in the inset. See text for details on the calculation.

the HOMO/LUMO gap becomes less pronounced, and a slight FM shift appears between the two spin bands.

For the sake of improving both the stability of convergence with Siesta and for pinning down a particular spin configuration (FM, or with one or more Fe magnetic moments inverted), we applied the FSM scheme [158] in the calculation. Imposing an (integer) spin moment per molecule fixes the number of electrons in the two spin channels and removes the possibility of spin flips, which are a major source of numerical instability, as there are many nearly degenerate states in the vicinity of the Fermi level in the molecule (and no symmetry constraints on these states in Siesta). The FSM procedure would normally split the common chemical potential into two separate ones, for majority- and minority-spin channels, which

Table 2. Local magnetic moments M at Fe and its neighbors.

Atom	R(a.u.)	$M(\mu_B)$, NRLMOL	$M(\mu_B)$, Siesta
Fe	2.19	3.85	3.91
O (apical)	1.25	0.20	0.30
O (bridge)	1.25	±0.01	±0.02
N	1.32	0.07	0.09

Note: NRLMOL results correspond to spin density integrated over sphere of radius R centered at corresponding atom; Siesta values result from Mulliken population analysis.

corresponds to an effective external magnetic field, and hence to an additional (Zeeman) term in the total energy, in analogy with Eq. (7). Because molecular magnets possess a HOMO-LUMO gap, the latter correction must only be considered if such gaps in two spin channels do not overlap.

Figure 13 shows the total energy values and energy gaps for FSM values of 30 μ_B (FM case), 20 and 10 μ_B (one and two local moments inverted, correspondingly), and 0 (alternate-spin AFM case). A linear change of the total energy while inverting one and then two local moments from the FM configuration is what would be expected from the Heisenberg model with "rigid" magnetic moments (in the sense that their S values do not depend on the total spin of the system), assuming moreover that only nearest-neighbors interactions between spins are important. An additional justification of the validity of the Heisenberg model comes from the observation that the magnitudes of local magnetic moments on the Fe atoms always remain close (within several percent) to 4 μ_B, and the partial DOS on Fe sites remains largely unaffected by the actual magnetic ordering. Similarly unaffected is a pattern of local magnetic moments at O and N neighbors of a particular Fe atom, always getting inverted as the latter experiences a spin flip. Keeping this in mind, and assuming Heisenberg-model spin Hamiltonian as in Section 1 with the S value of 5/2 (i.e., for the total spin that gets inverted), we arrive at the estimate for $-J$ of around 80 K (over both $30 \rightarrow 20$ and $20 \rightarrow 10$ μ_B flips). This is qualitatively correct (i.e., it indicates a preference toward AFM coupling) and even of the correct order of magnitude.

However, two observations can be made here. First, the "true" AFM configuration (with half of magnetic moments inverted on the ring) does not follow the linear trend (see Fig. 13) and lies actually higher in energy than the configuration with two spins inverted. The origin of this deviation is not yet clear to us at the moment. There are several possibilities: the zero-FSM configuration is, technically, the most difficult to converge, so some numerical instability can still play a role. In contrast, a true (mixed) quantum-mechanical ground state of a system with six coupled $S = 5/2$ spins may win over both our density-functional theory solutions that correspond to selected values $S_z = 0$ or $S_z = 5$ of the total spin. Moreover, the necessity to include magnetic interactions beyond first neighbors, not considered at the moment, might further complicate the situation. The second observation concerns the magnitude of the exchange parameter J and the fact that it is probably overestimated by a factor of ~ 4 in our calculation. The origin of this overestimation lies most probably in on-site correlations, which, if treated accurately beyond the standard schemes of the DFT, would primarily affect localized Fe $3d$ states, shifting the bulk of occupied states downward in energy, and the bulk of unoccupied states upward, expanding the energy gap, and—whatever scheme is used for estimating exchange parameters—substantially reducing their magnitude. This has been

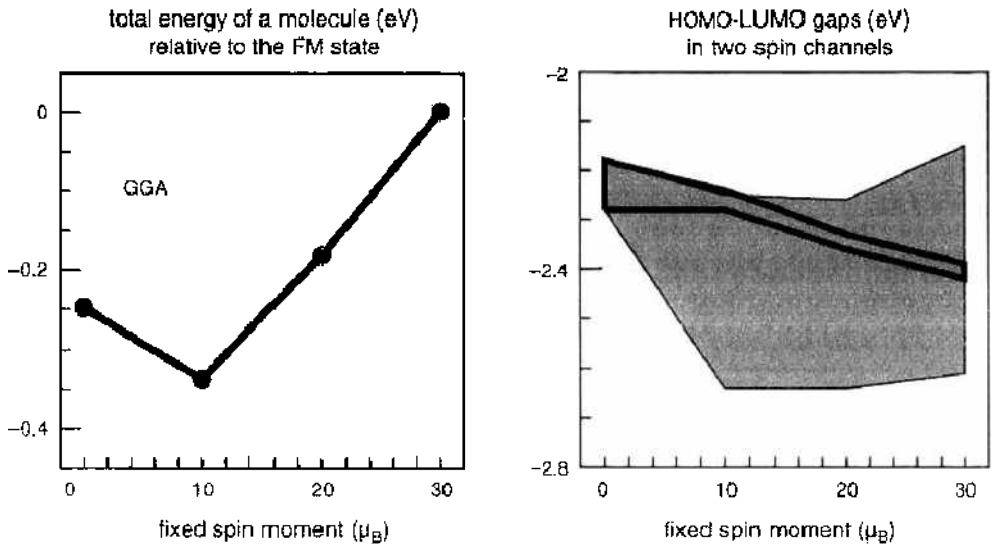


Figure 13. Total energy per molecule (left panel) and HOMO LUMO gap in two spin channels (right panel: shaded area=majority spin, thick lines=minority spin) from fixed spin moment calculations.

recently shown for another molecular magnet (Mn_{12}) by Boukhvalov et al. [166] (see the earlier discussion on Mn_{12} above and our following analysis of a model binuclear system).

Summarizing our analysis of the electronic structure of Li- and Na-centered ferric wheels, one can conclude that the local magnetic moments on Fe sites seem to be 4 μ_B, rather than 5 μ_B, as is often assumed. This implies a valence state closer to Fe(II) than to Fe(III), with a substantial covalent part in the Fe–O bonding. The local spin of $S = 5/2$ per iron site consistent with magnetization measurements is, however, recovered if one takes the magnetization of neighboring atoms into account. The ability to calculate Wannier functions in such systems may provide much more reliable estimates of projected moments than are currently offered by either Mulliken methods or methods based on moments within a sphere. The largest moment is on the apical oxygen atom, followed by smaller moments on nitrogen and the bridging oxygen atoms. This picture is well confirmed by a spatial distribution of spin density.

With respect to its magnetic interactions, this system can be mapped reasonably well onto the Heisenberg model; hence, we deal with *rigid* magnetic moments that are nevertheless delocalized—an interesting counterexample to the common belief that the Heisenberg model primarily applies to localized spins.

6.2. Ni_4

"Ni_4" is a shorthand notation for a molecular crystal $[Mo_{12}O_{30}(\mu_2\text{-}OH)_{10}H_2\{Ni(H_2O)_3\}_4] \cdot 14H_2O$, synthesized and characterized by Müller et al. [247]. This material crystallizes in a structure containing two formula units (shown in Fig. 14), related by the 180-degree rotation around an edge of the Ni_4 tetrahedron. The Ni–Ni distance is 6.6–6.7 Å, and magnetic interactions are mediated by a longer path than in the systems discussed above.

Magnetic properties are caused by Ni(II) ions in the $3d^8$ configuration ($s = 1$); the ground state is antiferromagnetic. An intriguing aspect of this compound is that the measured zero-field magnetic susceptibility can be very well mapped onto the Heisenberg model, whereas the measurements of magnetization cannot. The inclusion of different anisotropy terms in the Heisenberg model to improve the description of experiment had only limited success [248]. First-principles calculations have been performed using the Siesta method to access the electronic structure and estimate the magnitudes of magnetic interaction parameters.

Similar to the case of the ferric-wheel system discussed above, the FSM scheme was used for probing different spin configurations and comparing their total energies. The local DOS is practically indistinguishable for the cases of zero total moment (the AFM structure, which has indeed, in agreement with experiment, the lowest total energy) and for configurations with local magnetic moments inverted at one or two Ni atoms (yielding, in the last case, the FM configuration). The local moment per atom in these cases agrees with the $s = 1$ estimation derived from magnetization measurements. As it was discussed above for other magnetic molecules, the magnetic moment is not fully localized on the Ni ion; small but nonnegligible magnetization is induced on neighboring oxygen atoms, and even on more distant Mo atoms (Fig. 15, left panel). As the Ni–Ni interaction path is much longer than in other, earlier-discussed magnetic molecules (see inset in Fig. 15), the energy differences

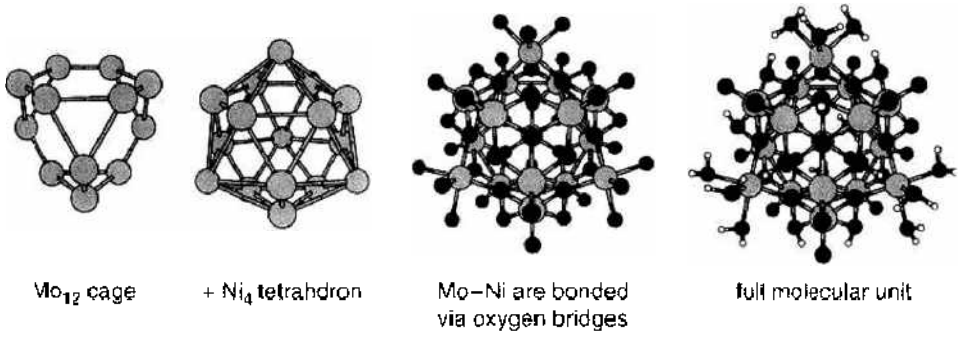


Figure 14. Buildup of the "Ni_4" molecular unit.

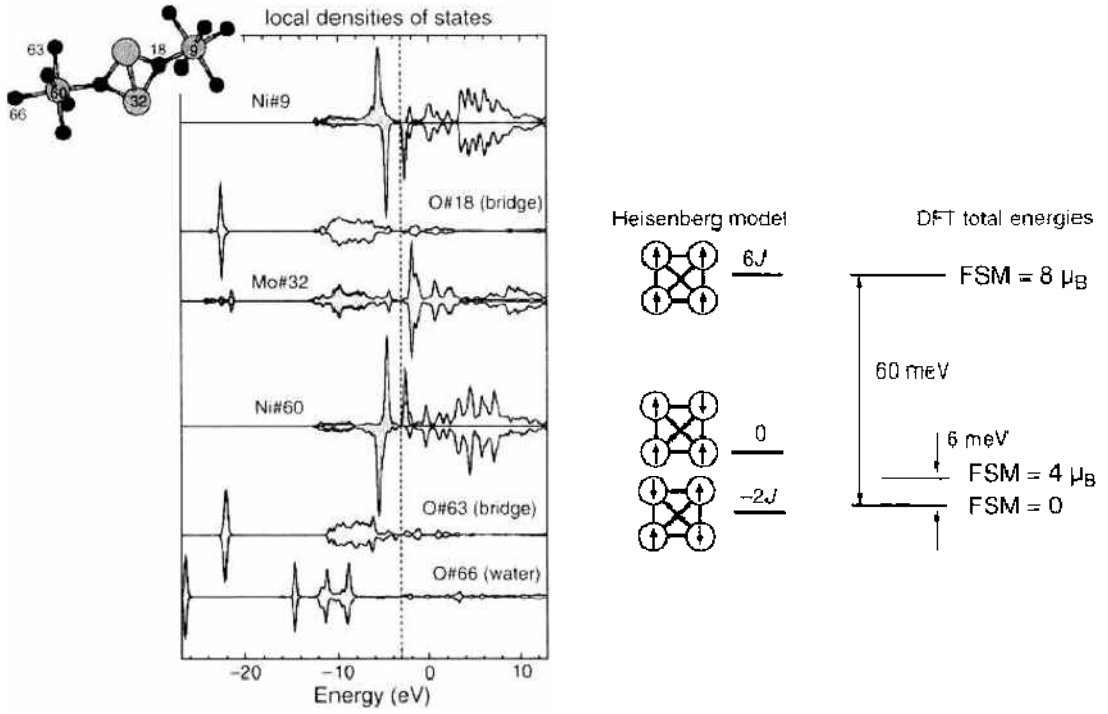


Figure 15. Left panel: local DOS of atoms at the Ni–Ni magnetic path. Right panel: a scheme of energy levels in different spin configurations of "Ni_4" according to the Heisenberg model and from first-principles calculations.

between configurations with FSM values of 0, 4, and 8 μ_B are small. These solutions are separated by other magnetic configurations that can be converged (2 and 6 μ_B) and correspond to a nonmagnetic configuration of one Ni atom, with unchanged and differently coupled $s = 1$ at three others (as schematically shown in Fig. 16, left panel). The energies of these intermediate configurations are substantially higher, and HOMO-LUMO gaps in two spin channels move apart, indicating the necessity for an external magnetic field (hence additional Zeeman energy) for stabilizing these artificial configurations. On the contrary, the three lowest-energy configurations have HOMO-LUMO gaps common for both spin directions (Fig. 16, right panel); therefore, the mapping to the Heisenberg model can be done directly without considering the Zeeman term.

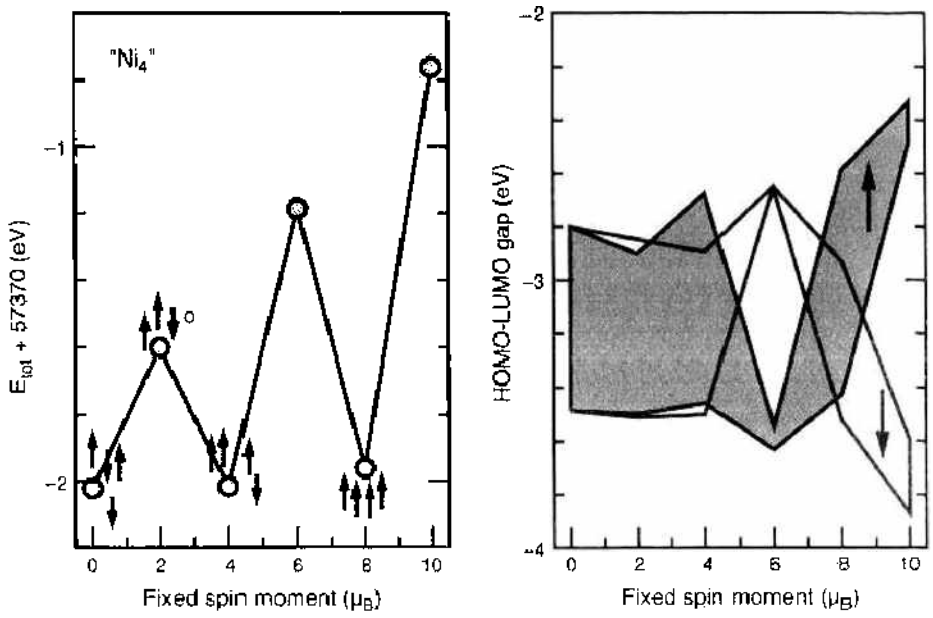


Figure 16. Total energy (left panel) and HOMO-LUMO gap (right panel) from FSM calculations of "Ni_4"

An attempt of such mapping is schematically shown in the right panel of Fig. 15; obviously, the sequence of energies of the configurations with one or two spins inverted (starting from the FM solution) is only in qualitative agreement with the Heisenberg model, but the numerical energy differences do not allow for the evaluation of a unique value of J, in contrast to the case of ferric wheel discussed above. At best, one can make a rough estimate of the order of magnitude of $-J$ that yields 30–90 K.

This failure indicates that the magnetic interactions in "Ni_4" are strongly anisotropic. However, an adequate mapping of first-principles results onto models including the anisotropy would require the inclusion of spin-orbit interaction in the calculation, and this is not yet available in Siesta. This feature is, however, included in NRLMOL, and some of the progress along these lines is outlined below.

6.3. Magnetic Anisotropy in Single-Molecule Magnets

As a modification of Eq. (7), which introduced the anisotropy in the simplest form, we distinguish in the following between axial and transverse anisotropy, with their corresponding parameters D and E. They enter the magnetic spin Hamiltonian (only second-order terms) as

$$\mathcal{H} = DS_z^2 + E(S_x^2 - S_y^2) \tag{59}$$

The values of the axial anisotropy D are available from a number of experiments for different SMMs, and for several SMMs first-principle calculations have been carried out with the use of the NRLMOL code. These results are summarized in Table 3.

In all the cases presented here, the calculated spin ordering is in agreement with experiment. The calculated D parameters for Mn_{12}, Mn_{10}, Mn_9, and the ferric star Fe_4, molecular magnets are in excellent agreement with experimental values. The only remarkable discrepancy is found for Fe_8, a system that seems to pose complications for the density-functional theory treatment. Apparently the density-functional theory may be unable to predict the ground-state density accurately enough because of important electronic correlations beyond the mean-field treatment or missing Madelung stabilization (absent in the isolated system).

The SMMs listed in Table 3 are in general characterized by a high-spin ground-state. However, a high spin state does not necessarily correlate with a high anisotropy barrier. The prefactor D is also very important. To increase the barrier, one has to understand and control D, which will be the main goal of future research in this area. In all cases in which the E parameter is not zero by symmetry, it has been predicted with similar accuracy as D (see Refs. [249–254] and others, listed in Table 3).

The results obtained build confidence in the predictive power of the formalism. It has been already mentioned that a microscopic understanding (based on the electronic structure of SMMs) of the magnetic anisotropy parameters is crucial for the rational design of single-molecule magnets.

Table 3. Comparison of the calculated by NRLMOL and experimental magnetic anisotropy parameter D for the single molecule magnets.

Molecule	S	D(K) Theory	D(K) Experiment
$Mn_{12}O_{12}(O_2CH)_{16}(H_2O)_4$	10	−0.56[a]	−0.56[b]
$[Fe_8O_2(OH)_{12}(C_6H_{15}N_3)_6Br_6]^{2+}$	10	−0.53[c]	−0.30[d]
$[Mn_{12}O_{12}(2,2'\text{-biphenoxide})_{16}Br_{12}]^{4-}$	13	−0.06[e]	−0.05[f]
$Co_4(CH_3C_5H_2N)_4(CH_3OH)_4Cl_4$	6	−0.64[g]	−0.7– −1.9[h]
$Fe_4(OCH_2)_6(C_2H_4ON)_6$	5	−0.56[i]	−0.57[j]
$Cr[N(Si(CH_3)_3)_2]_3$	3/2	−1.15[k]	−2.66[k]
$Mn_{19}C_{12}N_3H_{36}$	17/2	0.33	−0.32[l]
$Ni_4C_{16}Cl6H_{16}$	4	0.385	−0.40[l]
$Mn_{12}Cl_4(O_2CCH_2CH_3)_{16}(NC_5H_5)_4$	9/2	0.58[m]	−0.72[n]

Note: See theory references for computational details.
[a][72, 189]. [b][28, 249]. [c][204]. [d][58]. [e][209]. [f][250]. [g][210]. [h][251]. [i][212]. [j][252]. [k][253]. [l][254]. [m][217]. [n][71].

In the following text, we will discuss some selected recent results of the not-so-well-known single-molecule magnets.

6.3.1. Co_4 Magnet

A new Co-based ferromagnetic SMM with the complete chemical formula Co_4 $(hmp)_4(CH_3OH)_4Cl_2$ (hmp^- is the deprotonated hydroxymethylpyridine), has achieved great interest because of the high anisotropy energy. A simulation by Baruah and Pederson [210] resulted in the prediction of two new, energetically non competitive structural conformations with even higher anisotropy. Specifically, the magnetic anisotropy energy per Co atom was estimated from the experiment to be 25–50 K [255]. More recent experiments on a similar Co_4-cluster have found significantly smaller total anisotropy energies of about 29 K [251], in better agreement with the calculated values of 23, 160, and 50 K for the lowest-energy and for the two higher-energy phases found in the calculation. As already mentioned above, a large magnetic anisotropy is a prerequisite for potential applications of molecular magnets as "microdomains" for magnetic storage. An additional requirement, the existence of a net spin moment, is also satisfied here, with $S = 6$ per molecular unit in the parallel (high-spin) configuration, in all three isomers. Especially given that the earlier calculated results of Baruah and Pederson [210] compare more favorably with the more recent experimental results, a first-principle calculation might guide and stimulate practically relevant experimental research on this promising family of molecular magnets.

6.3.2. Fe_4-Star

This material (of which several analogs with different central atoms are known by now) realizes a net spin moment in a relatively compact and highly symmetric molecule as a result of AFM coupling of the peripheric Fe atoms to the central one. The structure of the Fe_4 ferric star is shown in Fig. 17. All iron atoms are in the Fe^{3+} state, and the resulting ferrimagnetic arrangement has total $S = 5$. Similar to the other molecular magnets, only those states within an energy window of about 5 eV around the Fermi level contribute to the magnetic anisotropy. The symmetry of the cluster allows for a rhombohedral E contribution to the spin Hamiltonian. The theoretical value ($|E| = 0.064$ K) obtained using the experimental geometry [252] as a starting point is in good agreement with the experimental one ($|E| = 0.056$ K) (the sign of E depends on the definition of the axis). This agreement is relatively stable with respect to geometry changes. Total anisotropy barriers change normally only by a few degrees K at most, although in some cases the agreement between theory and

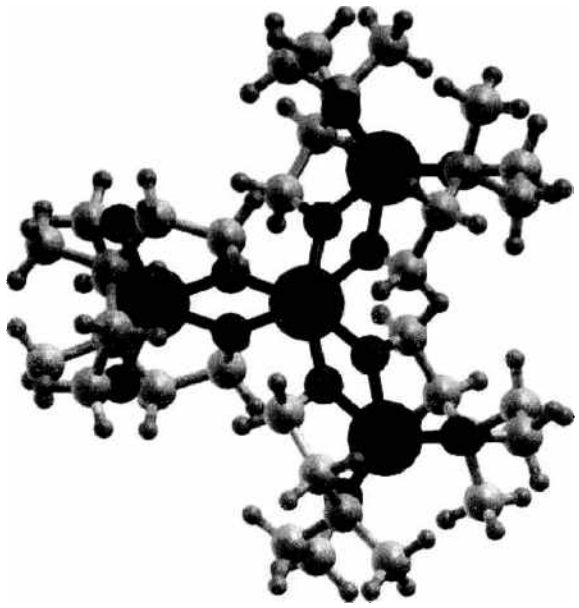


Figure 17. The molecular structure of the Fe_4-star. The Fe atoms are shown by large spheres.

experiment becomes worse by optimizing the molecular geometry in the calculations. This can be understood because the geometry optimization is done for a single isolated molecule, hence neglecting crystal packing effects and interactions in the molecular crystal, which are indeed important for the real molecular geometry.

6.3.3. Mn_{10} Cluster

In this system, 10 Mn atoms form a tetrahedron-like structure with Mn atoms at the corners and at the middle of the tetrahedron edges, all bridged by oxygen ions [250]. Two of the Mn atoms are coupled antiferromagnetically to all the rest. The calculation by Kortus et al. [209] indicates an ionic picture in which the first Mn has an Mn^{3+} ($S = 2$) state, whereas the other two are Mn^{2+} ($S = 5/2$). Because of the symmetry of the cluster, the two types of majority spin Mn atoms have a multiplicity of 4, whereas the minority spin Mn atom has a multiplicity of 2, resulting in the $S = 4 \times 2 + 4 \times 5/2 - 2 \times 5/2 = 13$ magnetic ground state. This magnetic core is further stabilized by organic rings that are also connected to the oxygen atoms. This molecular unit with the chemical formula $[Mn_{10}O_4(2,2'\text{-biphenoxide})_4Br_{12}]^{4-}$ is charged and compensated by another molecular cluster containing a single manganese atom, $[(CH_3CH_2)_3NH]_2[Mn(CH_3CN)_4(H_2O)_2]$.

The calculation confirms the experimental suggestion that the magnetic anisotropy is only a result of the functional unit containing 10 Mn atoms. The compensating cluster behaves paramagnetically with the Mn atom in +2 charge state and spin $s = 5/2$. As shown by Kortus et al. [209], the single Mn complex exhibits the easy-plane behavior with an energy well of only 0.1 K. The majority-spin gap in Mn_{10} is much smaller than the minority-spin one. Those matrix elements of Eq. (48) related to the occupied majority-spin states contribute in favor of an easy axis behavior, whereas the matrix elements from the occupied minority-spin channel favor easy plane. These tendencies compete and cancel each other to a large extent. Only as a result of the larger contribution from the occupied majority-spin channel does the complete Mn_{10} cluster end up as an easy-axis system (see Table 4). Therefore, in spite of the fact that Mn_{10} possesses a high-spin state (S is larger than in Mn_{12}-ac), the anisotropy barrier in this system is small.

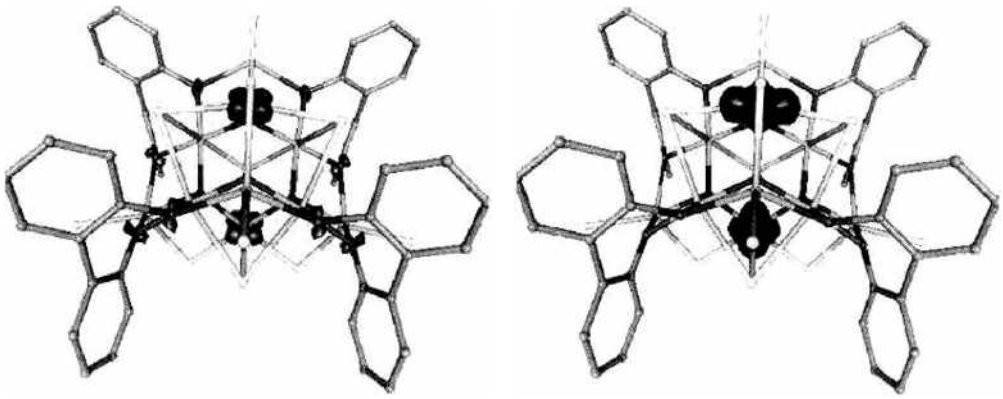
Kortus et al. [209] found that the removal of subsets of the Br ions will change the magnetic anisotropy drastically because of large perturbations of the electronic structure. However, neutralizing the electric field resulting from Br ions by an external potential in the calculations changed the anisotropy barrier by less than 1 K. Therefore, one can conclude that the electric fields created by the Br ions do not have any significant effect on the magnetic properties of the molecule, in contrast to chemical interactions.

One of the advantages of the first-principles approach is the possibility of controling in detail the interactions and states that are important for a certain physical property to gain a microscopic understanding. Equation (48) shows that the barrier is related to matrix elements between occupied and unoccupied orbitals in the majority and minority spin channels. In addition to the discrimination associated with spin pairing, we can analyze which electronic states mostly contribute to the matrix elements $M_{ij}^{\sigma\sigma'}$. In Fig. 18, we display plots of the square of the wavefunctions of the occupied majority state and the unoccupied minority state that contribute to the matrix element $M_{ij}^{\sigma\sigma'}$ with the largest absolute value. (Please note

Table 4. The contributions of the different spin channels (see Eq. 48) to the magnetic anisotropy parameter D and the magnetic anisotropy energy DS_z^2.

Occupied	Unoccupied	$D(K)$	$DS_z^2(K)$
Majority	Majority	−0.039	−6.6
Majority	Minority	−0.106	−17.9
Minority	Majority	0.034	5.7
Minority	Minority	0.055	9.3
All	All	−0.056	−9.5

Note: Adapted from [209].



(a) (b)

Figure 18. Isosurface (dark) at 0.005 e/a_B^3 of the square of the wavefunctions that contribute most to the matrix elements $M_{ii'}^{\sigma\sigma'}$ of Eq. (48). (a) occupied majority state; (b) unoccupied minority state for the Mn_{III} cluster. It is clearly visible that the matrix element connects majority and minority d-states at the same Mn atom.

that the value of the magnetic anisotropy energy is not determined by a single dominant contribution but results from the sum of many contributions with different signs.) It is clearly visible that the states of interest are d-states localized on the same Mn atom. In this case, the states are localized on the minority spin Mn atoms.

6.4. Some Results on the V_{15} Spin System

As already mentioned, the V_{15} system remains of great interest for studies on quantum coherence and relaxation phenomena, despite not having any sizeable magnetic anisotropy barrier [40, 41, 256, 257]. The dynamics of the magnetization relaxation depend on the spin-phonon interaction at finite temperatures, and an intrinsic phonon-bottleneck with a characteristic "butterfly" hysteresis has been demonstrated by Chiorescu et al. [256]. Because of several very recent experimental studies on this system, it became possible to check the quality of the electronic structure calculations. In a joint theoretical and experimental study by Boukhvalov et al. [187], the system has been investigated using the LSDA+U band structure calculations (the same computational method as referred to above, for calculations by the same group on Mn_{12}-ac [166]) and measured x-ray photoelectron and fluorescence spectra. Comparing experimental data with the results of electronic structure calculations, the authors conclude that the LMTO LSDA+U method provides a good description of the electronic structure of V_{15}.

Choi et al. [258] report the reflectance and optical conductivity of solid V_{15} over a wide energy range. The band centered at 1.2 eV is assigned as a V dd transition, and other features at 3.7, 4.3, and 5.6 eV are attributed to Op-Vd charge transfer excitations. The comparison of the results to recent electronic structure calculations [41, 187, 204] shows good agreement with all these calculations without clearly favoring any U value.

Chaboussant et al. [257] report an Inelastic Neutron Scattering study of the fully deuterated molecular compound. They deliver direct confirmation that the essential physics at low temperature is determined by three weakly coupled spin-(1/2) on a triangle. Interestingly, the experiment allowed us to determine the effective exchange coupling of 0.21 meV within the triangle and the gap between the two spin-(1/2) doublets of the ground state. This direct interaction had been predicted earlier by Kortus et al. [41], with a value of 0.55 meV.

The work by Kortus et al. [41] used an efficient coupled multilevel analysis that relied on fitting density-functional energies to mean-field Heisenberg or Ising energies to determine the exchange parameters. The approximate exchange parameters gleaned from the first N Ising configurations were used to find the next lowest energy Ising configuration and, subsequently, to improve the parameterization of the exchange parameters. The "self consistency" criterion in this approach was the check as to whether the predicted Ising levels

remain unchanged under the addition of data from new Ising configurations. This mapping of density-functional theory results on a classical Ising model allowed for the determination of the exchange parameters by considering only several spin configurations.

The data used to determine the exchange parameters from a least square fit to the mean-field solution of the Heisenberg Hamiltonian (Eq. 5) are displayed in Table 5. The fit is very good (with errors ranging from 0.1 to 1.55 meV) and leads to exchange parameters (in the notations of Fig. 4) of $J = 290.3$ meV, $J' = -22.7$ meV, $J'' = 15.9$ meV, $J1 = 13.8$ meV, $J2 = 23.4$ meV, and $J3 = 0.55$ meV, where positive numbers correspond to AFM and negative numbers to FM interactions. The FM interaction J' is a surprising result and deserves further discussion because it is qualitatively different from earlier assumptions based on entirely AFM interactions [40, 256]. A FM coupling is possible without polarizing the oxygens through a fourth-order process similar to super-exchange. In super-exchange, the intermediate state has the lowest d-orbital on the V' atom doubly occupied with up and down electrons. However, electrons can also hop to higher-energy d-orbitals. In this case, both parallel and antiparallel spins are allowed without violating the Pauli exclusion principle, and consistently, with the Hund's first rule, the parallel spin alignment is preferred. The super-exchange (within the same d-orbital) completely excludes electrons of the same spin, whereas the ferromagnetic process (different participating d-orbitals) merely favors FM alignment. Thus, a FM coupling is obtained if the V–O hopping matrix elements into the higher d-orbital are significantly larger than the matrix elements for the hopping of O electrons into the lowest-energy d-orbital. The occurrence of such interactions is possible in a low-symmetry system such as V_{15}. Even with this FM interaction, the spin Hamiltonian yields a $S = 1/2$ ground state composed largely of Ising configurations similar to the one depicted in Fig. 4. This Ising configuration was predicted from the Js from the earlier fits to density-functional theory energies and corresponds to the ground-state density-functional theory configuration (**I**).

Comparing the calculated susceptibility with experiment [256], one finds the that low-temperature behavior is not well reproduced and that the doublet-quadruplet gap $\Delta \approx 10$ K is significantly larger than the experimental value of $\Delta \approx 3.7$ K, whereas the high-temperature behavior shows that calculated value of J is too large.

Both of these discrepancies can be explained almost entirely by a J that is too large within the density-functional-based treatment. The large value of J can be attributed to both exchange processes through the oxygens and to direct exchange between the V. If direct exchange is important, the value of J will be influenced greatly by the overlap between the V atoms. Electronic correlations included in form of LDA + U may help to improve the agreement with experiment, because the overlap between the d-orbitals of the vanadium atoms will be decreased by shifting the occupied d-orbitals down in energy by U. Similarly,

Table 5. DFT energies (E in meV) of calculated Ising configurations, energies obtained from the fit, and $4\langle S_i^z S_j^z \rangle$ along each of the six bonds.

E	*Fit*	J	J'	J''	$J1$	$J2$	$J3$	Spin	Label	δ (K)
78.37	−78.44	−6	2	−2	6	−6	−1	1/2	I	0.8
73.39	−73.63	−6	2	−2	4	−4	1	1/2	II	
−35.48	−35.08	−6	−2	2	4	−4	−1	1/2	III	
−34.89	−34.53	−6	−2	2	4	−4	3	3/2	IV	
0.00	−0.79	−6	−6	6	6	−6	3	3/2	V	1.5
8.38	8.28	−6	−6	6	2	−2	−1	1/2	VI	1.3
28.14	28.08	−6	6	6	−6	6	3	3/2	VII	
126.32	126.14	−4	−4	6	4	−6	3	1/2	VIII	
129.17	128.88	−4	−4	2	6	−4	3	5/2	IX	
278.35	278.50	−2	−6	2	4	−4	3	3/2	X	
434.22	435.78	0	0	6	6	0	3	9/2	XI	1.6
760.75	760.76	6	6	6	−6	−6	3	9/2	XII	1.6
873.11	872.35	6	6	6	6	6	3	15/2	XIII	1.8

Note: Also included is the anisotropy shift δ for the $M_s = S$ state of each Ising configuration. A least square fit of this data leads to exchange parameters of $J = 290.3$, $J' = -22.7$, $J'' = 15.9$, $J1 = 13.8$, $J2 = 23.4$ and $J3 = 0.55$ meV. Adapted from [41].

self-interaction correction (SIC) [151, 152] will lower the magnitude of J because they will localize the V d-orbitals more, reducing the overlap of the wavefunctions.

Without including a direct exchange interaction between the vanadium atoms in the inner triangle ($J3 = 0$), reducing J to 70 meV and slightly reducing the difference between $J1$ and $J2$ yields the experimentally observed effective moment, although another set of only antiferromagnetic interactions [40] also fits the experimental results well. In fact, any set of parameters with the correct values of J and Δ given by simple perturbation theory

$$\Delta = \frac{3}{4}\frac{(J2 - J1)^2(J'' - J')}{J^2} + \frac{3}{2}J3 \tag{60}$$

will fit the experimental effective moment well. The already-mentioned problem of the parameter dependence on the assumed model arises here.

By including a possible direct interaction between the triangle vanadium atoms ($J3$) in the spin Hamiltonian, the agreement with experiment can be achieved by dividing all Js by a constant factor of 2.9. Scaling $J3$ down by a factor of 2.9 gives a value of 0.19 meV, in surprisingly good agreement with the corresponding value obtained from inelastic neutron scattering [257] of 0.221 meV. One possibility for deciding between different models could be the measurement of the spin ordering and the spin–spin correlation functions by, for example, neutron scattering.

6.5. Model Fe-Binuclear System

Binuclear metal-organic systems form a large, and very simple, group among molecular magnets. Even if their magnetic characteristics like ordering temperature and bulk magnetization are not necessarily outstanding, they help to grasp important physics of $3d$–$3d$ magnetic interaction mediated by an organic ligand, and thus offer a convenient model system. Moreover, an interesting effect of spin-crossover has been observed in some such systems, for instance, in $[Fe(bt)(NCS)_2]_2$-bpym (bt = 2, 2′-bithiazoline, bpym = 2, 2′-bipyrimidine): a switch from LS–LS to LS–HS to HS–HS configuration (LS: low spin; HS: high spin) at the increase of temperature, where the intermediate LS–HS state gets stabilized near 170 K because of an interplay between intermolecular and intramolecular magnetic interactions [25, 128, 259]. The possibility of optical switching between different magnetic states was demonstrated [260] and brought into discussion the prospects of using such systems as active elements in memory devices.

Our interest in binuclear systems is primarily that for model molecular magnets, to be treated with a method of recognized accuracy, and with the aim of looking at the effect of intraatomic correlation effects ("Hubbard U"). Starting from the real structure of $[Fe(bt)(NCS)_2]_2$-bpym (see Fig. 19, left panel), we "streamlined" it somehow to fit it into a compact unit cell for an accurate calculation by a band structure method with periodic boundary conditions (Fig. 19, right panel). This transformation preserved the bipyrimidine part between two Fe centers but "shortcut" the distant parts of ligands to make a connected structure. The calculation has been done with the FLEUR code [231], a realization of full-potential augmented plane wave technique. One can see that, in contrast to ferric wheels, the Fe atom is now octahedrally coordinated by nitrogen ions. A formal valence state in these compounds is routinely referred to as Fe(II). The HS and LS states were discussed as being represented by the $t_{2g}^4 e_g^2$ and t_{2g}^6 configurations, correspondingly [128]. Our calculation did not yet include the orbital transition of this type; we initialized only HS configurations and brought them into self-consistency in FM and AFM settings. The resulting partial DOS are shown in Fig. 20.

Certain similarities can be found with the Fe local DOS in ferric wheels—clear splitting into t_{2g}-like and e_g-like states in a nearly octahedral ligand field, full occupation of majority-spin Fe$3d$ states, and one electron per Fe atom trapped in the Fe$3d$–N$2p$ hybridized band of minority spin. The values of magnetic moments (total per Fe atom in the FM case, along with the local moment, integrated over the muffin-tin sphere) are listed in Table 6. The interatomic exchange parameters have been estimated from total energy differences between FM and AFM cases.

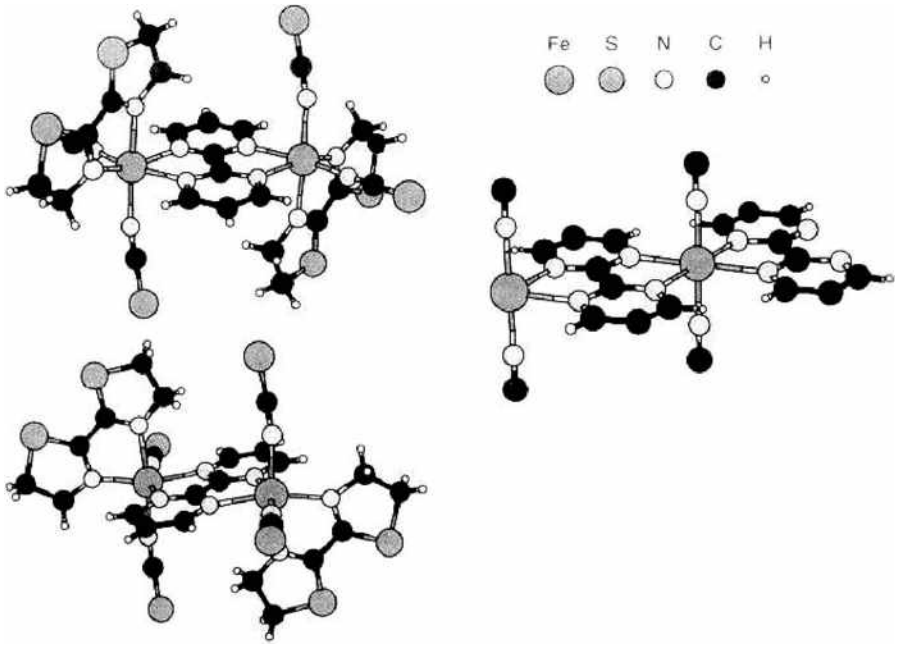


Figure 19. Two views of the $[Fe(bt)(NCS)_2]_2$-bpym molecule (left panel) and a simplified periodic Fe-binuclear system used in the FLEUR calculation (right panel).

Because the magnetic moment is largely localized at the Fe site, the inclusion of intraatomic correlations beyond the "conventional" density-functional theory might be important. The exchange parameters J depend on the spatial overlap of the d-orbitals on different Fe sites. It is well known that the d-orbitals within density-functional theory are

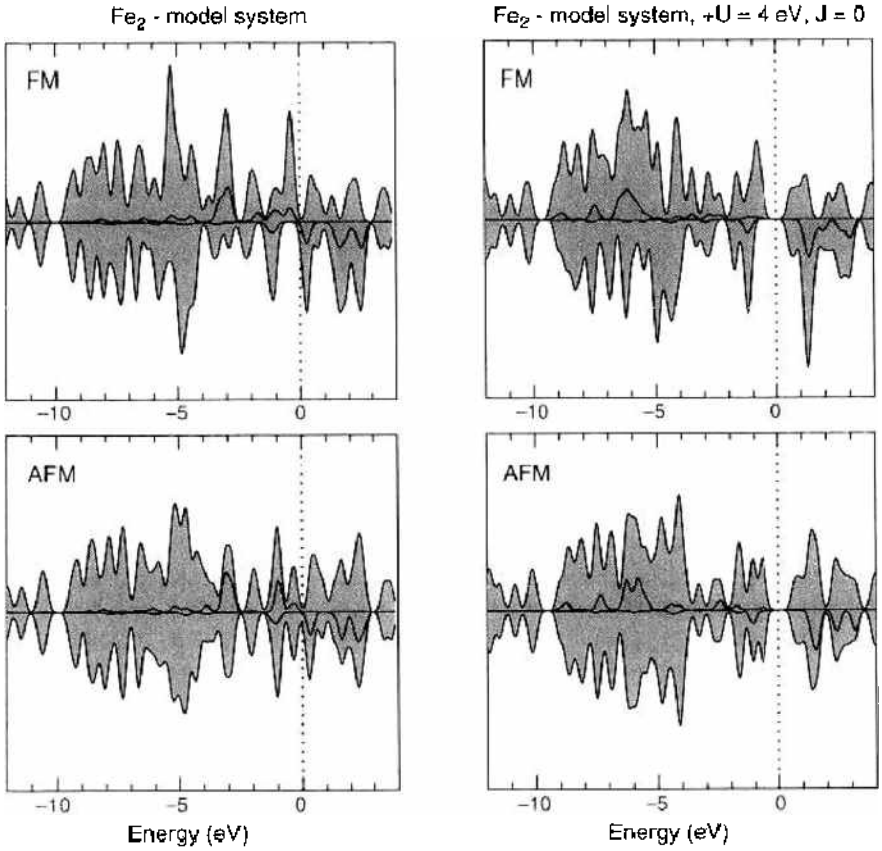


Figure 20. Densities of states in FM and AFM cases as calculated by FLEUR for the model Fe-binuclear system, in the DFT and in the LDA + U approach. Fe local DOS are shown as shaded areas.

Table 6. Magnetic moments and interaction parameters as estimated for a model Fe-binuclear system (Fig. 19) from calculations by FLEUR with and without Hubbard U.

		M(Fe)	M/Fe	ΔE	$J(S = 5/2)$
$U = 0$	FM	3.62	4.10	—	—
	AFM	3.61	—	102.5 meV	−190 K
$U = 4$ eV	FM	3.93	4.94	—	—
	AFM	3.92	—	76.8 meV	−143 K

not localized enough compared to experiment, and as a consequence, the J values will be overestimated. There are two main reasons for this shortcoming. First, possible on-site correlations as known from atomic physics are underestimated in the case of "conventional" density-functional theory. Second, density-functional theory is not free from spurious self-interactions resulting from the replacement of the point-like electrons by corresponding densities. Bringing in the atomic physics in the form of LDA + U (and adding a local orbital dependent atomic Coulomb interaction parameter U to DFT [153] or SIC [151, 152]) improve the results by lowering the d-orbitals in energy and therefore localizing them stronger. SIC only affects occupied states, whereas LDA + U plunges the occupied d-states and shifts the unoccupied ones to higher energies. By increasing, on average, the magnetic excitation energy across the spin majority–minority gap, both mechanisms help to effectively reduce the magnitude of J. To our knowledge, SIC have not yet been applied in calculations on molecular magnets (nor are we aware of any practical implementation of SIC in a full-potential code; i.e., beyond the muffin-tin or atomic sphere approximation. Pederson et al. (private communication) are actively working toward a practical implementation of SIC within the NRLMOL suite of codes. The LDA + U scheme is implemented in the FLEUR code (as in many others). This ansatz has, however, the disadvantage of not being truly first-principles one: It remains on the user to single out certain orbitals as localized and to choose an appropriate value for the "Hubbard U" parameter. For the Fe-binuclear system we have chosen an empirically reasonable value $U = 4$ eV; in principle, we were more interested in studying qualitative trends, as they deal with a model system anyway. One observes from Table 6 that the inclusion of intraatomic correlation somehow enhances the local magnetic moment at the Fe site and, to a much smaller extent, the total magnetic moment (in the FM configuration). Much more important, the J parameter is noticeably reduced as a result of the inclusion of correlation. These observations agree with what was earlier reported by Boukhvalov et al. [166] for the "Mn_{12}" system from the LDA + U calculation.

7. CONCLUSION

We attempted to give a broad overview of physical questions and technical problems one encounters in modern first-principles simulations in the rapidly growing field of molecular magnets. Our own presented results largely correspond to work still in progress, and they might be far from providing a definitive answer for particular systems. On the contrary, these results are likely to be refined and extended by subsequent studies. Our current results make us very confident in the predictive power of the presented methods. To explore the range of systems in which the presented first-principles methods give reliable results, further studies on more systems are required. A large number of calculations are being performed by other groups on many other systems, which we might have failed to name in this limited contribution. However, it is our hope that it may help the newcomers in the field to access the problems, the difficulties experienced, and the possibilities offered by different methods and practical schemes of first-principles calculation. Many additional calculations are required to obtain a complete understanding of the idealized behaviors of molecular magnets, and both new theoretical and new computational tools will be required to understand the real world that will define the operating environments in application of such systems.

ACKNOWLEDGMENTS

We thank the Deutsche Forschungsgemeinschaft for financial support (Priority Program SPP 1137 "Molecular Magnetism") and appreciate useful discussions with Mark R. Pederson, Roberta Sessoli, Jürgen Schnack, Stefan Blügel, Sorin Chiuzbăian, Manfred Neumann, Danil Boukhvalov, Vladimir Mazurenko, Vladimir Anisimov, and Mikhail Katsnelson. We are grateful to Lilia Boeri for critically reading the manuscript. The crystal structure data for calculations of some molecular magnets have been kindly provided by the group of Rolf Saalfrank of the University Erlangen-Nürnberg, Roberta Sessoli of the University of Florence, and Paul Kögerler of the Ames Laboratory. A.V.P. thanks Gustav Bihlmayer for introduction into the FLEUR code and for his help in performing the calculations on a Fe-binuclear system. A.V.P. was supported in part by the Research Council of the President of the Russian Federation (grant NSH-1026.2003.2) and by the Russian Foundation for Basic Research (project 02-02-16674).

REFERENCES

1. R. J. T. Morris and B. J. Truskowski, *IBM Sys J.* 42, 205 (2003).
2. O. Kahn, "Molecular Magnetism," Wiley, Singapore, 1993.
3. M. Verdaguer, *Polyhedron* 20, 1115 (2001).
4. W. Linert and M. Verdaguer, Eds., "Molecular Magnets," Springer-Verlag, Wien, 2003.
5. J. B. Goodenough, *Phys. Rev.* 100, 564 (1955).
6. J. B. Goodenough, *J. Phys. Chem. Solids* 6, 287 (1958).
7. J. Kanamori, *J. Phys. Chem. Solids* 10, 87 (1959).
8. J. van Vleck, *Phys. Rev.* 52, 1178 (1937).
9. Materials Research Society Bulletin 25, November (2000).
10. A. J. Banister, N. Bricklebank, I. Lavender, J. M. Rawson, C. I. Gregory, B. K. Tanner, W. Clegg, M. R. J. Elsegood, and F. Palacio, *Angew. Chem.* 35, 2533 (1996).
11. R. Sessoli, H. Tsai, A. Schake, S. Y. Wang, J. B. Vincent, K. Folting, D. Gatteschi, G. Christou, and D. N. Hendrickson, *J. Am. Chem. Soc.* 115, 1804 (1993).
12. J. R. Friedman, M. P. Sarachik, J. Tejada, and R. Ziolo, *PRL* 76, 3830 (1996).
13. L. Thomas, F. Lionti, R. Ballou, D. Gatteschi, R. Sessoli, and B. Barbara, *Nature (London)* 383, 145 (1996).
14. B. Barbara and L. Gunther, *Phys. World* 12, 35 (1999).
15. D. Gatteschi and R. Sessoli, *Angew. Chem. Int. Ed.* 42, 268 (2003).
16. M. N. Leuenberger and D. Loss, *Nature* 410, 789 (2001).
17. M. N. Leuenberger, F. Meier, and D. Loss, *Monatshefte Chem.* 134, 217 (2003).
18. M. Verdaguer, A. Bleuzen, C. Train, R. Garcıe, F. Fabrizi de Biani, and C. Desplanches, *Phil. Trans. R. Soc. Lond.* A 357, 2959 (1999).
19. M. Verdaguer, N. Galvez, R. Garde, and C. Desplanches, *Electrochem. Soc. Interface* 11, 29 (2002).
20. S.-I. Ohkoshi and K. Hashimoto, *Electrochem. Soc. Interface* 11, 34 (2002).
21. S. Ferlay, T. Mallah, R. Ouah' es, P. Veillet, and M. Verdaguer, *Nature* 378, 701 (1995).
22. S. M. Holmes and G. S. Girolami, *J. Am. Chem. Soc.* 121, 5593 (1999).
23. M. R. Pederson, A. Y. Liu, T. Baruah, E. Z. Kurmaev, A. Moewes, S. Chiuzbăian, M. Neumann, C. R. Kmety, K. L. Stevenson, and D. Ederer, *Phys. Rev. B* 66 (2002).
24. A. B. Gaspar, M. C. Muñoz, N. Moliner, V. Ksenofontov, G. Levchenko, P. Gütlich, and J. A. Real, *Monatshefte Chem.* 134, 285 (2003).
25. V. Ksenofontov, A. B. Gaspar, J. A. Real, and P. Gütlich, *J. Phys. Chem.* 105, 12266 (2001).
26. T. Lis, *Acta Crystallogr. Soc. B* 36, 2042 (1980).
27. A. Caneschi, D. Gatteschi, R. Sessoli, A. L. Barra, L. C. Brunel, and M. Guillot, *J. Am. Chem. Soc.* 113, 5873 (1991).
28. A. L. Barra, D. Gatteschi, and R. Sessoli, *Phys. Rev. B* 56, 8192 (1997).
29. A. Cornia, M. Affronte, A. G. M. Jansen, D. Gatteschi, A. Caneschi, and R. Sessoli, *Chem. Phys. Lett.* 322, 477 (2000).
30. I. Mirebeau, M. Hennion, H. Casalta, H. Andres, H. U. Güdel, A. V. Irodova, and A. Caneschi, *Phys. Rev. Lett.* 83, 628 (1999).
31. R. A. Robinson, P. J. Brown, D. N. Argyriou, D. N. Hendrickson, and S. M. J. Aubin, *J. Phys. Condens. Matter* 12, 2805 (2000).
32. H. J. Eppley, H.-L. Tsai, N. de Vries, K. Folting, G. Christou, and D. N. Hendrickson, *J. Am. Chem. Soc.* 117, 301 (1995).
33. A. M. Gomes, M. A. Novak, R. Sessoli, A. Caneschi, and D. Gatteschi, *Phys. Rev. B* 57, 5021 (1998).
34. K. Wieghardt, K. Pohl, I. Jibril, and G. Huttner, *Angew. Chem.* 23, 77 (1984).
35. R. Caciuffo, G. Amoretti, A. Murani, R. Sessoli, A. Caneschi, and D. Gatteschi, *Phys. Rev. Lett.* 81, 4744 (1998).
36. W. Wernsdorfer, T. Ohm, C. Sangregorio, R. Sessoli, D. Mailly, and C. Paulsen, *Phys. Rev. Lett.* 82, 3903 (1999).

37. W. Wernsdorfer and R. Sessoli, *Science* 284, 133 (1999).
38. A. Garg, *Europhys. Lett.* 22, 205 (1993).
39. A. Müller and J. Döring, *Angew. Chem.* 27, 1721 (1988).
40. D. Gatteschi, L. Pardi, A. L. Barra, A. Müller, and J. Döring, *Nature (London)* 354, 463 (1991).
41. J. Kortus, C. S. Hellberg, and M. R. Pederson, *Phys. Rev. Lett.* 86, 3400 (2001).
42. R. Sessoli, D. Gatteschi, A. Caneschi, and M. A. Novak, *Nature (London)* 365, 141 (1993).
43. L. Gunther and B. Barbara, eds., "Quantum Tunneling of Magnetization-QTM '94," NATO Advanced Studies Institute, Series E: Applied Sciences, Vol. 301, Kluwer, Dordrecht, 1995.
44. J. R. Friedman, M. P. Sarachik, J. Tejada, and R. Ziolo, *Phys. Rev. Lett.* 76, 3830 (1996).
45. A.-L. Barra, P. Brunner, D. Gatteschi, C. E. Schulz, and R. Sessoli, *Europhys. Lett.* 35, 133 (1996).
46. L. Gunther, *Phys. World* 3, 28 (1990).
47. D. A. Garanin and E. M. Chudnovsky, *Phys. Rev. B* 56, 11102 (1997).
48. F. Luis, J. Bartolomé, and J. F. Fernández, *Phys. Rev. B* 57, 505 (1997).
49. G. Bellessa, N. Vernier, B. Barbara, and D. Gatteschi, *Phys. Rev. Lett.* 83, 416 (1999).
50. M. N. Leuenberger and D. Loss, *Phys. Rev. B* 61, 1286 (2000).
51. E. M. Chudnovsky and D. A. Garanin, *Phys. Rev. Lett.* 85, 5259 (2000).
52. S. Hill, J. A. A. J. Perenboom, N. S. Dalal, T. Hathaway, T. Stalcup, and J. S. Brooks, *Phys. Rev. Lett.* 80, 2453 (1998).
53. E. del Barco, J. M. Hernandez, J. Tejada, N. Biskup, R. Achey, I. Rutel, N. Dalal, and J. Brooks, *Phys. Rev. B* 62, 3018 (2000).
54. A. B. Sushkov, B. R. Jones, J. L. Musfeldt, Y. J. Wang, R. M. Achey, and N. S. Dalal, *Phys. Rev. B* 63, 214408 (2001).
55. B. Parks, J. Loomis, E. Rumberger, D. N. Hendrickson, and G. Christou, *Phys. Rev. B* 64, 184426 (2001).
56. K. Park, M. A. Novotny, N. S. Dalal, S. Hill, and P. A. Rikvold, *Phys. Rev. B* 66, 144409 (2002).
57. K. Park, M. A. Novotny, N. S. Dalal, S. Hill, and P. A. Rikvold, *J. Appl. Phys.* 91, 7167 (2002).
58. M. Dressel, B. Gorshunov, K. Rajagopal, S. Vongtragool, and A. A. Mukhin, *Phys. Rev. B* 67, 060405 (2003).
59. A. Cuccoli, A. Fort, A. Rettori, E. Adam, and J. Villain, *Eur. Phys. J. B* 12, 39 (1999).
60. N. V. Prokof'ev and P. C. E. Stamp, *Phys. Rev. Lett.* 80, 5794 (1998).
61. J. F. Fernández and J. J. Alonso, *Phys. Rev. B* 62, 53 (2000).
62. X. Martínez-Hidalgo, E. M. Chudnovsky, and A. Aharony, *Europhys. Lett.* 55, 273 (2001).
63. J. F. Fernández, *Phys. Rev. B* 66, 064423 (2002).
64. A. Garg, *Phys. Rev. Lett.* 70, 1541 (1993).
65. E. M. Chudnovsky, *Phys. Rev. Lett.* 72, 1134 (1994).
66. D. A. Garanin, E. M. Chudnovsky, and R. Schilling, *Phys. Rev. B* 61, 12204 (1999).
67. W. Wernsdorfer, A. Caneschi, R. Sessoli, D. Gatteschi, A. Cornia, V. Villar, and C. Paulsen, *Phys. Rev. Lett.* 84, 2965 (2000).
68. R. Sessoli, A. Caneschi, D. Gatteschi, L. Sorace, A. Cornia, and W. Wernsdorfer, *J. Magn. Magnet. Mater.* 226, 1954 (2001).
69. D. A. Garanin and E. M. Chudnovsky, *Phys. Rev. Lett.* 87, 187203 (2001).
70. D. A. Garanin and E. M. Chudnovsky, *Phys. Rev. B* 65, 094423 (2002).
71. W. Wernsdorfer, N. Aliaga-Alcalde, D. N. Hendrickson, and G. Christou, *Nature* 416, 406 (2002).
72. W. Wernsdorfer, S. Bhaduri, R. Tiron, D. N. Hendrickson, and G. Christou, *Phys. Rev. Lett.* 89, 197201 (2002).
73. S. Hill, R. S. Edwards, S. I. Jones, N. S. Dalal, and J. M. North, *Phys. Rev. Lett.* 90, 217204 (2003).
74. A. Cornia, R. Sessoli, L. Sorace, D. Gatteschi, A. L. Barra, and C. Daiguebonne, *Phys. Rev. Lett.* 89, 257201 (2002).
75. A. Caneschi, D. Gatteschi, C. Sangregorio, R. Sessoli, L. Sorace, A. Cornia, M. A. Novak, C. Paulsen, and W. Wernsdorfer, *J. Magn. Magnet. Mater.* 200, 182 (1999).
76. E. M. Chudnovsky and D. A. Garanin, *Phys. Rev. Lett.* 89, 157201 (2002).
77. J. Tejada, R. Amigo, J. M. Hernandez, and E. M. Chudnovsky, *Phys. Rev. B* 68, 014431 (2003).
78. R. W. Saalfrank, I. Bernt, E. Uller, and F. Hampel, *Angew. Chem.* 36, 2482 (1997).
79. O. Waldmann, J. Schülein, R. Koch, P. Muller, I. Bernt, R. W. Saalfrank, H. P. Andres, H. U. Gudel, and P. Allenspach, *Inorg. Chem.* 38, 5879 (1999).
80. O. Waldmann, R. Koch, S. Schromm, J. Schülein, P. Müller, I. Bernt, R. W. Saalfrank, F. Hampel, and E. Balthes, *Inorg. Chem.* 40, 2986 (2001).
81. F. D. M. Haldane, *Phys. Rev. Lett.* 50, 1153 (1983).
82. J. Schnack and M. Luban, *Phys. Rev. B* 63, 014418 (2001).
83. O. Waldmann, *Phys. Rev. B* 65, 024424 (2002).
84. F. Meier and D. Loss, *Phys. Rev. B* 64, 224411 (2001).
85. O. Waldmann, R. Koch, S. Schromm, P. Müller, I. Bernt, and R. W. Saalfrank, *Phys. Rev. Lett.* 89, 246401 (2002).
86. E. C. Stoner, *Rev. Mod. Phys.* 25, 2 (1953).
87. C. Kittel, in "Proceedings of the Kamerlingh Onnes Conferences on Low-Temperature Physics," Suppl. Physica, Vol. 24, p. S88. (1958).
88. W. P. Wolf, *Phys. Rev.* 115, 1196 (1959).
89. A. Steane, *Rep. Prog. Phys.* 61, 117 (1998).
90. D. Bouwmeester, A. Ekert, A. Zeilinger, and D. Bouwmeester, Eds., "The Physics of Quantum Information: Quantum Cryptography, Quantum Teleportation, Quantum Computation," Springer Verlag, 2000.

91. G. Johnson, "A Shortcut Through Time: The Path to the Quantum Computer," Alfred A. Knopf, 2003.
92. P. W. Shor, SIAM *J. Comput.* 26, 1484 (1997).
93. L. M. K. Vandersypen, M. Steffen, G. Breyta, C. S. Yannoni, M. H. Sherwood, and I. L. Chuang, *Nature (London)* 415, 883 (2001).
94. L. K. Grover, *Phys. Rev. Lett.* 79, 325 (1997).
95. C. H. Bennett, E. Bernstein, G. Brassard, and U. Vazirani, SIAM *J. Comput.* 26, 1510 (1997).
96. L. K. Grover, *Phys. Rev. Lett.* 79, 4709 (1997).
97. P. W. Anderson, in "Solid State Physics," (F. Seitz and D. Turnbull, Eds.), Advances in Research and Applications, p. 99, Vol. 14, Academic Press, New York, 1963.
98. C. Zener, *Phys. Rev.* 82, 403 (1951).
99. P. W. Anderson and H. Hasegawa, *Phys. Rev.* 100, 675 (1955).
100. J. M. D. Coey, M. Viret, and S. von Molnár, *Adv. Phys.* 48, 167 (1999).
101. P. W. Anderson, *Phys. Rev.* 115, 2 (1959).
102. H. Weihe and H. U. Güdel, *Inorg. Chem.* 36, 3632 (1997).
103. C. Kollmar and O. Kahn, *Acc. Chem. Res.* 26, 259 (1993).
104. L. Onsager, *Phys. Rev.* 65, 117 (1944).
105. I. Dzyaloshinsky, *J. Phys. Chem. Solids* 4, 241 (1958).
106. T. Moriya, *Phys. Rev.* 120, 91 (1960).
107. S. Blundell, "Magnetism in Condensed Matter," Oxford Master Series in Condensed Matter Physics, Oxford University Press, Oxford, UK, 2001.
108. M. I. Katsnelson, V. V. Dobrovitski, and B. N. Harmon, *Phys. Rev. B* 59, 6919 (1999).
109. R. L. Mössbauer, *Naturwissenschaften* 45, 538 (1958).
110. R. L. Mössbauer, *Z. Physik* 151, 124 (1958).
111. N. N. Greenwood and T. C. Gibb, "Mössbauer Spectroscopy," Chapman and Hall, London, 1971.
112. A. Vértes, L. Korecz, and K. Burger, "Mössbauer Spectroscopy," Studies in Physical, and Theoretical Chemistry, Vol. 5, Elsevier Scientific Publishing Company, Amsterdam, 1979.
113. U. Gonser, ed., "Mössbauer Spectroscopy," Topics in Applied Physics, Vol. 5, Springer-Verlag, Berlin, 1975.
114. P. Gütlich, in "Mössbauer Spectroscopy" (U. Gonser, Ed.), Topics in Applied Physics, Vol. 5, pp. 53–96, Springer-Verlag, 1975.
115. P. Jena, *Phys. Rev. B* 17, 1046 (1978).
116. P. Blaha, K. Schwarz, and P. Herzig, *PRL* 54, 1192 (1985).
117. H. M. Petrilli, P. E. Blöchl, P. Blaha, and K. Schwarz, *PRB* 57, 14690 (1988).
118. P. Dufek, P. Blaha, and K. Schwarz, *PRL* 75, 3545 (1995).
119. Y. Zhang and E. Oldfield, *J. Phys. Chem. A* 107, 4147 (2003).
120. V. Schünemann and H. Winkler, *Rep. Progr. Phys.* 63, 263 (2000).
121. K. L. Taft, C. D. Delfs, G. C. Papaefthymiou, S. Foner, D. Gatteschi, and S. J. Lippard, *J. Am. Chem. Soc.* 116, 823 (1994).
122. S. M. Gorun, G. C. Papaefthymiou, R. B. Frankel, and S. J. Lippard, *J. Am. Chem. Soc.* 105, 3337 (1987).
123. Z. Zeng, Y. Duan, and D. Guenzburger, *Phys. Rev. B* 55, 12522 (1997).
124. Z. Zeng, D. Guenzburger, and D. E. Ellis, *Phys. Rev. B* 59, 6927 (1999).
125. L. Bottyan, N. S. Ovanesyan, A. A. Pyalling, N. A. Sanina, and A. B. Kashuba, *Hyperfine Interactions* 126, 149 (2000).
126. E. Coronado, J.-R. Galán-Mascarós, C.-J. Gómez-García, J. Ensling, and P. Gütlich, *Chem. Euro. J.* 6, 552 (2000).
127. J.-A. Real, H. Bolvin, A. Bousseksou, A. Dworkin, O. Kahn, F. Varret, and J. Zarembowitch, *J. Am. Chem. Soc.* 114, 4650 (1992).
128. V. Ksenofontov, H. Spiering, S. Reiman, Y. Garcia, A. B. Gaspar, N. Moliner, J. A. Real, and P. Gütlich, *Chem. Phys. Lett.* 348, 381 (2001).
129. C. P. Slichter, Ed., "Principles of Magnetic Resonance," Springer Series in Solid-State Sciences, Vol. 1, Springer-Verlag, Berlin, 1978.
130. J. A. Weil, J. R. Bolton, and J. E. Wertz, "Electron Paramagnetic Resonance: Elementary Theory and Practical Applications," Wiley, New York, 1994.
131. A. Abragam, "The Principles of Nuclear Magnetism," Clarendon Press, Oxford, 1978.
132. P. Sohár, "Nuclear Magnetic Resonance Spectroscopy," Vol. 1, CRC Press, Boca Raton, FL, 1983.
133. T. Goto, T. Kubo, T. Koshiba, Y. Fujii, A. Oyamada, J. Arai, K. Takeda, and K. Awaga, *Physica B* 284, 1227 (2000).
134. D. Sebastiani, *Modern Phys. Lett. B* 17, 1301 (2003).
135. B. Rakvin, D. Žilić, J. M. North, and N. S. Dalal, *J. Magnet. Resonance* 165, 260 (2003).
136. E. Bill, U. Beckmann, and K. Wieghardt, *Hyperfine Interactions* 144, 183 (2002).
137. K. Park, M. A. Novotny, N. S. Dalal, S. Hill, and P. A. Rikvold, *J. Appl. Phys.* 91, 7167 (2002).
138. K. Park, M. A. Novotny, N. S. Dalal, S. Hill, and P. A. Rikvold, *Phys. Rev. B* 65, 014426 (2002).
139. B. Parks, J. Loomis, E. Rumberger, D. N. Hendrickson, and G. Christou, *Phys. Rev. B* 64, 184426 (2001).
140. J. van Slageren, S. Vongtragool, B. Gorshunov, A. A. Mukhin, N. Karl, J. Krzystek, J. Telser, A. Müller, C. Sangregorio, D. Gatteschi, and M. Dresse, *Phys. Chem. Chem. Phys.* 5, 3837 (2003).
141. R. M. Dreizler and E. K. U. Gross, "Density Functional Theory. An Approach to the Quantum Many-Body Problem," Springer-Verlag, Berlin, 1990.

142. H. Eschrig, "The Fundamentals of Density Functional Theory," Teubner-Texte zur Physik, Vol. 32. B. G. Teubner Verlagsgesellschaft, Stuttgart, 1996.
143. P. Hohenberg and W. Kohn, *Phys. Rev.* 136, B864 (1964).
144. W. Kohn and L. J. Sham, *Phys. Rev.* 140, A1133 (1965).
145. G. Senatore and N. H. March, *Rev. Mod. Phys.* 66, 445 (1994).
146. W. M. C. Foulkes, L. Mitas, R. J. Needs, and G. Rajagopal, *Rev. Modern Phys.* 73, 33 (2001).
147. D. J. Singh, W. E. Pickett, and H. Krakauer, *Phys. Rev. B* 43, 11628 (1991).
148. J. P. Perdew, K. Burke, and M. Ernzerhof, *Phys. Rev. Lett.* 77, 3865 (1996).
149. O. Gunnarsson, M. Jonson, and B. I. Lundqvist, *Phys. Rev. B* 20, 3136 (1979).
150. I. I. Mazin and D. J. Singh, "First-Principles Calculations for Ferroelectrics: Fifth Williamsburg Workshop" (R. E. Cohen, Ed.). AIP Conference Proceedings, Vol. 46, pp. 251–264. American Institute of Physics, Woodbury, 1998.
151. A. Svane and O. Gunnarsson, *Phys. Rev. B* 37, 9919 (1988).
152. A. Svane and O. Gunnarsson, *Phys. Rev. Lett.* 65, 1148 (1990).
153. V. I. Anisimov, F. Aryasetiawan, and A. I. Lichtenstein, *J. Phys. Condens. Matter* 9, 767 (1997).
154. R. O. Jones and O. Gunnarsson, *Rev. Mod. Phys.* 61, 689 (1989).
155. W. Koch and M. C. Holthausen, "A Chemist's Guide to Density Functional Theory," Wiley-VCH, Weinheim, 2000.
156. M. Springborg, "Methods of Electronic-Structure Calculations: From Molecules to Solids," Wiley, Chichester, 2000.
157. K. Capelle, *J. Chem. Phys.* 119, 1285 (2003).
158. K. Schwarz and P. Mohn, *J. Phys. F Metal Phys.* 14, L129 (1984).
159. V. A. Gubanov and D. E. Ellis, *Phys. Rev. Lett.* 44, 1633 (1980).
160. L. M. Sandratskii, *Adv. Phys.* 47, 91 (1998).
161. T. Oguchi, K. Terakura, and A. R. Williams, *Phys. Rev. B* 28, 6443 (1983).
162. A. I. Liechtenstein, M. I. Katsnelson, and V. A. Gubanov, *J. Phys. F Metal Phys.* 14, L125 (1984).
163. A. I. Liechtenstein, M. I. Katsnelson, V. P. Antropov, and V. A. Gubanov, *J. Magn. Magnet. Mater.* 67, 65 (1987).
164. V. P. Antropov, M. I. Katsnelson, and A. I. Liechtenstein, *Physica B* 237-238, 336 (1997).
165. A. I. Liechtenstein, V. I. Anisimov, and J. Zaanen, *Phys. Rev. B* 52, R5467 (1995).
166. D. W. Boukhvalov, A. I. Lichtenstein, V. V. Dobrovitski, M. I. Katsnelson, B. N. Harmon, V. V. Mazurenko, and V. I. Anisimov, *Phys. Rev. B* 65, 184435 (2002).
167. H. J. F. Jansen, *Phys. Rev. B* 38, 8022 (1988).
168. X. F. Zhong, W. Y. Ching, and W. Lai, *J. Appl. Phys.* 70, 6146 (1991).
169. A. Shick, A. J. Freeman, R. Q. Wu, and L. J. Chen, *J. Magn. Magnet. Mater.* 177, 1216 (1998).
170. H. J. F. Jansen, *Phys. Rev. B* 59, 4699 (1999).
171. G. Schneider and H. J. F. Jansen, *J. Appl. Phys.* 87, 5875 (2000).
172. M. R. Pederson and S. N. Khanna, *Phys. Rev. B* 59, 693 (R) (1999).
173. M. R. Pederson and S. N. Khanna, *Phys. Rev. B* 60, 9566 (1999).
174. D. J. Singh, "Planewaves, Pseudopotentials and the LAPW Method." Kluwer Academic, Boston, 1994.
175. T. Helgaker and P. R. Taylor, in "Modern Electronic Structure Theory, Part II" (D. R. Yarkony, Ed.). Advanced Series in Physical Chemistry, Vol. 2, chap. 12, pp. 725–856. World Scientific, Singapore, 1995.
176. T. A. Arias, *Rev. Mod. Phys.* 71, 267 (1999).
177. T. L. Beck, *Rev. Mod. Phys.* 72, 1041 (2000).
178. H. Bachau, E. Cormier, P. Decleva, J. E. Hansen, and F. Martín, *Rep. Progr. Phys.* 64, 1815 (2001).
179. J. E. Pask, B. M. Klein, C. Y. Fong, and P. A. Sterne, *Phys. Rev. B* 59, 12352 (1999).
180. P. Blaha, K. Schwarz, G. K. H. Madsen, D. Kvasnicka, and J. Luitz, WIEN2k, Vienna, 2001 edition; P. Blaha, K. Schwarz, P. Sorantin, and S. B. Trickey, *Comput. Phys. Commun.* 59, 339 (1990).
181. P. E. Blöchl, *Phys. Rev. B* 50, 17953 (1994).
182. G. Kresse and D. Joubert, *Phys. Rev. B* 59, 1758 (1999).
183. O. K. Andersen, *Phys. Rev. B* 12, 3060 (1975).
184. O. K. Andersen, O. Jepsen, and M. Sob, in "Electronic Band Structure and Its Applications: Proceedings, Kanpur, India 1986" (M. Yussouff, Ed.). Lecture Notes in Physics, Vol. 283, pp. 1–57, Springer-Verlag, Berlin, 1987.
185. TBLMTO homepage. Available at: http://www.mpi-stuttgart.mpg.de/andersen/.
186. O. K. Andersen and O. Jepsen, *Phys. Rev. Lett.* 53, 2571 (1984).
187. D. W. Boukhvalov, E. Z. Kurmaev, A. Moewes, D. A. Zatsepin, V. M. Cherkashenko, S. N. Nemnonov, L. D. Finkelstein, Y. M. Yarmoshenko, M. Neumann, V. V. Dobrovitski, M. I. Katsnelson, A. I. Lichtenstein, B. N. Harmon, and P. Kögerler, *Phys. Rev. B* 67, 134408 (2003).
188. O. Gunnarsson, O. Jepsen, and O. K. Andersen, *Phys. Rev. B* 27, 7144 (1983).
189. M. R. Pederson and S. N. Khanna, *Chem. Phys. Lett.* 307, 253 (1999).
190. M. R. Pederson and K. A. Jackson, *Phys. Rev. B* 41, 7453 (1990).
191. K. A. Jackson and M. R. Pederson, *Phys. Rev. B* 42, 3276 (1990).
192. NRLMOL homepage. Available at: http://est-www.nrl.navy.mil/~nrlmol.
193. M. R. Pederson and C. C. Lin, *Phys. Rev. B* 35, 2273 (1987).
194. M. R. Pederson and K. A. Jackson, *Phys. Rev. B* 43, 7312 (1991).
195. A. A. Quong, M. R. Pederson, and J. L. Feldman, *Sol. Stat. Comm.* 87, 535 (1993).

196. M. R. Pederson, B. M. Klein, and J. Q. Broughton, *Phys. Rev. B* 38, 3825 (1988).
197. D. V. Porezag and M. R. Pederson, *Phys. Rev. B* 54, 7830 (1996).
198. A. Briley, M. R. Pederson, K. A. Jackson, D. C. Patton, and D. V. Porezag, *Phys. Rev. B* 58, 1786 (1998).
199. D. Porezag and M. R. Pederson, *Phys. Rev. A* 60, 2840 (1999).
200. M. R. Pederson, D. V. Porezag, J. Kortus, and D. C. Patton, *Phys. Stat. Sol. (b)* 217, 197 (2000).
201. M. R. Pederson, D. V. Porezag, J. Kortus, and S. N. Khanna, *J. Appl. Phys.* 87, 5487 (2000).
202. J. Kortus and M. R. Pederson, *Phys. Rev. B* 62, 5755 (2000).
203. M. R. Pederson, J. Kortus, and S. N. Khanna, in "Cluster and Nanostructure Interfaces" (P. Jena, S. N. Khanna, and B. K. Rao, Eds.), pp. 159–173, World Scientific Publishing, Singapore, 2000.
204. J. Kortus, C. S. Hellberg, M. R. Pederson, and S. N. Khanna, *Eur. Phys. J. D* 16, 177 (2001).
205. M. R. Pederson, J. Kortus, and S. N. Khanna, *J. Appl. Phys.* 91, 7149 (2002).
206. J. Kortus, T. Baruah, M. R. Pederson, S. N. Khanna, and C. Ashman, *Appl. Phys. Lett.* 80, 4193 (2002).
207. P. Bobadova-Parvanova, K. A. Jackson, S. Srinivas, and M. Horoi, *Phys. Rev. B* 66, 195402 (2002).
208. M. R. Pederson, N. Bernstein, and J. Kortus, *Phys. Rev. Lett.* 89, 097202 (2002).
209. J. Kortus, T. Baruah, N. Bernstein, and M. R. Pederson, *Phys. Rev. B* 66, 092403 (2002).
210. T. Baruah and M. R. Pederson, *Chem. Phys. Lett.* 360, 144 (2002).
211. T. Baruah and M. R. Pederson, *Int. J. Quant. Chem.* 93, 324 (2003).
212. J. Kortus, M. R. Pederson, T. Baruah, N. Bernstein, and C. S. Hellberg, *Polyhedron* 22, 1871 (2003).
213. H. Hellmann, "Einführung in die Quantentheorie." Deuticke, Leipzig, 1937.
214. R. P. Feynman, *Phys. Rev.* 56, 340 (1939).
215. P. Pulay, *Mol. Phys.* 17, 197 (1969).
216. K. Park, M. R. Pederson, and C. S. Hellberg, *Phys. Rev. B* 69, 014416 (2003).
217. K. Park, M. R. Pederson, S. L. Richardson, N. Aliaga-Alcalde, and G. Christou, *Phys. Rev. B* 68, 020405 (R) (2003).
218. J. M. Soler, E. Artacho, J. D. Gale, A. Garcia, J. Junquera, P. Ordejón, and D. Sánchez-Portal, *J. Phys. Condens. Matter* 14, 2745 (2002).
219. Siesta homepage. Available at: http://www.uam.es/siesta.
220. P. Ordejón, *Comput. Mater. Sci.* 12, 157 (1998).
221. D. Sánchez-Portal, P. Ordejón, E. Artacho, and J. M. Soler, *Int. J. Quant. Chem.* 65, 453 (1997).
222. A. V. Postnikov, P. Entel, and J. M. Soler, *Eur. Phys. J. D* 25, 261 (2003).
223. N. Troullier and J. L. Martins, *Phys. Rev. B* 43, 1993 (1991).
224. S. G. Louie, S. Froyen, and M. L. Cohen, *Phys. Rev. B* 26, 1738 (1982).
225. D. Sánchez-Portal, E. Artacho, and J. M. Soler, *J. Phys. Condens. Matter* 8, 3859 (1996).
226. J. Junquera, Ó. Paz, D. Sánchez-Portal, and E. Artacho, *Phys. Rev. B* 64, 235111 (2001).
227. G. S. Painter and D. E. Ellis, *Phys. Rev. B* 1, 4747 (1970).
228. A. Rosen, D. E. Ellis, H. Adachi, and F. W. Averill, *J. Chem. Phys.* 65, 3629 (1976).
229. C. Massobrio and F. Ruiz, *Monatshefte Chem.* 134, 317 (2003).
230. D. Gatteschi and L. Pardi, in "Research Frontiers in Magnetochemistry" (C. J. O'Connor, Ed.), pp. 67–86, World Scientific, Singapore, 2003.
231. FLEUR homepage. Available at: http://www.flapw.de.
232. K. L. Taft and S. J. Lippard, *J. Am. Chem. Soc.* 112, 9629 (1990).
233. A. Caneschi, A. Cornia, and S. J. Lippard, *Angew. Chem.* 34, 467 (1995).
234. A. Caneschi, A. Cornia, A. Fabretti, S. Foner, D. Gatteschi, R. Grandi, and L. Schenetti, *Chem.-Eur. J.* 2, 1379 (1996).
235. G. L. Abbati, A. Cornia, A. C. Fabretti, W. Malavasi, L. Schenetti, A. Caneschi, and D. Gatteschi, *Inorg. Chem.* 36, 6443 (1997).
236. B. Pilawa, R. Desquiotz, M. T. Kelemen, M. Weickenmeier, and A. Geisselmann, *J. Magn. Magnet. Mater.* 177, 748 (1997).
237. S. P. Watton, P. Fuhrmann, L. E. Pence, A. Caneschi, A. Cornia, G. L. Abbati, and S. J. Lippard, *Angew. Chem.* 36, 2774 (1997).
238. A. Caneschi, A. Cornia, A. C. Fabretti, and D. Gatteschi, *Angew. Chem.* 38, 1295 (1999).
239. J. van Slageren, R. Sessoli, D. Gatteschi, A. A. Smith, M. Helliwell, R. E. P. Winpenny, A. Cornia, A.-L. Barra, A. G. M. Jansen, E. Rentschler, and G. A. Timco, *Chem. Eur. J.* 8, 277 (2002).
240. E. Rentschler, D. Gatteschi, A. Cornia, A. C. Fabretti, A.-L. Barra, O. I. Shchegolikhina, and A. A. Zhdanov, *Inorg. Chem.* 35, 4427 (1996).
241. A. Lascialfari, Z. H. Jang, F. Borsa, D. Gatteschi, A. Cornia, D. Rovai, A. Caneschi, and P. Carretta, *Phys. Rev. B* 61, 6839 (2000).
242. E. K. Brechin, O. Cador, A. Caneschi, C. Cadiou, S. G. Harris, S. Parsons, M. Vonci, and R. E. P. Winpenny, *Chem. Commun.* 17, 1860 (2002).
243. G. L. Abbati, A. Cornia, A. C. Fabretti, A. Caneschi, and D. Gatteschi, *Inorg. Chem.* 37, 1430 (1998).
244. A. L. Dearden, S. Parsons, and R. E. P. Winpenny, *Angew. Chem. Int. Ed. Engl.* 40, 151 (2001).
245. A. V. Postnikov, S. G. Chiuzbăian, M. Neumann, and S. Blügel, *J. Phys. Chem. Solids* 65, 813 (2004).
246. A. V. Postnikov, J. Kortus, and S. Blügel, *Molecular Physics Reports* 38, 56 (2003).
247. A. Müller, C. Beugholt, P. Kögerler, H. Bögge, S. Bud'ko, and M. Luban, *Inorg. Chem.* 39, 5176 (2000).
248. A. V. Postnikov, M. Brüger, and J. Schnack, http://arXiv.org/abs/cond-mat/0404343.
249. K. M. Mertes, Y. Suzuki, M. P. Sarachik, Y. Paltiel, H. Shtrikman, E. Zeldov, E. Rumberger, D. N. Hendrickson, and G. Christou, *Phys. Rev. Lett.* 87, 227205 (2001).

250. A. L. Barra, A. Caneschi, D. Gatteschi, D. P. Goldberg, and R. Sessoli, *J. Solid State Chem.* 145, 484 (1999).
251. M. Murrie, S. J. Teat, H. Stoeckli-Evans, and H. U. Güdel, *Angew. Chem.* 42, 4653 (2003).
252. S. Schromm, O. Waldmann, and P. Müller (to be published).
253. D. C. Bradley, R. G. Copperthwaite, S. A. Cotton, K. D. Sales, and J. F. Gibson, *J. Chem. Soc. Dalton Trans.* 1, 191 (1973).
254. G. Rajaraman and R. E. P. Winpenny (private communication).
255. E.-C. Yang, D. N. Hendrickson, W. Wernsdorfer, M. Nakano, L. N. Zakharov, R. D. Sommer, A. L. Rheingold, M. Ledezma-Gairaud, and G. Christou, *J. Appl. Phys.* 91, 144 (2002).
256. I. Chiorescu, W. Wernsdorfer, A. Müller, H. Bögge, and B. Barbara, *Phys. Rev. Lett.* 84, 3454 (2000).
257. G. Chaboussant, R. Basler, A. Sieber, S. Ochsenbein, A. Desmedt, R. Lechner, M. Telling, P. Kögerler, A. Müller, and H.-U. Güdel, *Europhys. Lett.* 59, 291 (2002).
258. J. Choi, L. A. W. Sanderson, J. L. Musfeldt, A. Ellern, and P. Kögerler, *Phys. Rev. B* 68, 064412 (2003).
259. J.-F. Létard, J. A. Real, N. Moliner, A. B. Gaspar, L. Capes, O. Cador, and O. Kahn, *J. Am. Chem. Soc.* 121, 10630 (1999).
260. G. Chastanet, A. B. Gaspar, J. A. Real, and J.-F. Létard, *Chem. Commun.* 9, 819 (2001).

CHAPTER 10

Computer Simulation of Magnetic Nanolayer Systems

Willi Schepper
University of Bielefeld, Bielefeld, Germany

CONTENTS

SBN: 1-58883-049-7

Handbook of Theoretical and Computational Nanotechnology
Edited by Michael Rieth and Wolfram Schommers
Volume 7: Pages (563–621)

1. INTRODUCTION

The discovery of af-(antiferromagnetically) coupled Fe/Cr-films by Grünberg [1], the giant magnetic resistance (GMR) effect by Fert [2], and oscillatory exchange coupling by Parkin [3] has triggered immense research activities in the area of thin magnetic films. The oscillations in the coupling constants can be interpreted by the RKKY (Ruderman-Kittel-Kasuya-Yosida) theory, based on the interaction between two magnetic sheets embedded in a nearly free electron gas [4, 5].

GMR and spin valve (SP) elements consist of many af-coupled magnetic layers (e.g., Permalloy, Py: $Fe_{19}Ni_{81}$), which are separated from each other by nonmagnetic spacers (e.g., Cu).

Magnetic tunnel junctions (MTJ) or tunnelling magnetoresistance (TMR) devices consist of three magnetic layers: soft (layer 1; Py) and hard (layer 2; e.g., Co) magnetic layer, af layer (layer 3; e.g., MnIr). The direction of the magnetic moments in the hard magnetic layer is fixed by magnetically coupling to the af layer. The soft magnetic layer is separated by an insulator layer (e.g., Al_2O_3) from the Co layer, but it is coupled to the Co layer by the demagnetizing fields.

The TMR/GMR/SP devices opened up a number of applications, including magnetic random access memory (MRAM), magnetoresistive sensors, single-molecule detection, read heads for storage devices, and new magnetic recording media for storage densities beyond 100 Gb/in^2.

There are numerous theoretical approaches to *ab initio* calculations of the GMR using realistic band structures [6–9]. These methods can provide understanding of the physical origin of the GMR and material parameters influencing the GMR. Phenomenological methods have been very successful too, because they allow a direct comparison with experimental results $\Delta R/R(H_{ext})$. In Co/Cu multilayers the hysteresis and the irreversible decrease of the maximum GMR value has been studied [10] by phenomenologic methods. Micromagnetics has supplied many valuable contributions to the understanding of the magnetic behavior of such complex devices.

2. MICROMAGNETICS

W. F. Brown, Jr. established the principles of micromagnetism [11] and especially coined the term micromagnetics. Micromagnetic papers based on his ideas still appear more than 40 years later. He was a highly respectable scientist in the magnetic community, and a special

issue of *IEEE Transactions on Magnetics* was devoted to his honor for his 75th birthday [12]. A short biography [13] with a photograph is included in that issue. He was active until he turned 80 years old; his last paper was published posthumously [14].

First numerical methods have been applied by Brown and LaBonte [133, 134], Holz [135–137] and Hubert [138, 139]. Holz and Hubert used the Ritz method, minimizing the magnetic Gibb's free energy with respect to the free parameters of a functional ansatz for the distribution of the magnetization. This technique however does not allow a correct treatment of the stray fields.

Most numerical micromagnetic simulations rely on the finite difference method (FD). Finite element (FE) codes can treat more effectively the shape of the magnetic layers and their grain structure. Scientific activities concentrate nowadays on such areas as how to incorporate thermally activated magnetization reversal in the framework of the micromagnetic concept, how to expand the theory for the simulation of the influence of complex microstructures and large-scale systems (MRAM devices, sensors), and how to develop hybrid micromagnetic models including Monte Carlo approaches [15].

2.1. Energy Terms and Effective Field Contributions

All energy terms but the stray field energy depend only locally on the magnetization. There are several contributions to the effective field: external field (H_{ext}) and magnetocrystalline (H_{an}) or surface (H_{surf}) anisotropy at the grid point, the exchange energy (H_{ex}), and additionally, af coupling between two adjacent magnetic layers on both sides of the nonmagnetic spacer layer (H_{af}).

$$
\begin{aligned}
e_{tot} &= e_{ext} + e_{ex} + e_{an} + e_{dem} + \frac{e_{su} + e_{af}}{\Delta z} \\
e_{ext} &= -J_s(\mathbf{H_{ext}} \cdot \mathbf{m}) \\
e_{ex} &= A(\nabla \mathbf{m})^2 \\
e_{an} &= k_u[1 - (\mathbf{k} \cdot \mathbf{m})^2] \\
e_{su} &= k_{su}[1 - (\mathbf{n} \cdot \mathbf{m})^2] \\
e_{af} &= -J_l[1 - (\mathbf{m}_l \cdot \mathbf{m})] - J_q[1 - (m_l \cdot m)^2] \\
e_{dem} &= \frac{-J_s(\mathbf{H_{dem}} \cdot \mathbf{m})}{2} \\
E_{tot} &= \iiint_V (e_{ext} + e_{ex} + e_{an} + e_{dem})\, dV + \iint_S (e_{af} + e_{su})\, dS
\end{aligned}
\tag{1}
$$

The micromagnetic quantities include all three components of the magnetization **M** or the magnetic polarization **J**. J_s, M_s are the corresponding saturation values, $\mathbf{m} = \mathbf{M}/\mathbf{M}_s$ is the unit vector of **M** or **J**, $\mathbf{m}_l$ the unit vector of $\mathbf{M}_l$ in the nearest neighbor cell of the adjacent layer, A the exchange stiffness constant, k_u the anisotropy constant, and **k** the anisotropy direction. The terms (J_l, J_q) are the linear and biquadratic interlayer af coupling constants; negative values (J_l, $J_q < 0$) would refer to the ferromagnetic coupling case.

The volume and surface anisotropy defined here are uniaxial, directed along the easy direction *k* and the normal *n* to the boundary surface. Positive values of k_{su} make the normal surface an easy axis, and negative values make the surface an easy plane. The constants ($k_u \Delta z + k_{su} + J_l + J_q$) can be neglected in the energy equation ($dV = dS\Delta z$), because they are irrelevant in the search for the energy minimum and do not enter into the magnetic field contributions H_{co}; co is used here in place of ex, an, su, af. These values because of the different energy terms e_{co}, can be calculated all by the first equation $H_{co}(e_{co})$ and lead to the following results:

$$
\mathbf{H}_{co} = -\frac{\partial e_{co}}{J_s \partial \mathbf{m}}
$$

$$
\mathbf{H}_{ex} = \frac{2A}{h^2 J_s} \sum_{i=1,\ldots,N_n} \mathbf{m}_i,\ h = \Delta x, \Delta y, \Delta z
$$

$$\mathbf{H}_{an} = \frac{2k_u}{J_s}(\mathbf{k}\cdot\mathbf{m})\mathbf{k}$$

$$\mathbf{H}_{su} = \frac{2k_{su}}{J_s \Delta z}(\mathbf{n}\cdot\mathbf{m})\mathbf{n}$$

$$\mathbf{H}_{af} = -\frac{1}{J_s \Delta z}[J_1 + 2J_q(\mathbf{m}_1\cdot\mathbf{m})]\mathbf{m}_1$$

$$\mathbf{H}_{eff} = \mathbf{H}_{ext} + \mathbf{H}_{ex} + \mathbf{H}_{an} + \mathbf{H}_{su} + \mathbf{H}_{af} + \mathbf{H}_{dem} \tag{2}$$

The exchange field H_{ex} is only determined by the nearest-neighbor cells NN (e.g., 6). Sometimes coupling constants A_1, A_q [16] are used, referring to the stiffness constant A in $\mathbf{H}_{ex}$ instead of the above introduced J_1, J_q for the description of the linear and biquadratic coupling between adjacent layers. The relationship between these two differently defined parameter pairs is $A_{1,q} = -J_{1,q}\Delta z/2$, as the comparison between the fields $\mathbf{H}_{ex}$ and $\mathbf{H}_{af}$ in Eq. (2) shows. Section 3 deals in more detail with the calculation of the demagnetizing field $\mathbf{H}_{dem}$.

2.2. Dynamics of Micromagnetics—LLG

The LLG (Landau-Lifshitz-Gilbert) equation describes a precession motion of the magnetization around the effective field [first term in Eq. (3)]; [140–147] γ is the gyromagnetic ratio. The angle between magnetization and field would never change without consideration of losses or damping [second term in Eq. (3)].

$$\frac{\partial \mathbf{m}}{dt} = -\frac{|\gamma|}{1+\alpha^2}(\mathbf{m}\times\mathbf{H}_{eff}) - \frac{\alpha\gamma}{1+\alpha^2}[\mathbf{m}\times(\mathbf{m}\times\mathbf{H}_{eff})] \tag{3}$$

The gyromagnetic ratio γ and the damping constant α are those constants formerly used for the Gilbert equation ($\gamma_G, \alpha_G \rightarrow \gamma_1\alpha$). α is introduced to describe dissipative phenomena, like the relaxation of magnetic impurities or the scattering of spin waves on lattice defects [17]. At first, Eq. (3) describes the damped precession motion of a single spin at one definite grid cell i, but generally it represents a system of many coupled partial differential equations, because all the spins of the whole simulation volume can enter into the field H_{eff} at that picked up grid cell i. In this sense, Eq. (3) describes the damped precession motion of the whole spin ensemble.

Because α controls the time scale for the magnetization approach to the effective field direction, it governs how quickly a solution converges. Real materials can have values of α in the 0.01 range. However, choosing much larger values ($\alpha = 1$) can provide the same final solution to problems (as $\alpha = 0.01$ solutions) in a fraction of the computation time; this statement has been extensively tested for domain wall systems [142].

2.3. Solving Methods in Micromagnetics

All spins point into the direction of the effective field $\mathbf{H}_{eff}$ [Eq. (2)], when the minimum of the magnetic Gibb's free energy E_{tot} has been obtained, the spin system is in equilibrium. The method to get the hysteresis curve in micromagnetic calculations is therefore the repeated search for the equilibrium state. Starting from the preceding field step, a small change in the external field $\mathbf{H}_{ext}$ alters the energy surface slightly, and the system is a little away from the equilibrium state. An effective minimization method, or only the angle difference between the magnetic moment $\mathbf{m}_i$ at a grid point i and the effective field $\mathbf{H}_{eff,i}$, gives the direction to come to the new equilibrium state.

2.3.1. Effective Field—Successive Overrelaxation

The method is to calculate the effective field $\mathbf{H}_{eff}$ according to Eq. (2) at a grid point (i, j, k) and set the corresponding spin angles (φ, ϑ) into that direction; the angles of $\mathbf{H}_{eff}$ supply the new spin directions: $\varphi_{i,j,k}$, $\vartheta_{i,j,k}$ [18]. Initially, this is only the condition for a local minimum; therefore, the procedure has to be repeated until the average or the maximum of

all the magnetization changes $\Delta\mathbf{m}_{i,j,k}$ drop below a preset limit (e.g., 10^{-6}). The number of iterations can be reduced considerably by using successive overrelaxation with ω parameters ($\omega = 1.97$) close to the stability limit of $\omega = 2$ [19, 20].

$$\alpha_{i,j,k}^{new} := \alpha_{i,j,k}^{old} + \omega(\alpha_{i,j,k}^{new} - \alpha_{i,j,k}^{old})$$

α stands here for the angles φ, ϑ in the polar coordinate system describing the vector $\mathbf{m}$.

2.3.2. Angle Update by the LLG Equation

The LLG equation is often used in micromagnetics not only for simulating the dynamics of small magnetic particles during magnetization reversal (small α calculation) but also for the static case, in which $\mathbf{M}$ points into the direction of $\mathbf{H}_{eff}$ in the equilibrium case ($t \to \infty$, $\mathbf{m} \times \mathbf{H}_{eff} = 0 \to \partial\mathbf{m} = 0$). A large value of the damping constant α is useful to get more quickly to the searched energy minimum. The right side of Eq. (3) directly provides the update of the angles φ, ϑ. The suitable choice of a time step dt is still essential: It must happen on account of the stability criterion for the explicit difference formula of the partial differential equation (LLG) considered up to now. Because of this criterion, the largest possible time step is proportional to the square of the spatial step size (dx, dy, dz), with bitter consequences for the micromagnetic simulation—it is a very time consuming job. In the last few years, great efforts have been made to find a set of implicit difference equations to overcome this limit [21–23].

2.3.3. Energy Minimization

The method is the repeated minimization of E_{tot} for decreasing and increasing the applied field $\mathbf{H}_{ext}$. An effective minimization method has to find the path to a new minimum; when all the spins point again to the new effective field $\mathbf{H}_{eff}$, the system has obtained a new equilibrium state. The conjugate gradient method has been proven to be a very successful procedure; the derivatives of the energy function to the angles (φ, ϑ) are only necessary in addition to the energy value itself. These derivatives can be calculated analytically, and they are also necessary in the FE method [see Eq. (7) in Section 2.3.4]. The subroutines frprmn, brent, mnbrak, linmin from Press et al. [24] can be used for this minimization procedure.

A more advanced method reacts with different step sizes to the slope (small) and the bottom of the energy valley (large). The algorithm is prepared to adapt to these different regions of the energy function by switching between minimization of the energy and the square of its gradient (modified steepest descent and relaxation method MSDR [25]).

2.3.4. Finite Element Technique

The FE method has become a well-established method in many fields of computer-aided engineering, such as structural analysis, fluid dynamics, and electromagnetic field computation. The FE technique was introduced into micromagnetics by Fredkin and Koehler [26–35]. Early acquisitions of the FE technique to micromagnetics go back to Schrefl, Fidler, and Kronmüller [36–40]. There are three main steps during the solution, according to this method:

1. First, the domain of solution is discretized into finite elements. Depending on the dimension of the problem these can be triangles, squares, or rectangles in two dimensions, or tetrahedrons, cubes, or hexahedra for three-dimensional problems. This subdivision process is usually called triangulation.
2. The solution is approximated by piecewise continuous polynomials and the micromagnetic equations are discretized and split into a finite number of algebraic equations. This process shall be illustrated by the interpolation, using a linear function on a triangular finite element consisting of the nodes 1..3. The polarization $\mathbf{J}(r)$ within the element is the weighted average of the magnetization at the nodal points.

$$\mathbf{J}(r) = J_s \sum_{i=1}^{3} \mathbf{m}_i \phi_i(r) \tag{4}$$

The so-called shape function $\phi_i(r)$ equals 1 on node i and decreases linearly to 0 on all other nodes of the element.

$$\phi_1^{\triangle}(r) = \phi_1^{\triangle}(x, y) = a + bx + cy, \quad \phi_1^{\triangle} \begin{pmatrix} x_1, y_1 \\ x_2, y_2 \\ x_3, y_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \rightarrow a, b, c$$

$$\rightarrow \phi_1^{\triangle}(x, y) = \frac{x_2 y_3 - x_3 y_2 + (y_2 - y_3)x + (x_3 - x_2)y}{x_1(y_2 - y_3) + x_2(y_3 - y_1) + x_3(y_1 - y_2)}$$

Thus, the three coefficients (a, b, c) of ϕ_1 are determined by the geometry of the triangular finite element (three nodes); the same applies to the shape functions ϕ_2, ϕ_3. The four coefficients (a, b, c, d) of a square finite element $[\Phi_i^{\square}(x, y) = a_i + b_i x + c_i y + d_i xy]$ are also determined by the geometry (four nodes), and so on.

3. The aim is now to determine the still-unknown coefficients $\mathbf{m}_i$ of the polynomials in Eq. (4). Altogether there are N coefficients $\mathbf{m}_i$, where N is the total number of nodal points.

Micromagnetic FE calculations start from the discretization of the total magnetic Gibb's free energy [Eq. (1)]. Polar coordinates ϑ_i, φ_i for the polarization at node i are introduced to satisfy the constraint ($|\mathbf{J}_i| = J_s$), such that

$$\mathbf{J}_i = \begin{pmatrix} J_{x,i} \\ J_{y,i} \\ J_{z,i} \end{pmatrix} = J_s \begin{pmatrix} \sin\vartheta_i \cos\varphi_i \\ \sin\vartheta_i \sin\varphi_i \\ \cos\vartheta_i \end{pmatrix} \tag{5}$$

The principle of the FE method involves transforming the energy equations into an energy functional $E[\mathbf{J}]$:

$$E[\mathbf{J}] = \int e(\mathbf{J})\, dV$$

The functional has a numerical value at each point and is an integral representation throughout the entire volume of the variables that are functions of the geometry, material properties, the potential solution, and its derivatives.

The FE mesh is used to integrate the total magnetic Gibb's free energy [Eq. (1)] over the entire magnetic system. The energy integral is then replaced by a sum over cells (finite elements), and Eq. (4) is applied to perform the integration of the energy over each cell. When $\mathbf{J}(r)$ is approximated by piecewise polynomial functions on the FE mesh, the energy functional reduces to an energy function with the nodal values of the directions (ϑ_i, φ_i) of the unit vector $\mathbf{m}$ as unknowns. The total energy may be written as

$$E_t = E_t[\mathbf{J}(r)] = E_t(\vartheta_1, \varphi_1, \vartheta_2, \varphi_2, \ldots \vartheta_N, \varphi_N) \tag{6}$$

The minimization of Eq. (6) with respect to the $2N$ variables ϑ_i, φ_i provides an equilibrium distribution of the polarization $\mathbf{J}$. The necessary condition for the minimum leads to exactly $2N$ equations $\partial E_t/\partial\vartheta_i = \partial E_t/\partial\varphi_i = 0$ for the $2N$ unknowns. The minimization may be effectively performed using a conjugate gradient method that requires the gradients of the energy to select the search directions:

$$\frac{\partial E_t}{\partial \vartheta_i} = -V_i \mathbf{H}_{\mathrm{eff},i} \cdot \frac{\partial \mathbf{J}_i}{\partial \vartheta_i}, \quad \frac{\partial E_t}{\partial \varphi_i} = -V_i \mathbf{H}_{\mathrm{eff},i} \cdot \frac{\partial \mathbf{J}_i}{\partial \varphi_i} \tag{7}$$

where V_i is the volume of a box surrounding the nodal point i [41]. Equation (5) is used to supply the derivatives with respect to the angles.

2.4. Some General Aspects of Modeling

Physical modeling of magnetic materials should give insights into the basic processes involved and should be able to extrapolate results to new situations that the models were not

necessarily intended to solve [12]. Thus, for example, if a model is designed to describe a static magnetization curve (minimization of the magnetic Gibb's free energy; Section 2.3.3), it should also be able to describe aspects of magnetization dynamics (Section 2.2). Another aspect is, if the parameter set is adjusted to a curve of a better-understood large device (e.g., a TMR sensor $1 * 1\ \mu m^2$), the model should be able to describe the change in the characteristic for smaller dimensions. We could show that the experimentally observed shift and anisotropy of the characteristic in the smallest TMR devices at the moment ($50 * 50\ nm^2$) [42] could be understood in this sense with the parameter set fitted to the larger devices. The much larger stray fields of the smaller layers are responsible for both the observed shift and the asymmetry [43] of the TMR characteristic.

2.5. Access to Micromagnetic Programs

Three micromagnetic programs of the FD type [44–46] have been available for our investigations in addition to our own code. No free accessible FE program exists at the moment, but there are several research groups using proprietary codes [47–50].

3. DEMAGNETIZING FIELD

The magnetostatic field H_{dem} is a long-range interaction: the magnetic moments of the whole simulation volume enter into its calculation. It is still the most time-consuming part of the micromagnetic simulation, despite the applying tricks such as FFT (Section 3.1).

3.1. Convolution Procedure—FFT

The resulting magnetic field at the grid point i depends on the spins of all j_N cells, where N denotes the number of grid points in the discretization of the whole simulation volume.

$$\mathbf{H}_i(\mathbf{r}_i) = \sum_{j=1}^{j_N} \mathbf{H}_{ij}; \quad \mathbf{H}_{ij} = -4\pi M_s \mathbf{N}(\mathbf{r}_i - \mathbf{r}_j) \cdot \mathbf{m}_j \tag{8}$$

$$\mathbf{N} = \begin{pmatrix} N_{xx} & N_{xy} & N_{xz} \\ N_{yx} & N_{yy} & N_{yz} \\ N_{zx} & N_{zy} & N_{zz} \end{pmatrix} \tag{9}$$

The sum procedure [Eq. (8)] has to be done for each grid point i; therefore, the direct computation of the demagnetizing field from the magnetic volume and surface charges scales with j_N^2 in storage and computation time. The demagnetizing tensor $\mathbf{N}$ in Eq. (9) contains three rows and three columns. It is not necessary to recalculate all tensor elements $N(r_i - r_j)$ for every new field $H_x(r_i)$. The tensor elements depend on Eq. (8) only from the spatial difference $\mathbf{r}_i - \mathbf{r}_j$; that is, for a two-dimensional grid (indices: i_x, i_y), the same demagnetizing factor $N(x, y) = N(3, 2)$ enters into the calculation of, for example, $H_{i_x, i_y} = H_{2,5}$ from the cell (5,7), as for $H_{4,6}$ from (7,8), because $\Delta x = 3 = 5 - 2 = 7 - 4$ or $\Delta y = 2 = 7 - 5 = 8 - 6$. Therefore, the computation time can be already drastically reduced, if initially the N components are calculated and stored in a data array of fourfold (2^2) the number of grid points in the two-dimensional case, or eightfold (2^3) the number of grid points in the three-dimensional case. The factor for the extension of the required array size is really a little smaller for the two-dimensional square case (number of grid points i_N^2), $4 * (1 - 1/2i_N)^2 = 3.75$ ($i_N = 16$) to be precise.

Neglecting the term $j = i$ in Eq. (8) is only necessary for the dipole approximation (Section 3.2), which in this case becomes $r = 0$ in Eq. (10) and $N \to \infty$. The term can be included for the more advanced methods described in Section 3.3. The enormous effort of the convolution step [Eq. (8)] can be drastically reduced by the fast Fourier transformation (FFT) [51, 52]. A brief outline of this extremely successful method in micromagnetics shall be presented here as follows. The convolution of two functions in Eq. (8), at first restricted

to one dimension, denoted $N * m$, is defined by

$$N * m = \sum_{j=1}^{j_N} N(x_i - x_j) m(x_j)\, dx_j \Longleftrightarrow \mathbf{N}(\omega_i) * \mathbf{m}(\omega_i)$$

The convolution theorem says that the Fourier transform of the convolution is just the product of the individual Fourier transforms.

The convolution theorem is also valid in the three-dimensional case of Eq. (8). Therefore, the following three steps are necessary for calculating the demagnetizing fields making use of the convolution theorem:

1. The Fourier transforms of N and m have to be calculated, **N** fortunately only once at the beginning of a simulation project, **m** always after each update [e.g., after applying $\partial\mathbf{m}$ in Eq. (3) in the whole simulation volume].
2. The product $\mathbf{N} * \mathbf{m}$ of the Fourier transforms **N** (stored initially) and **m** (calculated again and again) has to be determined.
3. The product $\mathbf{N} * \mathbf{m}$ must be replaced by its inverse Fourier transform, which leads directly to the demagnetizing fields $\mathbf{H}_i$ disregarding prefactors.

The subroutines rlft3, fourn from Press et al. [24] can be used for the three-dimensional convolution process which has to be done after each update of the spin distribution (new **m**). The scaling rule for Eq. (8) is then $j_N \ln j_N$ instead of j_N^2. FFT is already faster if the simulation lattice exceeds a very small $7 * 7$ mesh ($j_N = 49$) [20]. FFT is an extremely successful method and has enabled to simulate interesting sensor devices. Unfortunately, the algorithm is restricted to a regular lattice, and it cannot be applied to a mesh with a different step size or the unstructured mesh of the FE algorithm.

3.2. Dipole Approximation

The influence of the demagnetizing fields can be calculated by the dipole approximation, in which each cell is replaced by a dipole of strength $M_s V_c$ at its center [53, 54].

$$-4\pi\mathbf{N}(x, y, z) = \frac{3V_c}{r^5}\begin{pmatrix} x^2 - r^2/3 & xy & xz \\ xy & y^2 - r^2/3 & yz \\ xz & yz & z^2 - r^2/3 \end{pmatrix} \qquad (10)$$

$$r = \sqrt{x^2 + y^2 + z^2};\quad V_c = \Delta x \Delta y \Delta z$$

The term V_c is the cell volume and Δx its length, with its width Δy, and its thickness Δz. This is a tensor matrix symmetrical with respect to the diagonal. All the tensor components can be deduced from only the first two elements of the first column. First, the third element of the first column can be obtained by the relation $N_{xz}(x, y, z) = N_{xy}(x, z, y)$. The elements of the second column can be derived from the first on account of the valid relations: $N_{yx}(x, y, z) = N_{xy}(x, y, z)$, $N_{yy}(x, y, z) = N_{xx}(y, x, z)$, $N_{yz}(x, y, z) = N_{xz}(y, x, z)$, and so on.

Equations (8) and (10) can also be applied to get the field outside of a microsphere $r > R$, uniformly magnetized along the x-direction, R is the sphere radius [54, Section 6.1.2]. Only the dipole volume must be replaced by the sphere volume $V_c = 4\pi R^3/3$. We return to some interesting results of Eq. (10) in Section 12.

3.3. Analytical Approach by Advanced Methods

The calculation of the elements of the demagnetizing tensor **N** can be done in the FD algorithm by advanced methods based on the analytic solution for a homogeneously magnetized rectangular prism (grid cell) [53, 55–60]. The demagnetizing fields in these methods are much more precise in comparison with the dipole approximation (Section 3.2).

3.3.1. Magnetic Potential and Surface Charges

The method of calculation of the demagnetizing tensor elements in the case of a ferromagnetic rectangular prism homogeneously magnetized in the x-direction is outlined below.

The origin of a Cartesian coordinate system is defined at the center of this prism. More specifically, it is assumed, as in Ref. [56], that the prism extends over the volume $-a < x < a$, $-b < y < b$, and $-c < z < c$ (see Fig. 1). If this prism is saturated along x, a surface charge is created on its faces $x = \pm a$. The potential resulting from this charge can be calculated by integrals on these surfaces, and the magnetic field is the gradient of that potential.

$$\psi = M_s \int_{-c}^{c} \int_{-b}^{b} \left[\frac{1}{r(-x, y', z')} - \frac{1}{r(x, y', z')} \right] dy' dz' \tag{11}$$

$$\begin{aligned} r(x, y', z') &= \sqrt{(x+a)^2 + (y-y')^2 + (z-z')^2} \\ \mathbf{H}_{\text{dem}}(x, y, z) &= -\nabla\psi = -4\pi M_s \, \mathbf{N} \cdot \mathbf{m} \end{aligned} \tag{12}$$

where ψ is the magnetic potential [Eq. (11)], the components of the demagnetizing field are produced by Eq. (12); $\mathbf{N}_x$ corresponds to the first column of the demagnetizing tensor $\mathbf{N}$ [Eq. (9)].

$$\begin{aligned} f(x,y,z) &= \arctan \frac{(y+b)(z+c)}{(x+a)\sqrt{(x+a)^2+(y+b)^2+(z+c)^2}} \\ g(x,y,z) &= \ln \frac{\sqrt{(x+a)^2+(y+b)^2+(z+c)^2}+z+c}{\sqrt{(x+a)^2+(y+b)^2}} \\ 4\pi N_{xx}(x,y,z) &= f(-x,y,z)+f(x,y,-z)+f(-x,y,-z)+f(x,y,z)+f(x,-y,z) \\ &\quad +f(-x,-y,z)+f(x,-y,-z)+f(-x,-y,-z) \\ 4\pi N_{xy}(x,y,z) &= g(-x,y,z)+g(-x,y,-z)+g(x,-y,z)+g(x,-y,-z) \\ &\quad -g(x,y,z)-g(x,y,-z)-g(-x,-y,z)-g(-x,-y,-z) \end{aligned} \tag{13}$$

$$N_{xz}(x,y,z,a,b,c) = N_{xy}(x,z,y,a,c,b)$$

$$H_x(x,y,z) = -4\pi M_s N_{xx}, \quad H_y(x,y,z) = -4\pi M_s N_{xy}, \quad H_z(x,y,z) = -4\pi M_s N_{xz} \tag{14}$$

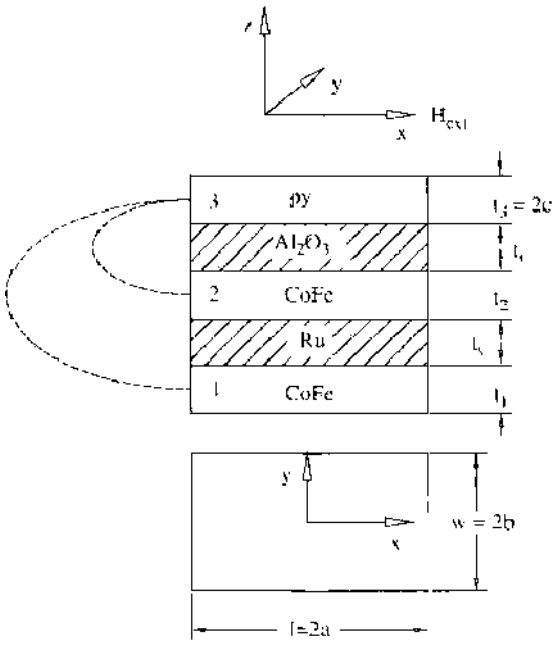


Figure 1. Sketch of a TMR device with the coordinate system. H_{ext} points into the x direction, coordinate axes (x, y) in the layer (length l, width w), axis z in the direction of the layer stack (thicknesses t_i). Dimensions $(2a, 2b, 2c)$ for the calculation of the demagnetizing field of the sensor layer (Py) according to Eqs. (13) and (14). The magnetic behavior of the sensor layer three is determined not only by its own stray field but also by the foreign stray fields (dotted field lines) of the AAF (layers 1, 2)

The tensor component N_{xx} results from Eq. (13) as a sum over eighth terms. The representation $g(x, y, z)$ is easier for the programming code, and the arctan-function was used here [43] instead of $G(\mathbf{r}|a, b, c) = \ln[\sqrt{(a \cdot x)^2 + (b - y)^2 + (c - z)^2} + b - y]$, and the arccot-function in the original paper [56]. The z-component in Eq. (13) N_{zz} did not have to be recalculated, as it could be reduced to the already calculated element N_{yy} by permuting the variables and cell dimensions. The still remaining tensor components for the magnetization in the y- and z-direction [second and third column in Eq. (9)] can be obtained in a similar way; for example, $N_{yy}(x, y, z, a, b, c) = N_{xx}(y, x, z, b, a, c)$. We will come back to some interesting results of Eqs. (13) and (14) for TMR devices in Section 11.

It can be concluded from the ln-terms in Eq. (13) that the magnetic stray field components (H_y, H_z) have a logarithmic singularity at each corner of the rectangular prism, and the enormous increase of the stray field H_y at the four corners is shown in Fig. 4 (left diagram). Naturally, this singularity is a consequence of the micromagnetic continuum approach, and it would be avoided when considering discrete atomic moments. As a result of the singularity, the magnetization at the corner is oriented parallel to the diagonal of a square grid cell.

It should be still remarked that Eq. (13) also can simply be used to calculate directly the own stray field within a magnetic layer; in this case the dimensions of the grid cell (2a,2b,2c) must only be identified with the layer dimensions (l, w, t) (see Fig. 1 for the explanation of the dimensions). This simple calculation is not possible for the more advanced methods [53, 57] described in Section 3.3.2. The corresponding demagnetizing factors cannot be used directly within the cell because of the additional average process. Therefore, the time-consuming convolution process [Eq. (8)] is absolutely necessary in this case to calculate the own stray field of a magnetic layer.

3.3.2. Spatial Average of Stray Fields

The so-calculated magnetic field caused by a source cell j in Eq. (8) still depends on the position within the target cell i. Newell and Schabes et al. [53, 55, 57] have introduced, therefore, an additional average of the stray field over the target cell. The typical fourfold integrals show the following equations for the tensor elements N_{xx}, N_{xy}:

$$4\pi V_c N_{xx} = 2f(X, Y, Z) - f(X + 2a, Y, Z) - f(X - 2a, Y, Z); \quad V_c = \Delta x \Delta y \Delta z = 8abc$$

$$4\pi V_c N_{xy} = g(X - a, Y + b, Z) - g(X + a, Y + b, Z) - g(X - a, Y - b, Z) - g(X + a, Y - b, Z)$$

$$f(X, Y, Z) = \int_{Z-c}^{Z+c} \int_{Y-b}^{Y+b} \int_{-c}^{c} \int_{-b}^{b} \frac{dy'\, dz'\, dy\, dz}{\sqrt{X^2 + (y - y')^2 + (z - z')^2}}$$

$$g(X, Y, Z) = \int_{Z-c}^{Z+c} \int_{X-a}^{X+a} \int_{-c}^{c} \int_{-b}^{b} \frac{dy'\, dz'\, dx\, dz}{\sqrt{x^2 + (y - y')^2 + (z - z')^2}}$$

The equations correspond to the case of a ferromagnetic rectangular prism homogeneously magnetized and saturated in the x-direction. The first equation reduces the integration over two faces with different x coordinates to only one integrand function; y', z' are the coordinates at the cell faces, and the integration considers the integration over the magnetic surface charges in the source cell. The field component H_x resulting only from this integration [(Eq. 13)] would still depend on the coordinates y, z of the target cell; therefore, an additional integration over y, z has been introduced by Schabes and Aharoni [57]. Unfortunately, their results are restricted to the case of cubes as grid cells. The formula [Eq. (15)] for the noncubic case, including different edge lengths, has been represented later by Newell and coworkers [53].

$$f(x, y, z) = \frac{y}{2}(z^2 - x^2)\,\mathrm{arsinh}\frac{y}{\sqrt{x^2 + z^2}} + \frac{z}{2}(y^2 - x^2)\,\mathrm{arsinh}\frac{z}{\sqrt{x^2 + y^2}} - xyz\arctan\frac{yz}{xR} - (2x^2 - z^2)\frac{\sqrt{x^2 + z^2}}{6} + \frac{x^3}{3} + \frac{R}{6}(2x^2 - y^2 - z^2)$$

$$-(2x^2-y^2)\frac{\sqrt{x^2+y^2}}{6}+\frac{y}{2}x^2\,\operatorname{arsinh}\frac{y}{x}+\frac{z}{2}x^2\,\operatorname{arsinh}\frac{z}{x}$$

$$g(x,y,z)=xyz\,\operatorname{arsinh}\frac{z}{\sqrt{x^2+y^2}}+\frac{y}{6}(3z^2-y^2)\,\operatorname{arsinh}\frac{x}{\sqrt{y^2+z^2}}+\frac{y^3}{6}\,\operatorname{arsinh}\frac{x}{y}$$
$$-\frac{z^3}{6}\arctan\frac{xy}{zR}-\frac{zy^2}{2}\arctan\frac{xz}{yR}-\frac{zx^2}{2}\arctan\frac{yz}{xR}+\frac{xy}{3}\sqrt{x^2+y^2}$$
$$-\frac{xy}{3}R+\frac{x^3}{6}\,\operatorname{arsinh}\frac{y}{x}+\frac{x}{6}(3z^2-r^2)\,\operatorname{arsinh}\frac{y}{\sqrt{x^2+z^2}}$$

$$\begin{aligned}4\pi V_c N_{xx}(x,y,z)&=8f(x,y,z)-4f(x-\Delta x,y,z)-4f(x+\Delta x,y,z)-4f(x,y-\Delta y,z)\\&-4f(x,y,z-\Delta z)+2f(x,y-\Delta y,z-\Delta z)-4f(x,y+\Delta y,z)\\&+2f(x,y+\Delta y,z-\Delta z)+2f(x,y-\Delta y,z+\Delta z)-4f(x,y,z+\Delta z)\\&+2f(x,y+\Delta y,z+\Delta z)+2f(x+\Delta x,y-\Delta y,z)\\&-f(x+\Delta x,y-\Delta y,z-\Delta z)+2f(x+\Delta x,y,z-\Delta z)\\&+2f(x+\Delta x,y+\Delta y,z)-f(x+\Delta x,y+\Delta y,z-\Delta z)\\&+2f(x+\Delta x,y,z+\Delta z)-f(x+\Delta x,y-\Delta y,z+\Delta z)\\&-f(x+\Delta x,y+\Delta y,z+\Delta z)+2f(x-\Delta x,y-\Delta y,z)\\&+2f(x-\Delta x,y,z-\Delta z)-f(x-\Delta x,y-\Delta y,z-\Delta z)\\&+2f(x-\Delta x,y+\Delta y,z)-f(x-\Delta x,y+\Delta y,z-\Delta z)\\&-f(x-\Delta x,y-\Delta y,z+\Delta z)+2f(x-\Delta x,y,z+\Delta z)\\&-f(x-\Delta x,y+\Delta y,z+\Delta z)\end{aligned}$$

$$\begin{aligned}4\pi V_c N_{xy}(x,y,z)&=2g(x-\Delta x,y,z-\Delta z)-g(x-\Delta x,y+\Delta y,z-\Delta z)\\&-g(x-\Delta x,y-\Delta y,z-\Delta z)-g(x+\Delta x,y-\Delta y,z+\Delta z)\\&+2g(x+\Delta x,y,z+\Delta z)-4g(x+\Delta x,y,z)-4g(x,y,z+\Delta z)\\&+8g(x,y,z)+2g(x+\Delta x,y-\Delta y,z)+2g(x,y-\Delta y,z+\Delta z)\\&-4g(x,y-\Delta y,z)+2g(x+\Delta x,y,z-\Delta z)-4g(x,y,z-\Delta z)\\&-g(x+\Delta x,y-\Delta y,z-\Delta z)+2g(x,y-\Delta y,z-\Delta z)\\&+2g(x-\Delta x,y,z+\Delta z)-4g(x-\Delta x,y,z)-g(x-\Delta x,y-\Delta y,z+\Delta z)\\&+2g(x-\Delta x,y-\Delta y,z)-g(x+\Delta x,y+\Delta y,z+\Delta z)\\&+2g(x+\Delta x,y+\Delta y,z)+2g(x,y+\Delta y,z+\Delta z)-4g(x,y+\Delta y,z)\\&-g(x+\Delta x,y+\Delta y,z-\Delta z)+2g(x,y+\Delta y,z-\Delta z)\\&-g(x-\Delta x,y+\Delta y,z+\Delta z)+2g(x-\Delta x,y+\Delta y,z)\end{aligned}$$

$$N_{ij}(x,y,z,\Delta x,\Delta y,\Delta z)=N_{ij}(x,z,y,\Delta x,\Delta z,\Delta y);\quad V_c=\Delta x\Delta y\Delta z \tag{15}$$

A representation of Newells formulas was used here more effectively for the programming code [Eq. (15)] instead of the complicated fivefold nested representation in the original paper [53]. Good agreement with the FE methods has been achieved using Newells formula [Eqs. (15)] [44]; other programs [45] apply proprietary formulae based on an extension of the Schabes and Aharoni paper [57]. The formula in the original paper of Joseph et al. [56] use for each field component H_x, H_y [Eq. (13)] eight elementary functions (arctan,ln). Newells's more advanced formula [Eqs. (15)] needs for the same task at least 27 calls of the indefinite integral function consisting of each of 10 ($8 \rightarrow 10*27 = 270$) elementary functions (arsinh,arctan,ln); the effort of the Schabes and Aharoni paper [57] is approximately the

same. Fortunately, these quantities must be calculated only once, at the beginning of a micromagnetic simulation session.

The above remarks on the remaining tensor components (second and third column in [Eq. (9)] are also valid in this case; they can be obtained by permuting the corresponding variables: $N_{yy}(x, y, z, \Delta x, \Delta y, \Delta z) = N_{xx}(y, x, z, \Delta y, \Delta x, \Delta z)$ or $N_{yz}(x, y, z, \Delta x, \Delta y, \Delta z) = N_{xz}(y, x, z, \Delta y, \Delta x, \Delta z)$.

3.3.3. Comparison with the Dipole Approximation

It is interesting to ask how the more exact treatment of the interaction energy [Eqs. (13) and (15)] between uniformly magnetized cubes compares to the dipole approximation (Section 3.2). Two rectangular prisms are considered for simplicity; one at the origin ($i = 0$), the other one displaced in the x direction by i lattice constants ($x = i\Delta x$). Δx is the length of the grid cell in the x direction, and the dimensions in the other directions are Δy and Δz. Equation (10) gives for $r = x \to x^2 - r^2/3 = 2x^2/3$ with a field $H_{di} = 2M_s V_c/x^3$ to be compared with $H_{re} = M_s N_{xx}(x, 0, 0)/V_c$ in Eqs. (13) and (15). Equation (16) contains the necessary geometry factors for the calculation of the ratio $H_{di}/H_{re}(i)$; it must still be extended by a factor V_c if the right sides of Eq. (15) ($V_c N_{xx}$) are directly entered in Eq. (16):

$$\frac{H_{di}}{H_{re}}(i) = -\frac{V_c}{\Delta x^3}\frac{2}{i^3 N_{xx}(i\Delta x, 0, 0)}; \quad \frac{V_c}{\Delta x^3} = \frac{\Delta y\, \Delta z}{\Delta x^2} \tag{16}$$

Some numerical results of Eq. (16) are listed in Table 1, which shows the ratio H_{di}/H_{re} versus the displacement index i. The dipole approximation obviously overestimates the field in the cube case; the error is biggest for direct neighboring grid cells ($i = 1$) and decreases rapidly with the cell distance – it already falls below 0.3% for the third grid cell ($i = 3$). The maximum amount of 18.1% (first line) agrees well with the value of 17.9% (second line) determined by Eq. (15) [53], and the values in line 2 agree exactly with those given in the paper of Schabes and Aharoni [57]. The deviations from the dipole approximation are much larger when the thickness of the grid cell is larger (third line) or smaller (lines 4, 5). The dipole approximation obviously underestimates the field in the rectangular case: When the thickness of the grid cell is smaller than its length and width (lines 4, 5), the deviations are just especially large in the case of very thin layers, that is the very interesting case for applications. That means that the dipole approximation really should not be applied in such cases.

3.3.4. Comparison at Large Distances

At large distances ($r = \sqrt{x^2 + y^2 + z^2} \gg a, b, c$), the demagnetizing field must agree with the field of a dipole at the center of the grid cell, and the cell volume $V_c = 8abc$ enters the dipole strength $\mu_c = M_s V_c$. Equation (10) can be useful to check the asymptotic behavior of Eq. (13). This shall be proved in the case of $y = z = 0$, $x \to \infty$. In this limit, r goes in Eq. (10) to x, and $4\pi N_{xx} \to -2V_c/x^3$. The study of the asymptotic behavior of Eq. (13) is a little more complicated. For $a \ll x$, the arguments of the eight arctan-functions in Eq. (13) are entirely $\ll 1$, there are four contributions to Eq. (13) almost cancelling out [sum in

Table 1. Comparison of the field components H_x or the interaction energy densities between two grid cells, as calculated from the dipole approximation [H_{di}, Eq. (10)] and the more realistic models of a rectangular prism homogeneously magnetized in the x-direction (H_{re}).

line	$\Delta z/\Delta x$	$i = 1$	2	3	4	5	6	7
1	1	1.18083	1.01303	1.00264	1.00084	1.00035	1.00017	1.00009
2	1	1.17878	1.02435	1.00521	1.00168	1.00069	1.00034	1.00018
3	2	2.13914	1.35198	1.16291	1.09278	1.05966	1.04152	1.03054
4	0.5	0.73074	0.91919	0.96106	0.97744	0.98536	0.98976	0.99244
5	0.2	0.45852	0.88645	0.94826	0.97053	0.98102	0.98677	0.99026

Table numbers correspond to the ratio H_{di}/H_{re} [Eq. (16)], the displacement of the two prisms is $i\Delta x$, and H_{re} resulting from Eq. (13) (line 1) and the formulae in [Eq. (15)] ([53], lines 2–5). Parameter: Equal length and width values ($\Delta x = \Delta y$) of the grid cell; its thickness Δz is larger in line 3 and smaller in lines 4, 5.

the brackets of Eq. (17)], and the result confirms the already discussed approximation of Eq. (10):

$$4\pi N_{xx} = 4bc\left[\frac{1}{(x+a)^2} + \frac{1}{(a-x)(x-a)}\right] \approx -4bc\frac{4a}{x^3} = -\frac{2V_c}{x^3} \tag{17}$$

3.3.5. Numerical Aspects

The arctan functions in Eq. (13) have been expanded in a power series (Section 3.3.4) to analyze the behavior at large distances; the equation shows also that the values of the tensor components N_{xx}, N_{xy}, N_{xz} are much smaller than the contributing integral terms f,g. Therefore, the numerical values of the demagnetizing fields depend critically on the precision to which the functions f,g are obtained.

These remarks are all the more valid on the more advanced methods described in Section 3.3.2 and Eq. (15). The values of the tensor components are many orders of magnitude smaller than the contributing integral terms. This is the result of a very delicate cancellation of the terms involved to yield the N_{xx}, N_{xy} integrals. Therefore, the different functions have to be calculated to sufficient accuracy. A way out of this dilemma is to use the advanced methods only up to a displacement of, say, 20 grid cells. Beyond it the dipole approximation supplies better results; there are no such cancellation problems with Eq. (10). Table 1 in Section 3.3.3 has shown that the dipole results can be achieved for a displacement of already 7 grid cells.

3.4. Boundary Element Method (BEM)

The boundary element method (BEM) introduced by Fredkin and Koehler [28] can be used in a more general case; for example, in FD with varying step size, and mainly in FE [61, 62];

$$\Delta\psi = \begin{cases} 4\pi\nabla\cdot\mathbf{M} & \text{inside} \\ 0 & \text{outside} \end{cases}; \quad \psi = \begin{cases} \psi_1 + \psi_2 \\ \psi_2 \end{cases} \tag{18}$$

$$\psi_i|_S = \psi_o|_S, \quad \lim_{r\to\infty}\psi = 0$$

$$\left.\frac{\partial\psi_i}{\partial n}\right|_S - \left.\frac{\partial\psi_o}{\partial n}\right|_S = 4\pi\mathbf{n}\cdot\mathbf{M} \tag{19}$$

$$4\pi\ \psi_2(x) = \int_S \psi_1(x')\frac{\partial}{\partial n(x')}\frac{1}{|x-x'|}dS' + \psi_1(x)\left(\frac{\Omega(x)}{4\pi} - 1\right) \tag{20}$$

$$\mathbf{H}_{dem} = -\nabla\psi$$

The magnetic scalar potential Ψ is split for each magnetic layer into two parts $\Psi_1 + \Psi_2$ [Eq. (18)]; Ψ_1 is only valid inside the magnetic layer, and it is obtained by solving the Poisson equation with the Neumann boundary condition [Eq. (19)], using the Galerkin method, which yields a sparse symmetric system of linear equations that can be solved numerically with the biconjugate gradient method. The term Ψ_2 is a solution of the Laplace equation with the Dirichlet boundary condition at the surface S of the magnetic layer, Ψ_2 is determined from Ψ_1 by means of an integral over the surface [Eq. (20)] considering the limiting conditions at large distances ($x \to \infty$), $\mathbf{n}(\mathbf{x})$ is the unit normal at the surface directed out of the magnetic layer, and $\Omega(x)$ is the solid angle subtended by S at $\mathbf{x}$. This method has the advantage that it is unnecessary to discretize the area between the magnetic layers.

3.5. Dynamic Alternating Direction Implicit Method DADI

In the meantime, special methods have been presented, allowing the integration of the Poisson equation even faster than FFT, as the authors [16] argue. Other experts in the field have reported that this statement cannot be accepted generally, in accordance with their experience.

The used dynamic alternating direction implicit method (DADI) is an iterative, operator-splitting technique in which a fictitious time step [Eq. (21)] is added to the Poisson equation [Eq. (18)]:

$$-\frac{\partial \psi}{\partial t} = \Delta\psi - 4\pi\nabla \cdot \mathbf{M} \tag{21}$$

The authors introduced a stable implicit difference scheme to solve the elliptical equation, enlarged with the partial derivation with respect to time. Their formulation relaxes stability constraints, which usually limit the size of the time step. The authors found that only one iteration of DADI is needed for each time step of the LLG equation [Eq. (3)].

4. RECENT PAPERS IN MICROMAGNETICS

A new book [63] and a recent review [64] are very recommendable for people who are interested in this matter in more detail. The standard book of Hubert and Schäfer [17] is also very useful as an introduction to many aspects of domains in magnetic layers. Some recent papers are reported on here to enable access by the reader to the newest literature in this rapidly developing research field.

The switching of small particles, thin-film elements, and nanowires are increasingly important in magnetic storage and magnetoelectronic devices. Recording data densities larger than 10 Gbits/in^2 is possible when the media have thermally stable, magnetically decoupled grains with grain sizes less than 10 nm. To increase data storage density (i.e., to decrease the bit size and to increase the signal-to-noise ratio), higher-symmetry textures and narrow grain size distributions are necessary. Fidler et al. [65] studied the magnetization reversal processes using a three-dimensional hybrid finite element/boundary element micromagnetic model. Typical examples for numerical micromagnetic simulations are shown, in which the role of a granular microstructure precipitates on hysteresis properties, such as in FePt nanocrystals, granular CoCrPtX thin films for longitudinal magnetic recording, and modern bulk rare earth permanent magnets. In nanostructured materials, the coercive field, remanence, and switching properties can be tailored by controlling grain size, texture, shape, defects, and intergranular coupling. Numerical micromagnetic simulations are used to fully understand new magnetic switching phenomena observed in nanosized magnetic particles, granular films, and structures, and for the design of future magnetic recording devices with storaging densities beyond 100 Gb/in^2.

Arrays of magnetic nanowires are possible candidates for patterned magnetic storage media. For these nanowires, and also for other future magnetoelectronic devices, the understanding of domain wall motion and mobility is important for the controlled switching of the nanostructure. Wieser et al. investigated numerically the motion of domain walls in ferromagnetic, cylindrical nanowires by solving the Landau-Lifshitz-Gilbert equation. The typical micromagnetic field terms must be completed by the driving magnetic field [66]. Depending on the nanowire diameter, either transverse domain walls or vortex walls are found. The transverse domain wall is observed for diameters smaller than the exchange length. For low damping, the domain wall mobility decreases with a decreasing damping constant. With increasing diameter, a crossover to a vortex wall enhances the domain wall mobility drastically. The main difference is the dependence on damping: For a vortex wall the domain wall mobility can be drastically increased for small values of the damping constant, up to a factor of $1/\alpha^2$. Porter and Donahue [67] analyzed domain wall motion in thin, narrow strips by micromagnetic simulations. Their result leads to a simplified analytical model. The model accurately predicts the same domain wall velocity as full micromagnetic calculations, including dependence on strip width, thickness, and magnitude of applied field pulse. Domain wall momentum and retrograde domain wall motion are both observed and explained by the analytical model.

Magnetization reversal in a cylindrical ferromagnetic particle seems to be a simple textbook problem in magnetism. However, at a closer look, the magnetization reversal dynamics in a cylinder are far from being trivial. The difficulty arises from the central axis, where the magnetization switches in a discontinuous fashion. Micromagnetic computer simulations allow for a detailed description of the evolution of the magnetic structure on the

subnanosecond time scale. The switching process involves the injection of a magnetic point singularity (Bloch point) into the cylinder [68]. Further point singularities may be generated and annihilated periodically during the reversal process. This results in the temporary formation of micromagnetic drops (i.e., isolated, nonreversed regions). This surprising feature in dynamic micromagnetism is the result of different mobilities of domain wall and Bloch point.

Thiaville et al. [69] study how micromagnetic calculations can be applied to processes that involve a singularity of the magnetization field (Bloch point). Permalloy thin-film disks supporting a vortex magnetic configuration were considered, to allow a comparison with recent experiments. The structure of the Bloch point at rest in the middle of the core of the vortex is studied first, comparing the evolution of the calculation results under decreasing mesh size to analytical results. The reversal of the core of the vortex under a field applied perpendicularly to the disk plane is then investigated. The researchers apply two different procedures to evaluate switching fields and processes: direct micromagnetic time-dependent calculation, and the evaluation of the energy barrier that separates the two orientations of the vortex core in the configuration space, using a path method. Both methods show the occurrence of Bloch points during reversal. Special attention is paid to the extrapolation toward zero mesh size of the numerical results. The calculations are compared to MFM (magnetic force microscopy) experiments [70]. The authors conclude that defects and thermal agitation are likely to assist Bloch-point injection, hence lowering the switching fields.

5. STONER-WOHLFARTH MODEL

5.1. Energy Equations

There are no spatial dependencies (no domains) within a layer neglecting the demagnetizing field. That means, all the spins point into the same direction, and the directions vary only from one layer to the next. In this case, the integration over the volume in Eq. (1) reduces simply to a multiplication with the layer volume (St); thereby, S is the layer area, and t is the layer thickness. The energy per unit area of layer i is then given by

$$\frac{E_i}{S} = -t_i k_{ui}(m_i \cdot m_i)^2 - J_{si} t_i (H_{ext} \cdot m_i) + J_{ii}(m_i \cdot m_i) + J_{qi}(m_i \cdot m_i)^2 + \frac{1}{2} J_{si} t_i \sum_{\substack{j=1 \\ j \neq i}}^{N} (H_{ij}^{(st)} \cdot m_i)$$

The term $H_{ij}^{(st)}$ is the stray field from the layer j into the layer i. The model is basically an extension of the Stoner-Wohlfarth's model [71] in the single-domain limit, including biquadratic coupling and allowing different values both of the directions of the spins as well as of the anisotropy constants in each layer. The external field and the easy axis of the magnetization are in the film plane. For convenience, we chose $\varphi = 0$ if the related magnetization points into the direction of H_{ext}. The energy per unit area of a multilayer structure, consisting of N layers, then is given by

$$\begin{aligned} \frac{E}{S} = &- \sum_{i=1}^{N} t_i k_{ui} \cos^2(\varphi_i - \xi_i) - \sum_{i=1}^{N} H_{ext} J_{si} t_i \cos\varphi_i + \sum_{i=1}^{N-1} J_{ii} \cos(\varphi_i - \varphi_{i-1}) \\ &+ \sum_{i=1}^{N-1} J_{qi} \cos^2(\varphi_i - \varphi_{i+1}) + \frac{1}{2} \sum_{i=1}^{N} J_{si} t_i \sum_{\substack{j=1 \\ j \neq i}}^{N} H_{ij}^{(st)} \cos(\varphi_i - \varphi_j) \end{aligned} \tag{22}$$

where J_i is the linear exchange constant and J_q the quadratic term, k_{ui} is the uniform anisotropy constant, ξ_i the easy direction, and t_i the thickness of layer i. This is already a very realistic consideration of the situation in complex multilayered structures. The angles φ for each individual magnetic layer i are allowed to have different values. Equation (22) allows a large number of degrees of freedom. The conjugate gradient method can be used to minimize this energy equation. The subroutines frprmn, brent, mnbrak, linmin from Press et al. [24] can be used for this minimization procedure. For every new field H_{ext}, the previous angles φ_i shall be used as a new initial distribution, that still can be corrected by the method of largest descent.

If the angles in a multilayer stack are known, the magnetization curve $M(H_{ext})$, and also the GMR or TMR characteristic (Section 6) result. The magnetization M is given by a simple sum over the angles φ_i:

$$M(H_{ext}, \xi) = \sum_{i=1}^{N} M_{si} \frac{t_i}{t_s} \cos \varphi_i; \quad t_s = \sum_{i=1}^{N} t_i \tag{23}$$

The M value in Eq. (23) is dependent on the preset easy direction ξ. Therefore, an average of M over ξ has still to be done in an outer loop, (e.g., within the region $0 \leq \xi \leq \pi$).

5.2. Fast-Solving Algorithm

Often the foreign stray field term in Eq. (22) is neglected. A system of equations of the special tridiagonal form results from the energy Eq. (22) by derivation with respect to the unknown angles:

$$\begin{aligned} \frac{\partial E}{\partial \varphi_i} &= 0 = f_i = H_{ext} J_{si} t_i \sin \varphi_i + 2k_{ui} t_i \sin(\varphi_i - \xi_i)\cos(\varphi_i - \xi_i) + A_i|_{i>1} - A_{i-1}|_{i>1} \\ A_i &= \sin(\varphi_{i+1} - \varphi_i)[J_{li} + 2J_{qi}\cos(\varphi_{i+1} - \varphi_i)] \end{aligned} \tag{24}$$

The ith equation contains two sin-terms except for the two edge equations ($i = 1$, N), as every layer i is coupled to the preceding $(i-1)$ and the following $(i+1)$ sheet. Mathematically, the derivation of every cos-term in the energy equation causes two contributions. The equations are strongly coupled, as a positive term in equation i is followed by a negative contribution in equation $i+1$. The nonlinear system of equations $\vec{\mathbf{f}}(\vec{\varphi}) = 0$ was solved according to an iterative procedure.

$$\begin{aligned} \mathbf{f}' \Delta\vec{\varphi} &= -\vec{\mathbf{f}} \\ \vec{\varphi}_{l+1} &= \vec{\varphi}_l + \omega \Delta\vec{\varphi}; \epsilon = \sum_{i=1}^{N} |\Delta\varphi_i| > 10^{-7} \end{aligned} \tag{25}$$

$$\vec{\varphi} = \begin{pmatrix} \varphi_1 \\ \varphi_2 \\ \vdots \\ \varphi_N \end{pmatrix}; \quad \vec{\mathbf{f}} = \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_N \end{pmatrix}; \quad \mathbf{f}' = \begin{pmatrix} \frac{\partial f_1}{\partial \varphi_1} & \frac{\partial f_1}{\partial \varphi_2} & \frac{\partial f_1}{\partial \varphi_3} & \cdots & \frac{\partial f_1}{\partial \varphi_N} \\ \frac{\partial f_2}{\partial \varphi_1} & \frac{\partial f_2}{\partial \varphi_2} & \frac{\partial f_2}{\partial \varphi_3} & \cdots & \frac{\partial f_2}{\partial \varphi_N} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_N}{\partial \varphi_1} & \frac{\partial f_N}{\partial \varphi_2} & \frac{\partial f_N}{\partial \varphi_3} & \cdots & \frac{\partial f_N}{\partial \varphi_N} \end{pmatrix}$$

$$B_i = \frac{\partial A_i}{\partial \varphi_{i+1}} = J_{li}\cos(\varphi_{i+1} - \varphi_i) + 2J_{qi}[2\cos^2(\varphi_{i+1} - \varphi_i) - 1]$$

$$\frac{\partial f_i}{\partial \varphi_j} = \begin{cases} B_{i-1}|_{i>1} & j = i-1 \\ H_{ext} J_{si} t_i \cos\varphi_i + 2k_{ui} t_i \cos[2(\varphi_i - \xi_i)] - B_i|_{i<N} - B_{i-1}|_{i>1} & j = i \\ B_i|_{i<N} & j = i+1 \\ 0 & \text{otherwise} \end{cases}$$

The Jacobian-matrix $\mathbf{f}'$ consists of N rows, with each row having at most three elements. It has the special tridiagonal form with nonzero elements only on the diagonal plus or minus one column. The two edge equations ($i = 1$, N) have only two elements. The equations can be solved with regard to the special tridiagonal form, using the fast Crout-algorithm [72].

For every iteration step, the system of Eqs. (24) and (25) for the determination of the $\Delta\varphi_i$ values is solved: 50 equations with 50 layers. After this, the angles must be updated according to Eq. (25), with $\omega = 1$. In the first three steps, the ω value was reduced to 0.1 to

avoid slipping in secondary minima. The iterative loop is continued until the sum ϵ over the absolute angle changes $|\Delta\varphi_i|$ drops below a default limit in the corresponding iteration step.

A crucial point in the minimization of the energy according to Eq. (25) is the choice of the start values for $\varphi_i|_{i=1\ldots N}$. It turns out that the simple analytic approach outlined in Section 10.2 can provide start values for the angles that lead to a fast and reliable detection of the minimum energy. For every new field H_{ext}, the previous distribution for the spins was used as a new initial distribution $\vec{\varphi}_i$, which still was corrected by the method of largest descent [Eq. (26)].

$$\vec{\varphi}_{i+1} = \vec{\varphi}_i - \gamma \nabla E_i(\varphi); \quad \nabla E_i = \vec{\mathbf{f}}_i(\varphi) \tag{26}$$

This method is an iterative procedure too. It was stopped if the energy from one iteration step to the other declines no more but increases again. The value of γ was chosen in such a way that the program spent about 10 steps in this loop. The introduction of this step has reduced the sensitivity of the program to the initial values considerably [72].

6. RESISTANCE MODELS FOR TMR AND GMR DEVICES

Using micromagnetic calculations for the spin distribution is only the first step in the simulation of magnetoresistance curves. The second one is a model for the GMR/TMR contribution to the conductance in dependence of the spin angles φ_i.

6.1. Double-Layer Models

6.1.1. TMR

The conductance G_i of the current path at the position i of the element area is proportional to the angle difference $\Delta\varphi_i$ on both sides of the insulator interface (Al_2O_3) in the simplest case.

$$G_i = g_0(1 + p_m \cos\Delta\varphi_i); \quad p_m = p_1 p_2$$

This equation goes back already to Ref. [73]; p_1, p_2 correspond to the effective spin polarization for the tunneling electrons in the magnetic layers on both sides of the insulator. The total conductance results from a sum over all grid cells i.

$$\begin{aligned} \frac{G}{g_0} &= n(1 + p_m \overline{c_\varphi}) = \frac{1}{Rg_0} \\ \frac{\Delta R}{R_\infty} &= p_m \frac{1 - \overline{c_\varphi}}{1 + p_m \overline{c_\varphi}}; \quad \overline{c_\varphi} = \frac{1}{n}\sum_i^n \cos\Delta\varphi_i \end{aligned} \tag{27}$$

The parameter p_m depends directly on the TMR maximum t_m.

$$\Delta\varphi_i = \begin{cases} 0 \\ \pi \end{cases} \rightarrow \overline{c_\varphi} = \begin{cases} 1 \\ -1 \end{cases} \rightarrow \mathrm{TMR} = \begin{cases} 0 \\ t_m \end{cases} \rightarrow t_m = \frac{2p_m}{1 - p_m} \rightarrow p_m = \frac{t_m}{2 + t_m} \tag{28}$$

The TMR characteristic can be calculated with Eqs. (27) and (28), with t_m as a parameter given by experiments. The neglect of the denominator of Eq. (27) ends up in the following simpler equation, which will be commented on in Section 6.1.2.

$$p_m, t_m \ll 1 \rightarrow \frac{\Delta R}{R_\infty} = \frac{t_m}{2}(1 - \overline{c_\varphi}) \tag{29}$$

6.1.2. GMR/SP

In these devices the magnetoresistive effect is based on different scattering processes of the spin-up and spin-down electrons on both sides of the nonmagnetic interface (e.g., Cu). Therefore, the conductance is proportional to the angle difference ($\Delta\varphi_{i,k} = \varphi_{i,k+1} - \varphi_{i,k}$) of

the spins in adjacent magnetic layers in the simplest case:

$$G_i = G_n + G_m\left(1 + g_m \cos^2 \frac{\Delta\varphi_i}{2}\right)$$

The conductance G_n of the nonmagnetic layer is constant because of the conductivity σ_n. The conductance G_i of the magnetic layer contains two parts: a constant one (G_m) and another one dependent on the difference of the magnetization directions of adjacent layers as $\cos^2(\Delta\varphi_i/2)$. We have shown that this approach leads to the following equation [72]:

$$\frac{\Delta R}{R_\infty} = g_m \frac{1 - \overline{c_\varphi^2}}{1 + g_m \overline{c_\varphi^2}}; \quad \overline{c_\varphi^2} = \frac{1}{n}\sum_{i=1}^{n} \cos^2 \frac{\Delta\varphi_i}{2} \tag{30}$$

The GMR characteristic can be calculated with Eq. (30), with the GMR maximum g_m as a parameter given by experiments:

$$\Delta\varphi_i = \begin{cases} 0 \\ \pi \end{cases} \rightarrow \overline{c_\varphi^2} = \begin{cases} 1 \\ 0 \end{cases} \rightarrow \text{GMR} = \begin{cases} 0 \\ g_m \end{cases}$$

The angle dependence corresponds to the af-order parameter $p_{af} = [1 - \cos(\Delta\varphi_i)]/2$ used from Holloway [10] for the GMR-contribution of the cell i.

$$p_{af} = 1 - \cos^2 \frac{\Delta\varphi_i}{2} = \frac{1}{2}(1 - \cos\Delta\varphi_i)$$

The sum of p_{af} over all cells leads to the following simpler equation:

$$g_m \ll 1 \rightarrow \frac{\Delta R}{R_\infty} = \frac{g_m}{2}(1 - \overline{c_\varphi}) = g_m \overline{p_{af}}, \quad \overline{p_{af}} = \frac{1}{2n}\sum_{i}^{n}(1 - \cos\Delta\varphi_i) \tag{31}$$

This often-used equation [10, 74] corresponds directly to Eq. (29) in the TMR case. Both equations result from the neglect of the denominators in Eqs. (27) and (30) and are only valid in the limit of small (t_m, $g_m \ll 1$) values. Actual t_m values of 19% (t_m) or 30% (g_m) can be achieved easily with the present technology and are obviously not $\ll 1$; therefore, today's application of these simpler equations, Eqs. (29) and (31), is in question.

6.2. GMR Multilayers

Mixed multilayers simultaneously containing the first and the second maximum of the af coupling have been investigated experimentally [75, 76] and have been well understood through Stoner-Wohlfarth calculations [72]. A formula for the GMR valid in such complex multilayer systems can be derived compatible with the previous Section 6.1.2. The result shall only be reported here, for the detailed derivation is referred to our previous paper [72]:

$$\begin{aligned} \frac{\Delta R}{g_m R_\infty} &= \frac{1 - \overline{g_\varphi^2}/\overline{g}}{1 + g_n \overline{g_\varphi^2}} \\ g_n \overline{g_\varphi^2} &= \frac{1}{N-1}\sum_{i=1}^{N-1} g_n g_i \cos^2 \frac{\varphi_{i+1} - \varphi_i}{2} \\ g_n \overline{g} &= \frac{1}{N-1}\sum_{i=1}^{N-1} g_n g_i; \quad g_n g_i = \begin{cases} 0.1 & w \\ 0.2 & s \end{cases} \\ g_m &= g_n \overline{g} = \frac{\Delta R}{R_\infty}(H_{ext} = 0) \end{aligned} \tag{32}$$

The layer index i is here used as in Section 5. The factor g_i accounts for the amplitude of the conductivity variation or the relative change of the resistance of one double layer between zero field and saturation: $g_n g_i$ represents the GMR maximum obtained with that af

coupling amplitude g_1, and $g_n g_1$ should be adapted to experimental observed values. In our calculations, we used, for example, 20% for the first (s) and 10% for the second (n) maximum of the RKKY coupling in agreement with experimental results [75]: g_m is the GMR maximum for the mixed multilayer.

In the special case of alternating signs of the angles φ_i and, additionally, neglecting the denominator in Eq. (32), the mean value $\overline{g_\varphi^2}/\overline{g}$ equals the square of the mean value $\overline{c_\varphi}$ of $\cos\varphi$.

$$\varphi_{i+1} = -\varphi_i \rightarrow \varphi_{i+1} - \varphi_i = -2\varphi_i \rightarrow \frac{\overline{g_\varphi^2}}{\overline{g}} = \overline{c_\varphi^2} \rightarrow \frac{\Delta R}{g_m R_\infty} = 1 - \overline{c_\varphi^2} = 1 - \left(\frac{M}{M_s}\right)^2 \qquad (33)$$

In this case, the GMR ratio can be simply expressed by the relative magnetization M/M_s. However, this often-used [75, 77–79] Eq. (33) is no more valid in general ($\overline{g_\varphi^2}/\overline{g} \neq \overline{c_\varphi^2}$), particularly if the conditions (alternating signs of the angles φ_i, neglect of the denominator) are not valid.

6.3. GMR—Quantum Statistical Treatment

We have shown [80, 81] that a combined theoretical approach to the GMR effect in magnetic multilayers is able to provide good agreement with experimentally obtained GMR characteristics, as well as for the GMR maximum g_m. This approach is first based on the quantum statistical theory and the transport properties within the Kubo linear response formalism, where a difficult but necessary task is to find the correct Green function matching at the interfaces. Second, numerical calculations have been performed to explore the magnetic reversal process of each individual magnetic layer of these multilayers on the basis of an extended Stoner-Wohlfarth model [71] in the single domain limit [72]. The Stoner-Wohlfarth model [minimization of the energy [Eq. (22)] by the fast-solving algorithm [Eq. (25)] described in Section 5.2] in addition to the GMR model [Eq. (30)] has been used to fit all simulation parameters to the experimental curves. Thus, the field dependence of the angles between adjacent magnetization vectors is known and can directly be used as input for the quantum statistical calculation of the GMR versus field characteristics. In summary, it could be shown that a quantum statistical treatment, together with the numerically determined orientation of magnetic moments, is a very powerful tool to reliably predict the GMR characteristics even of complex multilayered structures. The calculations can agree perfectly with the experimental curves.

7. AIM OF SIMULATIONS—MAGNETIC SENSOR AND BEAD DESIGN

Micromagnetic modeling is a well-established and useful tool to gain information about the mechanisms ruling the hysteretic properties of technologically interesting magnetic systems. The calculations can help to optimize the molecule design of the beads and the sensor devices [19, 72, 82, 83] and can contribute to answering the essential question of what is going on magnetically. Micromagnetic modeling is best suited to analyzing objects with dimensions that range from a few micrometers down to the nanometer scale.

The modeling algorithm combines essentially two steps: first, either micromagnetics (Section 2) or more simple, an extended Stoner-Wohlfarth model (Section 5), and second a GMR/TMR model (Section 6) coupling the magnetoresistance to the angles of the magnetic moments. Modeling magnetic systems is a very interesting tool in understanding what is going on in complex magnetic systems. That is the important condition to changing the sensor stack and achieving the wanted magnetic characteristic.

The next logical step is to include the signal source. In particular, we have presented some interesting results for the system bead sensor, which is very important for single-molecule detection [83]. Now it is time to have the courage to simulate the signals in a more complex sensor chain, consisting of, for example eight sensors; this will be done in the next future. Such systems are essential candidates for controlling the movement of molecules in small channels.

Another very promising application field of micromagnetics could be the bead design [84, 85]. In an on-chip laboratory, beads with different magnetic characteristics are necessary for the different assignments, especially the process of detection and manipulation or movement of biologically interesting molecules. The magnetic behavior is mostly influenced by the magnetic shell; there are a lot of open questions that must be answered for the new bead generation. Interesting magnetic materials for the bead shells could be Fe_2O_3, Co, FeCo, and FeCoPt, besides the nowadays mainly used magnetite Fe_3O_4.

8. SINGLE-LAYER ELEMENTS

8.1. Stray Fields of Magnetic Platelets

Stray fields of magnetic layers play an important role in the design of the characteristics of magnetic devices. Figs. 2, 3, 4, and 5 show the stray field of a Py platelet, magnetized and saturated in the x-direction. In Fig. 2 the length of the arrows is set constant (left); it represents the magnitude ($|H_{dem}|$) of the field (right). The right figure shows that the stray field is concentrated in a small strip at the left and right edges of the layer. This statement is confirmed in the three-dimensional view of H_x in Fig. 3.

Figure 3 shows a three-dimensional view of the H_x component of the stray field of the Py platelet concentrated at the left and right edges. The left diagram shows the own stray field ($z = 0$); the right one slows the fields at a distance ($z = t$) of the layer thickness t, which are penetrating into the adjacent layers (foreign stray field). The right drawing demonstrates the field decrease with distance.

Figure 4 shows a three-dimensional-view of the H_y and H_z components of the stray field of the Py platelet at a distance ($z = t$) of the layer thickness t, which are penetrating into the adjacent layers (foreign stray field). The left figure shows the large stray field values of H_y at the four corners of the rectangular prism; the simpler formula Eq. (13) shows at the corners logarithmic singularities. The right drawing shows that the component H_z is also concentrated at the left and right edges and is penetrating there as H_x in narrow stripes into the adjacent layers. It shall still be mentioned that $H_z = 0$ for $z = 0$.

Figure 5 shows the $H_x(x/l)$ component of the stray field from a Py platelet. The stray field component H_x is concentrated at the left and right edges (see drawings in Figs. 2 and 3). The small strip of high field values increases with decreasing layer lengths, as the comparison between the broken ($l = w = 200$ nm) and solid ($l = w = 50$ nm) curve demonstrates. The comparison between the solid curve [Eqs. (13)] and the Newell formula [Eqs. (15)] ([53], marked with squares, lattice $50 * 50$) shows that these different formulas [Eqs. (13), (15)] agree very well in the central region of the magnetic layer; the jump is smeared out in the Newell formula [Eq. (15)] because of the additional average over the target cell. The heights

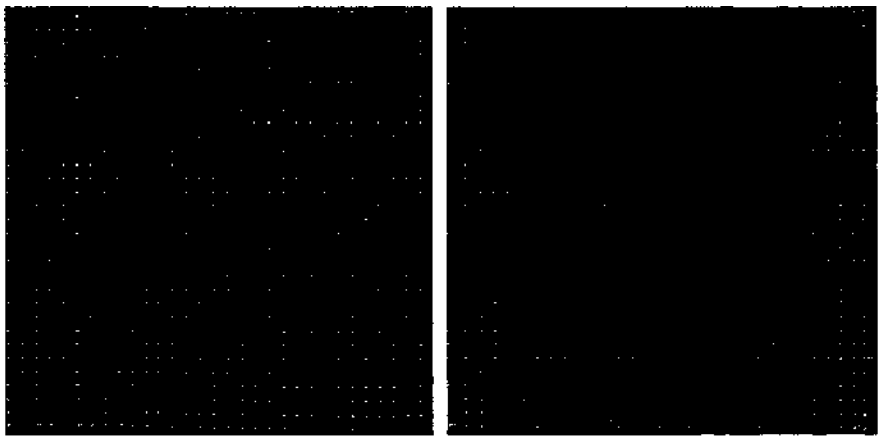


Figure 2. Own stray field according to Eq. (15) of a Py platelet, magnetized, and saturated in the x-direction. The length of the arrows is set constant (left); it represents the magnitude ($|H_{dem}|$) of the field (right). The right figure shows that the stray field is concentrated at the left and right edges. Parameters: $M_s = 800$ emu/cm^3, $l = w = 50$ nm, $t = 3$ nm, $z = 0$, lattice $32 * 32$.

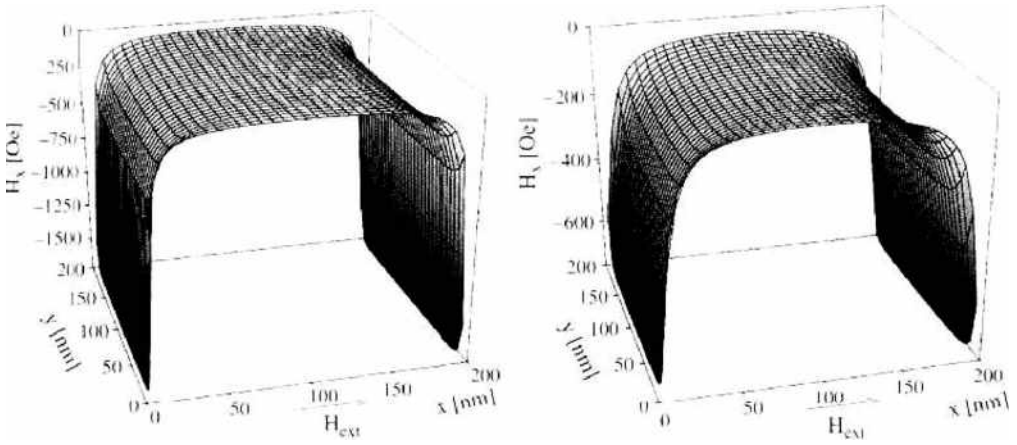


Figure 3. 3D-view of the H_x component of the stray field of a Py platelet according to Eq. (15), magnetized and saturated in the x-direction. The stray field component concentrated at the left and right edges is penetrating into the next layer. The graphs are calculated at a distance $z = 0$ (left) and $z = t = 3$ nm (right), t is the layer thickness. Parameters: $M_s = 800$ emu/cm^3, $l = w = 200$ nm, lattice $32 * 32$.

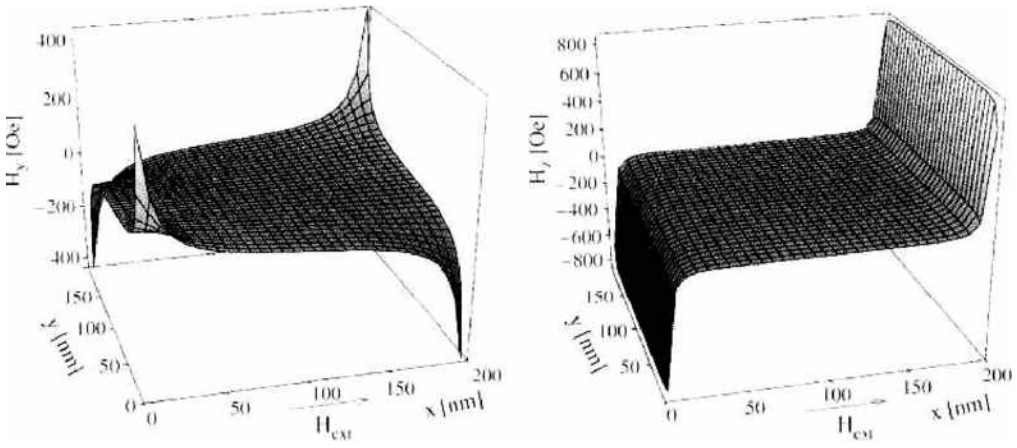


Figure 4. 3D-view of the H_y (left) and H_z (right) components of the stray filed at a distance $z = t$. All the other data as in Fig. 3.

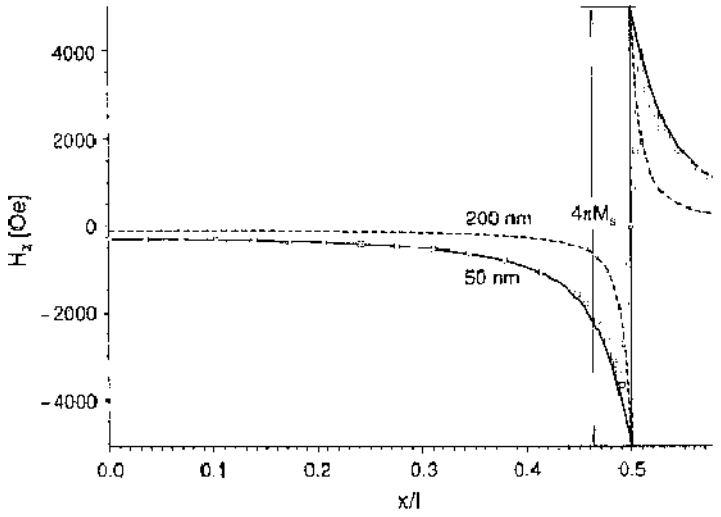


Figure 5. $H_x(x/l)$ component of the stray field at the layer center $y = z = 0$. Py platelet, magnetized and saturated in the x-direction. The stray field component H_x is concentrated at the left and right edges, the region of high field values decreases with increasing layer lengths (comparison between solid and broken curve). The jump in the curve (height $4\pi M_s$) at the position $x = l/2$ is a result of the surface charge at the edge faces: see Eq. (19). Calculation of Eq. (13) ($l = w = 50$ nm: solid, $l = w = 200$ nm: broken) and Eq. (15) (marked with squares, $l = w = 50$ nm, lattice $50 * 50$). Constant parameters: $M_s = 800$ emu/cm^3, layer thickness $t = 3$ nm.

4π of the jump of the curve H_x/M_s at the positions $x = \pm a$, correspond to the surface charges at the edge faces, see Eq. (19). One arctan function reproduces at the position $x = -a$, a jump of $\pi = 2 * \pi/2$ because of the factor $(x + a)$ in the denominator of f_1 in Eq. (13); there are four such contributions with the negative sign of x in the equation for N_{xx}.

8.2. Landau and Diamond Configurations

Thin-film nanoscale elements with a curling magnetic structure (vortex) are a promising candidate for future nonvolatile data storage devices. The magnetic ground state of such devices can be a vortex in dependence on the lateral and vertical size. The magnetization continuously curls around the particle center in such devices, drastically reducing the stray field energy and avoiding domain wall energy. The magnetization in the vortex core turns into the surface normal. Their properties are strongly influenced by the spin structure in the vortex core. The dimensions of the particles are too large to form a single-domain state because it would cost a relatively high stray field energy. However, they are also too small to form domains like those found in macroscopic pieces of magnetic material because the additional cost of domain wall energy cannot be compensated for by the reduction of stray field energy.

8.2.1. Bloch Line Width, Comparison with Experiments

The Landau configuration existing out of four domains occurs in not-too-thin square platelets. In the center of the square the spins must turn into the z-direction because of energetic reasons in a very small core region. Wachowiak et al. [86] used spin-polarized scanning tunneling microscopy on nanoscale iron islands to probe for the first time the internal spin structure of magnetic vortex cores. Using tips coated with a layer of antiferromagnetic chromium, the researchers obtained images of the curling in-plane magnetization around and of the out-of-plane magnetization inside the core region. Okuno et al. [70] have also confirmed the existence of perpendicular magnetization spots by MFM; they were able to visualize the direction of the magnetization, up or down in permalloy dots in the diameter range 0.11 μm.

Figure 6 shows a pronounced four-domain structure with spins parallel to the x- and y-axes in the outer regions with 90° walls between them. In the center region the spins formed into centered circles turn certainly into a vortex (left spin pattern in Figure 6), but they are still aligned in plane ($M_z = 0$). In the center core domain, the spins turn into the out-of-plane direction (z-direction perpendicular to the figure plane), and the in-plane component vanishes, $M_y = 0$ (Fig. 7).

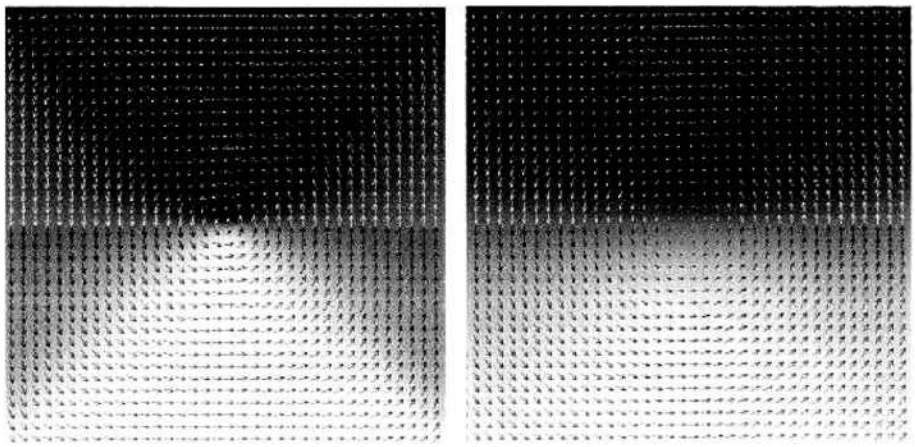


Figure 6. Landau configuration (four domains) of a Py platelet in the plane ($x = y = 100$ nm, $z = 10$ nm), area presented in the left figure is seven times enlarged (edge length shown 28.6 nm). **Parameters:** thickness 20 nm (z-axis perpendicular to the square layer plane), x-axis to the right (length 200 nm), y-axis upward (width 200 nm), layer grid $255 * 255$. Material parameters (Py): $M_s = 800$ emu/cm^3, $k_u = 10^5$ erg/cm^3, $A = 1.05$ $\mu erg/cm$. Energies: $E = 100.4$ perg, $E_{ex} = 63.1$ perg, $E_{demag} = 36.8$ perg, $E_{an} = 0.4$ perg.

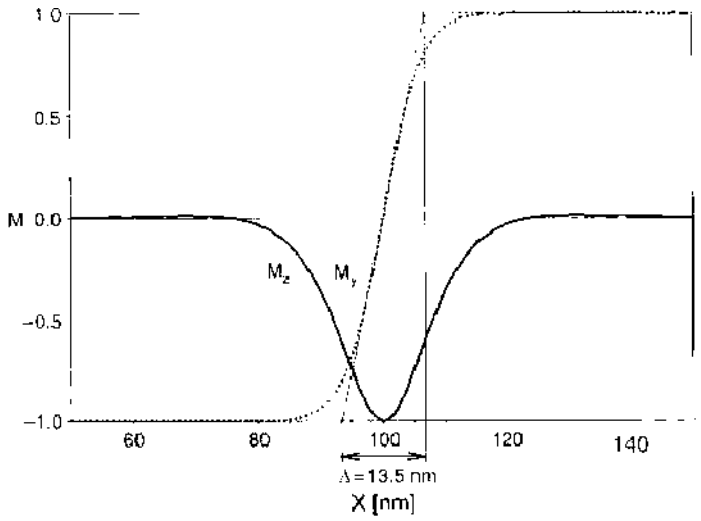


Figure 7. Magnetization components $M_z(x)$ (dotted), $M_y(x)$ (solid), $M_y(x) = 0$, $H_{ex} = 0$, extent $\Delta = 13.5$ nm of the inner domain in the square center ($x = 100$ nm), defined with the slope of the in-plane component M_y. The extent by this definition is a little smaller compared to the half-width of the out-of-plane component M_z.

8.2.2. Theoretical Approach

The first theoretical work was done by Feldtkeller and Thomas [87], who used a simple function $M_z(\rho)$ to find the extent of the cylindrical domain.

$$m_z(\rho) = \exp\left(\frac{-2\rho^2}{b^2}\right)$$

Hubert and Schäfer [17] used at first the generalized ansatz [Section 3.65(C), pp. 263–267] instead of only the first term [87]. The stray field and total energy can be impressed with the ansatz $M_z(\rho)$. The open parameters (b_i, c_i; e.g., $i_m = 6$) were determined by minimizing the total energy (Rayleigh-Ritz method).

$$m_z(\rho) = \sum_{i=1}^{i_m} c_i \exp\left(\frac{-2\rho^2}{b_i^2}\right), \quad \sum_{i=1}^{i_m} c_i = 1$$

$$E_{str} = \frac{\pi}{8} K_d \int_0^\infty (1 - e^{-u}) \left(\sum_{i=1}^{i_m} c_i b_i^2 e^{-u^2 b_i^2} \right)^2 du$$

Another improvement is to take into account the two-dimensional model including the z-dependence, which allows the Bloch line to get wider in the bulk and stay narrow at the surfaces. The exchange energy needs a two-dimensional integration in this case, whereas the stray field energy can be calculated as before.

$$\lambda_{ex} = \sqrt{\frac{A}{K_d}}, \quad K_d = \frac{J_s M_s}{2} \tag{34}$$

$$\Delta = 2\kappa\lambda_{ex} \tag{35}$$

The prefactor 2 considers the diameter instead of the radius. There is additionally a dependence of the layer thickness t (here 20 nm) compared to the exchange length or Bloch line width $\lambda_{ex} = 5.1$ nm. There are two limits for the prefactor: $\kappa = 1.32$ [87] for a small and $\kappa = 1.5$ [17] for a large layer thickness t. The first value has been confirmed by the newer calculations [17]. The simpler ansatz of the first paper overestimates the dependence of the layer thickness t. Their value $\kappa = 2.13$ [87] for $t = 4\lambda_{ex}$ is much too large.

The improved ansatz also shows a z-dependence of the extent Δ, which is larger in the center $z = t/2$ compared with the surface. It shows also the reverse magnetization in the outer region, seen in Fig. 7. This feature compensates a part of the flux welling up from

the core; it could not be described with the original function employed in Ref. [87]. The micromagnetic parameters give the following value for the extent Δ of the inner domain $A = 1.05$ μerg/cm, $M_s = 800$ G, $\kappa = 1.32 \rightarrow \Delta = 13.5$ nm, which agrees very well with the result (13.5 nm) of the micromagnetic calculation (Fig. 7). The results confirm the earlier theoretical predictions that the size and the shape of the vortex core as well as its magnetic field dependence are governed by only two material parameters, the exchange stiffness A and the saturation magnetization M_s, which determines the stray field energy.

8.2.3. Bloch Line Types in a Square Platelet

All spin patterns were from the circular Bloch line type (Fig. 6) until now. That type is not the only possible configuration showing a vortex core in the center. Figure 8 shows a sketch of the main features of the two possible structures, with the circular Bloch line type in the left diagram (Fig. 8a). There also exists another structure, with similar behavior, from the cross Bloch line type, as sketched in Fig. 8(b); Fig. 9 shows the corresponding spin pattern. There are practically no differences for the exchange energy (63.1 $\leftrightarrow$ 56.6) but large differences (36.8 $\leftrightarrow$ 280.7) in the demagnetizing field shown in Fig. 10, in which the arrows represent the magnitude (length) and direction (angle) of H_{dem}. The density of the demagnetizing field is concentrated at the four edges (left drawing) in the circular Bloch line type, whereas in the cross Bloch line type the two axes through the center (right drawing) are also regions of large demagnetizing field values, which is the reason for the three times larger total energy. The energy density is largest in the center domain in both cases.

Table 2 summarizes the results for a square Py platelet $200 * 200 * 20$ nm^3. There are large differences (>3) in the energy for the two Bloch line vortex states, which is really conspicuous—normally the low-remanence states should lie next to each other. There are only insignificant differences in the high-remanence states between C and S (7%), and the flower state shows again the largest energy.

8.2.4. Diamond and Cross-Tie States in a Rectangular Platelet

Figure 11 shows the spin patterns for the diamond (seven domains) and cross-tie (four domains) states, and Fig. 12 for the high-remanence states (three domains) of a rectangular Py platelet of the size $2000 * 1000$ nm^2 with their different demagnetizing fields. Table 3 summarizes the corresponding energy values.

In this case there are only small differences (10%) in the energy values of the two low-remanence vortex states (diamond-state, cross-tie), in contrast to Table 2. The high-remanence states (C, S, flower) occur in the same sequence as in Table 2.

8.2.5. Limit Between Vortex and Single Domain State

Patterned arrays of magnetic materials are of interest for MRAM and magnetic recording. They have been increasingly investigated because the results are fundamental for the development of thin-film elements. Various states of magnetization at remanence have been discussed in the literature. The remanent state of magnetization depends on the geometry of

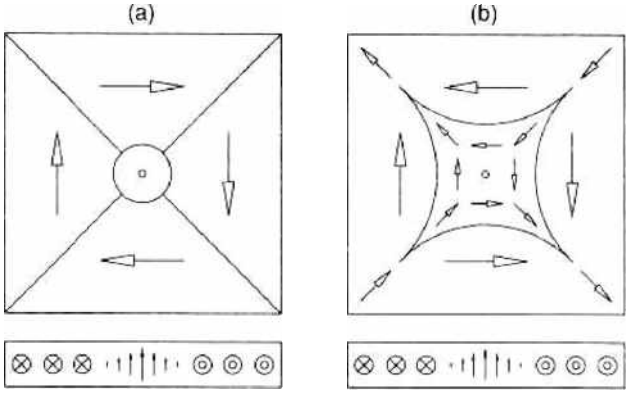


Figure 8. Sketch of a four-domain structure in thin films with a circular (a) and cross (b) Bloch line in the center.

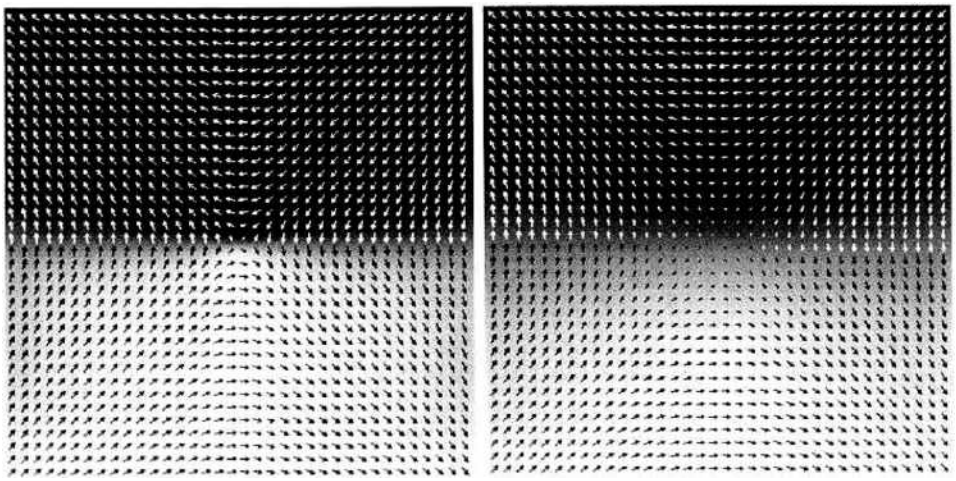


Figure 9. Landau configuration (four domains) of the same Py platelet as in Fig. 6 from the cross Bloch line type. area in the lower figure is seven times enlarged. Energies: E = 337.8 perg, E_{ex} = 56.6 perg, E_{demag} = 280.7 perg, E_{an} = 0.4 perg.

the nanodot, given by the width w and length l, the height (thickness) t, and the aspect ratios. The exchange length λ_{ex} [Eq. (34)] can be used to normalize the geometric dimensions.

Goll et al. used three-dimensional micromagnetic FE modeling to simulate and energetically compare zero-field magnetization patterns of ferromagnetic thin-film elements of square shape. The geometry of the sample (edge length a, thickness t) and its intrinsic material parameters are systematically varied. On the basis of the results, the corresponding phase diagram is set up that identifies a quasihomogeneous single-domain phase (C, S, and flower states) and a vortex phase (Landau state). For the transition between the two phases a material- and edge-length-dependent critical film thickness is found. The numerical results can be confirmed qualitatively, using a simple analytical model calculation. They have set up phase diagrams for a square ferromagnetic thin platelet, depending on both thickness and edge length and thickness and magnetic hardness by means of micromagnetic numerical FEM calculations identifying a multidomain phase and a single-domain phase. The critical layer thickness for which the phase transition between the two phases takes place is shifted to significantly larger values when the hardness Q increases, whereas an increasing edge length a of the platelet reduces the critical layer thickness [88].

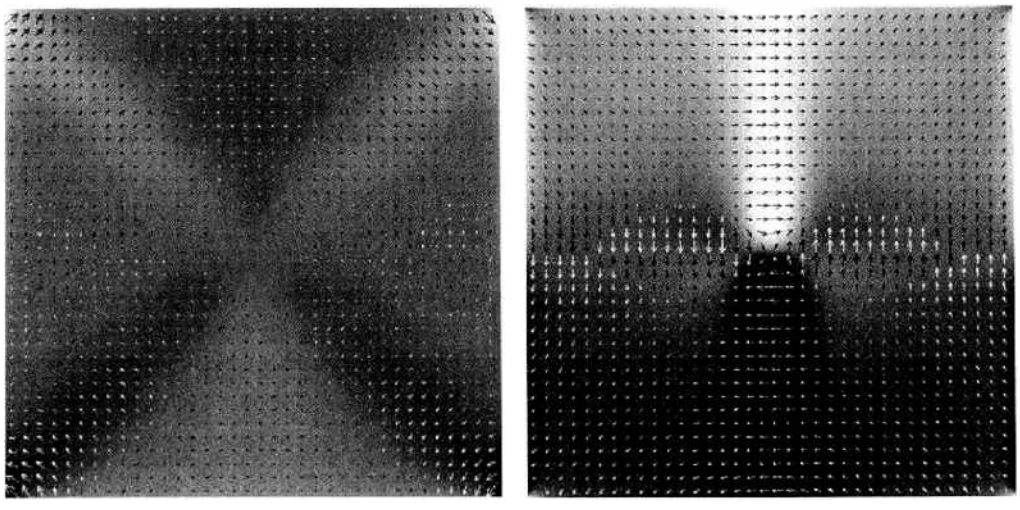
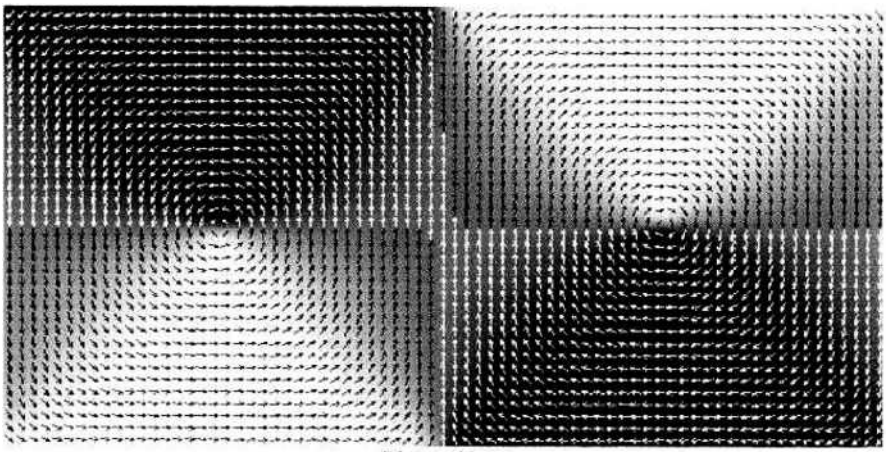


Figure 10. Demagnetizing fields in the Landau configuration (four domains) of the same Py platelet as in Fig. 6: the arrows represent the magnitude (length) and direction (angle) of H_{demag}. Parameter: Bloch line type.

Table 2. Energies for a square Py platelet $200 * 200 * 20$ nm^3, micromagnetic parameters as in Fig. 6, $H_{ext} = 0$.

State	Energies in [perg]			
	E_{ex}	E_{an}	E_{demag}	E_{ms}
State: circular Bloch line type	100.4	63.1	36.8	0.4
C-state	214.0	28.8	184.7	0.4
S-state	229.8	27.3	201.9	0.5
Flower-state	254.6	23.5	230.5	0.7
State: cross Bloch line type	337.8	56.6	280.7	0.4

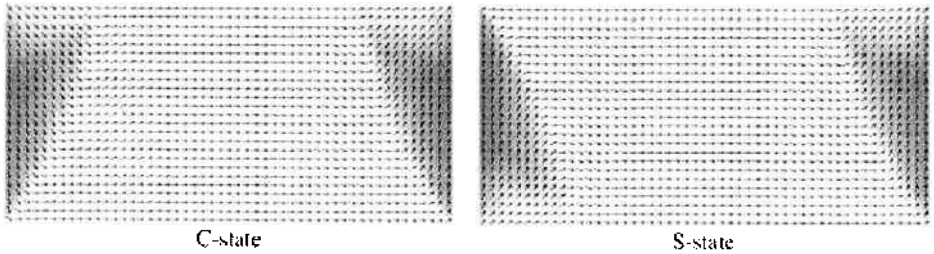


Diamond state

Cross-tie-state

Figure 11. Diamond (seven domains) and Cross-tie-states (four domains) of a rectangular Py platelet in the plane ($z = 7.5$ nm), $2000 * 1000$ nm^2, layer grid $512 * 256 * 4$. Simulation parameters as in Table 3.

Hoffmann and Steinbauer present an analytical approximation for calculating the magnetic energy in circular nanostructures for the vortex phase (closed magnetic flux) and the single-domain phase [89]. Comparing these energies in remanent state leads to a magnetic phase boundary between the single-domain phase and the vortex phase, given by a critical thickness.



C-state S-state

Figure 12. High-remanence states (three domains) of a rectangular Py platelet in the plane ($z = 5$ nm), $2000 * 1000$ nm^2, layer grid $256 * 128 * 2$. Simulation parameters as in Table 3.

Table 3. Remanence $m_r = M_r/M_s$ and energies for a rectangular Py platelet $2000 \times 1000 \times 20$ nm^3, $H_{ext} = 0$.

State	M_r/M_s	Energies in [perg]			
		E_{tot}	E_{ex}	E_{demag}	E_{an}
Diamond-state	−0.0037	588.1	271.7	296.1	20.3
Cross-tie-state	0.0007	655.0	292.2	348.3	14.5
C-state	0.8570	1347.3	87.4	1251.8	8.1
S-state	0.8573	1348.6	87.9	1252.6	8.1
Flower-state	0.9336	1583.5	82.1	1497.5	3.9

Simulation parameters: $M_s = 800$ emu/cm^3, $k_u = 10^3$ erg/cm^3, $A = 1.05$ μerg/cm, the data agree approximately with the μMAG standard problem 1 ($k_u = 3 \cdot 10^3$ erg/cm^3, $A = 1.3$ μerg/cm). Spin patterns in Figs. 11 (low-remanence states) and 12 (high-remanence states).

which depends on the radius of the circular nanostructure. Furthermore, the researchers compared the analytical results with micromagnetic simulations using OOMMF [44] and experimental data and verified their analytical results.

Figure 13 shows for a square Py platelet of edge length $a = 200$ nm the total energy E_{tot} for the Landau-state and C-state as a function of thickness t. The graph shows only a small part of the energy curves, and the different gradients of these states attract our attention. The intersection between the two curves leads to the critical thickness $t = 5.14$ nm: the spin system will take the C state to the left of this value and the Landau state to the opposite side. This critical thickness $t = 5.14$ nm is one point in the phase boundary graph in Fig. 14. Figure 14 can be arranged by repeating the described procedure for several edge lengths a (marked with squares). The following equation fits our data: $-0.1857176873411 8 + 1252.3078786307/a - 47535.731070258/a^2 + 2041814.1004943/a^3$. The dotted curve as comparison corresponds to a result from Hoffmann et al. [Eq. (26) in Ref. [89]], derived analytically for a very flat circular cylinder with uniaxial anisotropy k_u. The radius R has been set to $a/2$.

$$t = \frac{R}{3.37}(\nu - 1)^{1.15}, \quad \nu = \exp\left\{\frac{4A}{2.4J_s M_s R^2}\left[\ln\left(1 + \frac{R}{a_0}\right) + \gamma\right] + \frac{k_u}{2.4J_s M_s}\right\}$$

8.3. Dependencies on the Edge Shape and Grain Structure

The hysteresis curves depend not only on the material parameters and the geometry (width, length, thickness) of the layer but also on the grain structure and the shape and roughness

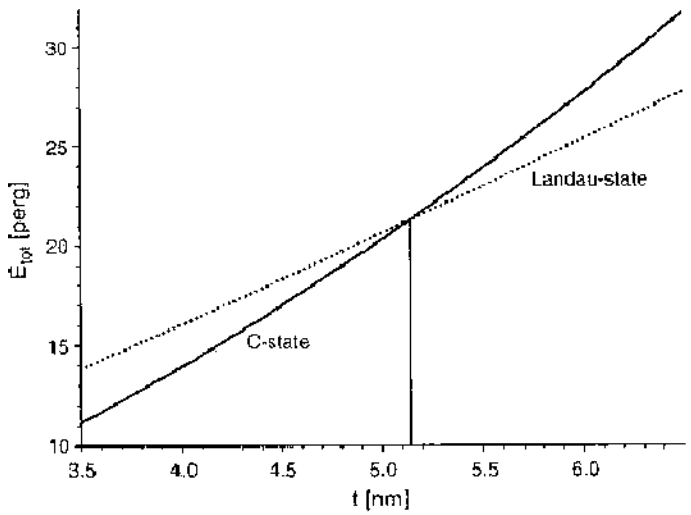


Figure 13. Total energy E_{tot} versus thickness t for the Landau state and C state, square Py platelet of edge length $a = 200$ nm, critical thickness t between vortex state and single domain state: $t = 5.14$ nm.

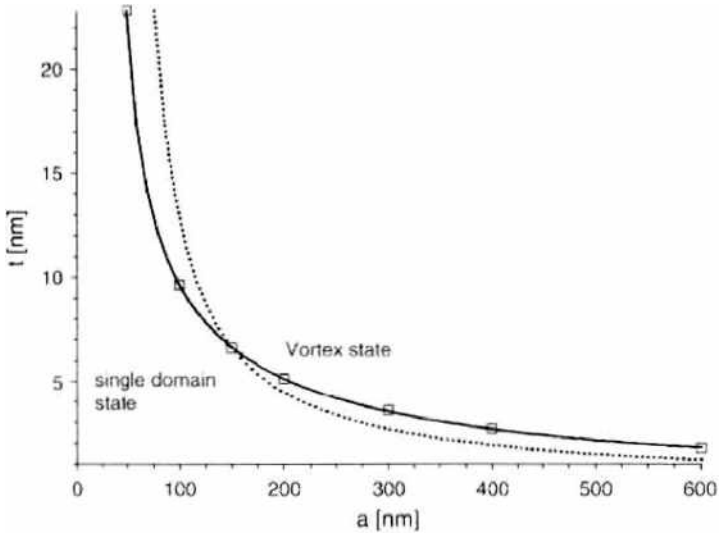


Figure 14. Critical thickness t of a square Py platelet of edge length a (solid), calculated points marked with squares. The figure shows the regions where vortex state and single domain state are dominant. The dotted curve as comparison corresponds to an analytical result derived from Hoffmann et al. [Eq. (26) in Ref. [89]] for a very flat circular cylinder with uniaxial anisotropy k_u. Simulation parameters: $A = 1.05$ μerg/cm, $a_0 = 0.3$ nm, $M_s = 800$ G, $k_u = 10^5$ erg/cm^3, $\gamma = 0.577216$).

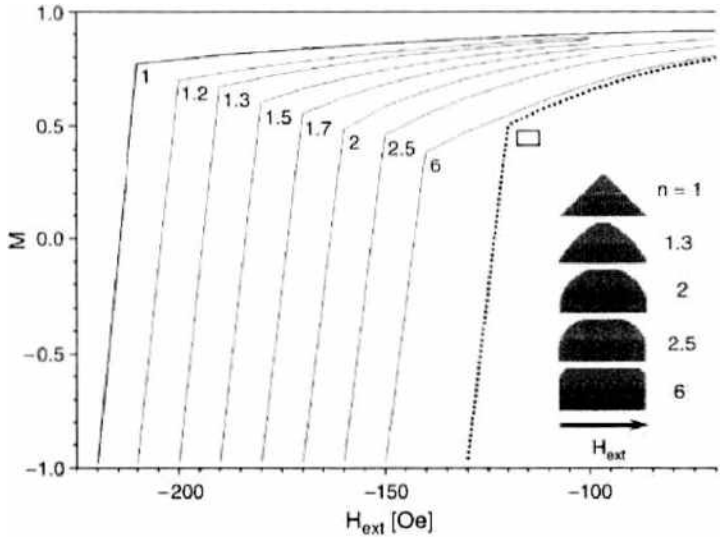


Figure 15. Left branch of a hysteresis curve M(H) for a Py platelet initially magnetized in the C state with different shapes modeled by the shape parameter n [Eq. (36)]; dotted curve corresponds to a rectangular platelet (aspect ratio $w/l = 2$) as comparison. The different shapes start with the triangle ($n = 1$, base line 400 nm, height 200 nm) across a semicircle ($n = 2$, radius 200 nm) up to the rectangle (length 400 nm, width 200 nm). layouts of some elements ($n = 1$, 1.3, 2, 2.5, 6) are shown in the inset. Simulation parameters: $x_0 = 400$ nm, $y_0 = 200$ nm, $t = 3$ nm, $M_s = 900$ erg/cm^3, $A = 1.05$ μerg/cm, $k_u = 100$ erg/cm^3, grain size 35 nm, randomly distributed easy axes.

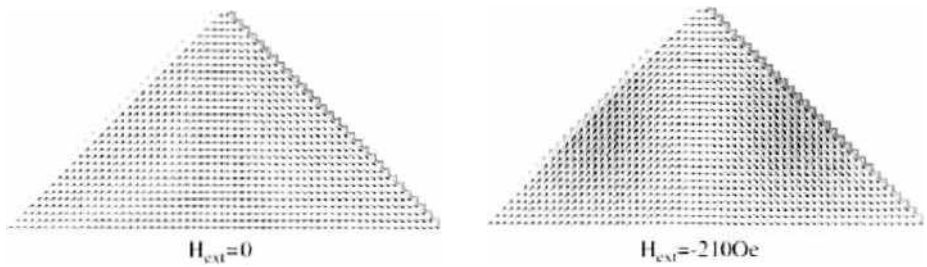


Figure 16. Spin patterns for the triangle shape, zero field $H_{ext} = 0$ and just before magnetization reversal ($H_{ext} = -210$ Oe), initial state $H_{ext} = 250$ Oe. Simulation parameters as in Fig. 15.

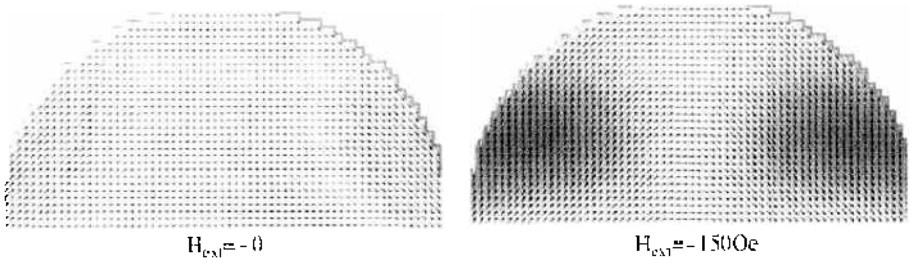


Figure 17. Spin patterns for the semicircle shape, zero field $H_{ext} = 0$ and just before magnetization reversal ($H_{ext} = -150$ Oe), initial state $H_{ext} = 250$ Oe. Simulation parameters as in Fig. 15.

of the edge. The shape was modeled by the following equation for the upper edge ($y \geq 0$) of the layer.

$$\left|\frac{x}{x_0}\right|^n + \left|\frac{y}{y_0}\right|^n = 1, \quad -x_0 \leq x \leq x_0, \quad 0 \leq y \leq y_0 \tag{36}$$

This equation was used by Scheinfein and Arrott to describe the curvature of MRAM elements in the form of a trapezoidally distorted generalized ellipse ($n = 2.7$). They introduced the "path method" to use results of a few dynamical micromagnetic calculations to predict the switching astroids that are very important for MRAMs [90]. Different upper edges of a platelet can be modeled simply by this just one equation, starting with the triangle ($n = 1 \rightarrow$ base line $2x_0$, height y_0) across half of an ellipsis ($n = 2 \rightarrow$ half axes x_0, y_0) up to the rectangle ($n \gg 1 \rightarrow$ length $2x_0$, width y_0).

Figure 15 shows the results obtained for elements with different shapes modeled by the shape parameter n [Eq. (36)]: the C state is the initial magnetization in all cases. An ideal hysteresis curve for a MRAM dot distinguishes itself by large remanence and coercivity values. The triangle shape is the best and the rectangle is the worst from this point of view. The corresponding spin patterns for zero field $H_{ext} = 0$ and just before magnetization reversal are shown in Figs. 16 and 17 for the triangle and semicircle case. The element was initially magnetized with $H_{ext} = 250$ Oe. All elements have a distinct C state, independent of their shape. The magnetization reversal takes place by changing the 90° domain wall at the saturation fields into 180° domain walls in the center of the element, which are squeezed after the coercive field is reached, at the same time 90° closure domain walls occur at the upper edge. This magnetization configuration turns out to be quite stable. There has been brisk research activity in the last years in the search of the right shape [91–93].

9. MAGNETIC NANODOT ARRAYS

Applications of magnetic nanodot arrays include magnetic random access memory, high-density magnetic recording media, and magnetic sensors. Interest in the magnetic properties of small ferromagnetic particles has increased steadily with the improvement of the fabrication techniques. Advances in lithographic patterning (e.g., interference technique of two laser beams) allow us to produce magnetic particles with dimensions of 100 nm and smaller. Co, Ni, CoP, CoNi [94, 95], CoPt, and FePt have been investigated as promising material candidates, the latter are characterized by the high magnetic anisotropy $L1_0$ ordered structure [96, 97]. These materials are of crucial importance in modern technologies such as extremely high-density data storage, spin electronics, memory devices, and high-performance magnetic materials. High magnetocrystalline anisotropy is needed to create a barrier to thermally activated switching of the magnetization.

The aim of the enormous research activity is to find out the conditions for nanomagnets with minimum dimensions, which, however, carry a spatially and temporally stable magnetization. The behavior of nanodot arrays depends on the dot dimensions (height h, base diameter b, aspect ratio $R = h/b$, shape) and also the period of the pattern or the distance between the dots. Micromagnetic models [98] can help to study and understand the fundamental properties.

9.1. Small Nanodots Magnetized Horizontally

Exploiting techniques of self-assembling and the above-reported advances in lithographic patterning allows powerful tools in designing the magnetic behavior of absolutely new devices with the design of patterning as an additional method besides the choice of material and geometry. Deposition onto appropriate substrates allows the self-assembling of isolated magnetic particles, separated by the organic ligand shell, into monolayered superlattices [84, 85]. Hence, the magnetic properties of these superlattices are determined by the magnetization and effective anisotropy of the underlying nanocrystal size distribution and by the resulting dipolar interactions between neighboring nanocrystals.

One application of these modern techniques is as nanodot clusters of small dots magnetized in-plane. Some interessant results for new magnetization curves are represented in Fig. 18. In usual magnetic devices all the spins are linked by the strong exchange coupling, and the characteristic is determined by magnetic domains and walls between them. Nanodot clusters are only dominated by the much weaker coupling of the demagnetizing fields of the involved dots and therefore allow regions of uniform magnetization much smaller as usual domains. Figure 19 shows the spin patterns accompanying with the hysteresis curve in Fig. 18. Dot clusters play an important role in the magnetization reversal process.

9.2. Nanopillars Magnetized Vertically

The simulation starts with a Co cluster magnetized out-of-plane with the dimensions (h_d = 50 nm high, base diameter b_d = 75 nm) in line 1 of Table 4; the dot axis is simultaneously the easy axis. The nanodot cluster can be characterized by the aspect ratio $R = h_d/b_d = 2/3$, and the coverage ratio $\kappa = b_d/d_d = 2/3$, where d_d is the dot distance. The question is now what will happen with increasing κ, line 2 gives an answer. Complex multidomain structures in the whole volume occur instead of the simple flower or vortex state restricted to the top and bottom cross sections of the dots. Dots with alternating sign appear in the cluster additionally with vanishing axial remanence (line 2); these states are completely useless for vertical data storage.

One possibility for improving the situation would be designing higher dots. At first this leads to increasing demagnetizing factors according to Eq. (13) and, as a consequence, to higher axial remanence [Eq. (12)]. Line 3 in Table 4 shows the successful results. The other solution can be enlarging the anisotropy—the Co value is too low for retaining the vertical magnetization (line 2), and it had to be slightly increased from 4 to 5.1 [10^6 erg/cm^3] (line 4). Much larger anisotropy constants up to 7 10^7 erg/cm^3 have already been found for the equiatomic ordered $L1_0$ phase of CoPt and FePt alloys [96].

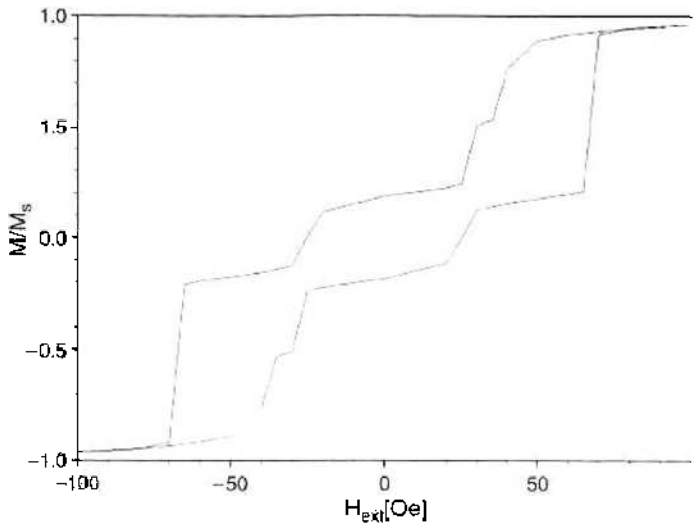


Figure 18. $M(H_{ext})$ for an $5 * 5$-nanodot cluster magnetized in-plane. Simulation parameters: M_s = 800 emu/cm^3, A = 1.05 μerg/cm, k_u = 1000 erg/cm^3, simulation volume $500 * 500 * 5$ nm^3, grid $128 * 128 * 1$.

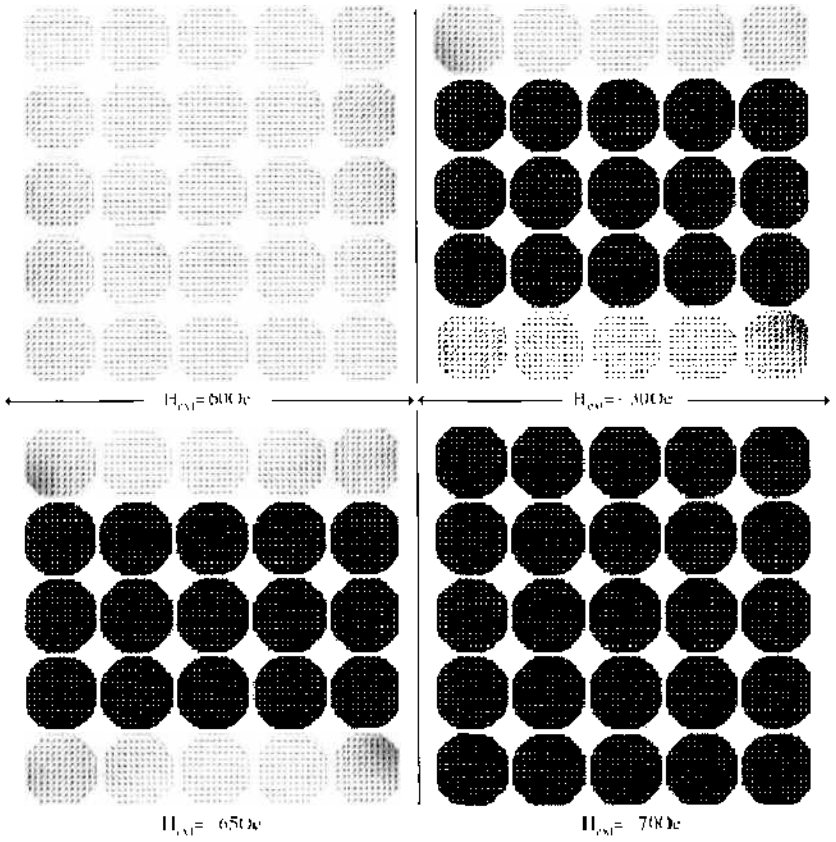


Figure 19. Spin patterns of a 5 * 5-nanodot cluster magnetized in-plane, initially positively saturated, initial state $H_{ext} = 100$ Oe. Simulation parameters as in Fig. 18.

Figure 20 shows the spin patterns for the data of the nanodot cluster in Table 4, line 3. The top ($z = 0$) and bottom ($z = h_d$) plane are presented in Fig. 20—a typical vortex state. Cross sections through the cluster center at a constant x value are drawn in Fig. 21. The magnetization points into the z-direction are in the pillar center, as shown in Fig. 21: presented is the cross section through the center dot ($x = 150$ nm, $i_x = 32$). The spin patterns for the other cross sections ($i_x = 9, 56$ or $i_y = 9, 56$) look quite similar. The spin directions turn into the xy-plane to the pillar top ($z = h_d$) and bottom ($z = 0$) and form a vortex state. The vortex centers in the Fig. 20 would coincide with the circle centers without interaction between the dots, that is the case for large dot distances ($b_d \ll d_d$). This behavior still only shows the dot in the center in Fig. 20. The vortex centers of the surrounding dots are shifted to the dot edge according to the interaction between the dots by the demagnetizing fields. A flower state can also occur, as shown in Fig. 22; the patterns are calculated with the parameters of Table 4, line 4. The anisotropy energy E_{an} is smaller in spite of the larger anisotropy value k_u because the flower state is more favorable in account of the easy axis. The transition from the vortex state to the flower state occurs for higher anisotropy.

Table 4. Remanence M_r and energies for a 3 * 3 nanodot array.

Line	h_d [nm]	Ratio $\kappa = b_d/d_d$	Aspect ratio $R = h_d/b_d$	i_z	k_u [erg/cm^3]	M_r/M_s	energies in [nerg] E_{an}	E_{ex}	E_{dem}	$\tilde{E}_{an}$
1	50	0.6666	0.6666	8	$4 \cdot 10^6$	0.9523	11.498	0.407	10.321	0.770
2	50	0.7827	0.5926	8	$4 \cdot 10^6$	−0.101	13.647	0.804	11.136	1.706
3	130	0.7827	1.5407	32	$4 \cdot 10^6$	0.9509	28.487	0.599	25.365	2.523
4	50	0.7827	0.5926	8	$5.1 \cdot 10^6$	0.9501	16.826	0.481	15.039	1.307

Vertical magnetization becomes unstable with increasing base diameter (line 2), stability can be recovered by either a larger dot height h_d (line 3) or a material with larger anisotropy (see k_u line 4). Simulation parameters: $M_s = 1414$ emu/cm^3, $A = 3.05$ μerg/cm, simulation volume $300 * 300$ nm$^2 * h_d$, grid $64 * 64 * i_z$.

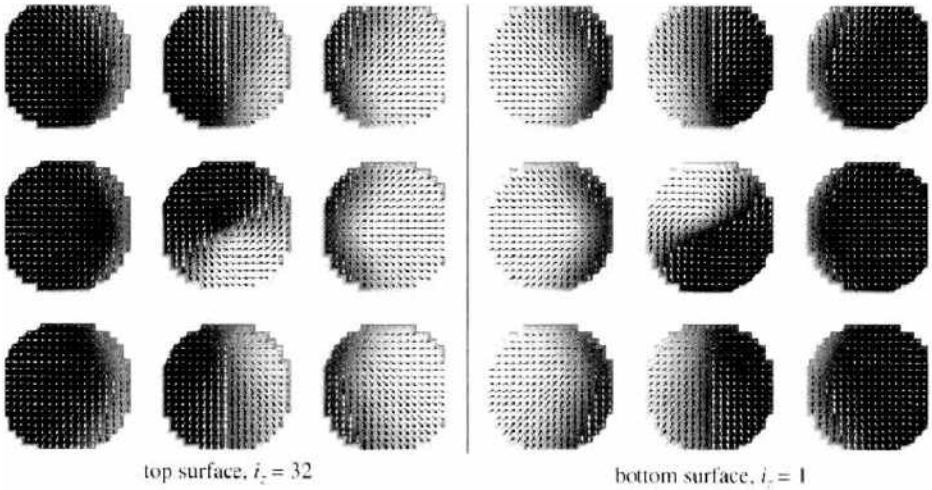


Figure 20. Spin patterns in the basis area of the cylinder for the remanent case, spin distributions for the upper (left) and lower (right) xy-plane of the $3 * 3$-nanodot cluster, axial remanence and energies in Table 4, line 3. Simulation parameters: $M_s = 1414$ emu/cm^3, $A = 3.05$ μerg/cm, $k_u = 4 \times 10^6$ erg/cm^3, simulation volume $300 * 300 * 300$ nm$^3 * h_d$, grid $64 * 64 * 32$, dot height $h_d = 130$ nm, dot diameter $b_d = 84.38$ nm $\rightarrow$ aspect ratio $R = h_d/b_d = 1.5407$.

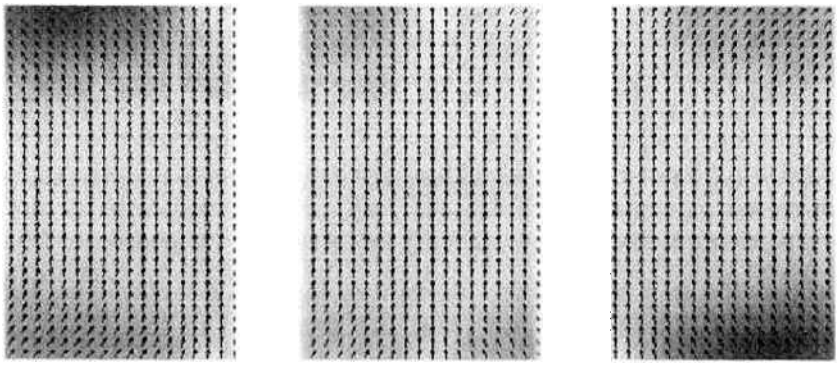


Figure 21. Spin patterns in the yz-plane of the cylinder for the remanent state, cross section through the nanodot cluster center: $i_x = 32$. The other interessant cross sections (i_x, $i_y = 9$, 56) really look quite similar; further representations therefore could be renounced. Simulation parameters as in Fig. 20.

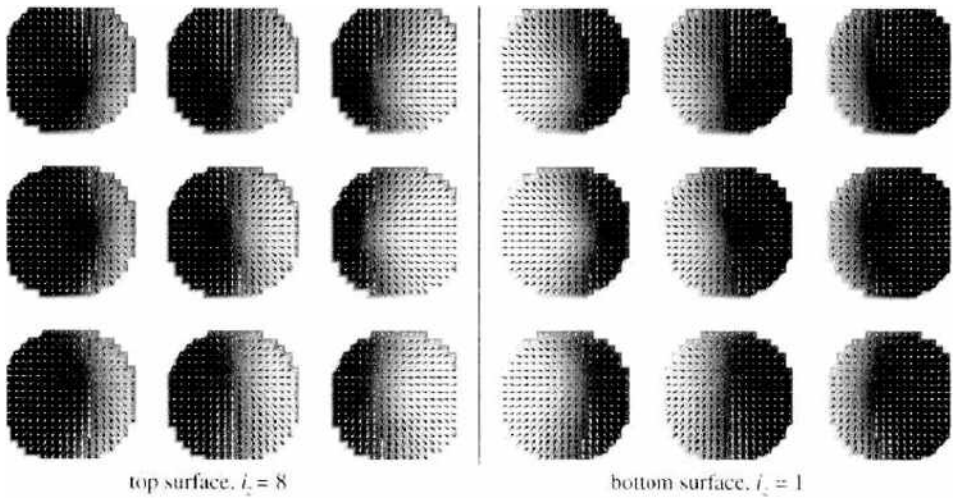


Figure 22. Spin patterns in the basis area of the cylinder for the remanent case, spin distributions for the upper (left) and lower (right) xy-plane of the $3 * 3$-nanodot cluster, axial remanence and energies in Table 4, line 4. The spin patterns in the yz-plane look quite similar to Fig. 21. Simulation parameters: $k_u = 8 \times 10^6$ erg/cm^3, simulation volume $300 * 300 * 50$ nm^3, grid $64 * 64 * 8$. The other parameters as in Fig. 20.

10. GMR DEVICES

A detailed discussion of GMR curves resulting from the anisotropy-term in the limits of a simple model has been presented earlier [99, 100], and the complex influence of the cubic anisotropy constants on the hysteresis loops of single-domain particles has been studied with a Monte-Carlo algorithm recently [101].

For sensor applications, the dependence of the resistance on the external field should be mostly linear. Additional requirements concerning amplitude and saturation field depend strongly on the type of application. GMR systems, in which these characteristics can be modified in a relatively simple way would be therefore highly desirable. It has been shown experimentally [75] that multilayers containing two types of af-coupling are good candidates for this purpose. In this work, we discuss the results of a phenomenological numeric algorithm developed for the simulation of the magnetization and the GMR of such complex multilayered structures. The calculations are compared with experimental results. Within this work we focus on the calculation of magnetization curves and GMR characteristics of copper/permalloy multilayers combining cells of the first and second antiferromagnetic exchange interaction. We show that analytical considerations combined with phenomenological simulations can provide a good description of the measured $\Delta R/R$ versus H dependence and can lead to predictions of new shapes of such GMR curves.

In the simplest case the layers alternate between being weakly (J_{in_i}) or strongly (J_{st}) af coupled, a typical layer sequence is $(stw)_{10}$ (s: strong, w: weak), consisting of 20 layers or more. The strengths of coupling can be adjusted by the layer thickness (t_{Cu}) of the nonmagnetic spacer.

10.1. Simulation Model

In this calculation the external field and the easy axis of the magnetization are in the film plane. Because no hysteresis has been observed in the (Cu/Py) multilayers, no anisotropy term in the energy equation is considered. The model is basically an extension of the Stoner-Wohlfarth's model [71] in the single-domain limit including biquadratic coupling and allowing both different values of the directions of the spins in each layer, as well as of the coupling constants.

10.2. Multilayer Model with Equal Thicknesses of Magnetic Layers

The arrangement of the layers, the magnetization directions, and the related angles φ_i between the external field and the magnetizations are sketched in Fig. 23. For convenience, we chose $\varphi_i = 0$ if the related magnetization points in the direction of H_{ext}. In a simple

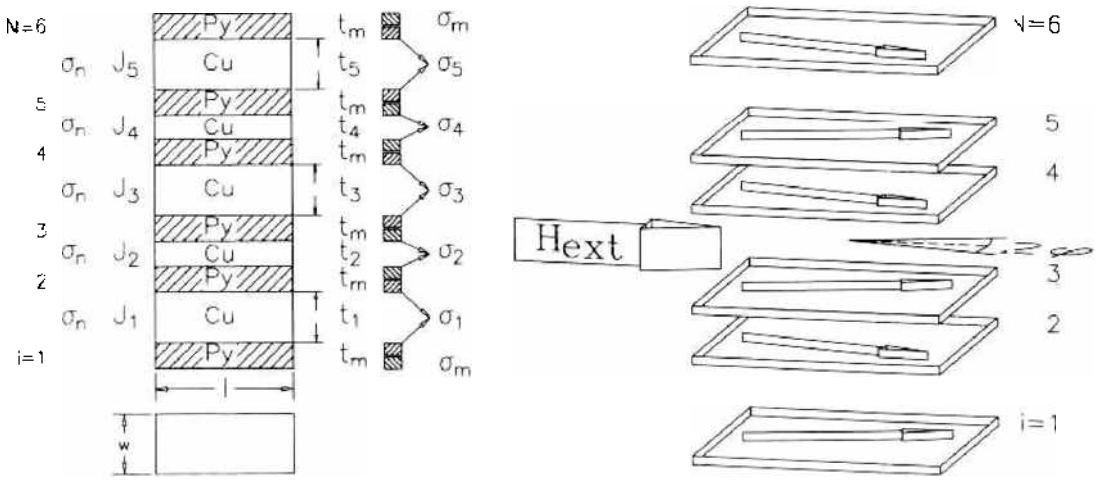


Figure 23. The simplest mixed multilayer $(stw)_{10}$, layer thickness t and conductivity σ (t_n, σ_n) in the nonmagnetic and (t_m, σ_m) in the magnetic layers, three-dimensional view of the magnetic field and spin directions, direction of the external field H_{ext} (broken line) and projection of the spins, $\varphi_i \approx \pm\varphi$, $|\varphi_{i+1} - \varphi_i| \approx 2\varphi$.

approach H_{ext} aligns the spins in all the layers to the angles $\pm\varphi(H_{ext})$; only the sign of the angle changes from one layer to the next [100]. Therefore, the coupling term in the energy equation depends on 2φ. For the simplest arrangement (i.e., N being even and the sequence weak/strong/.../strong/weak), the minimum of the energy

$$E = -NH_{ext}J_s t_m \cos\varphi + \left[\frac{N}{2}J_{ww} + \left(\frac{N}{2} - 1\right)J_{ss}\right]\cos 2\varphi\left(1 + \frac{J_q}{J_l}\cos 2\varphi\right)$$

gives an implicit relation between φ and H_{ext}:

$$J_s t_m H_{ext}(\varphi) = \frac{2}{N}\cos\varphi[NJ_{ww} + (N-2)J_{ss}]\left(1 + 2\frac{J_q}{J_l}\cos 2\varphi\right) \tag{37}$$

Because of the choice of the coordinates, the magnetization is given by

$$M(\varphi)/M_s = \cos\varphi \tag{38}$$

and M and H_{ext} can be simply calculated, if we choose a certain value of φ within the range $0 \le \varphi \le \pi/2$. The saturation field strength H_S is defined by the condition $\varphi = 0$.

$$J_s t_m H_s = H_{ext}(0) = \frac{2}{N}[NJ_{ww} + (N-2)J_{ss}]\left(1 + 2\frac{J_q}{J_l}\right) \tag{39}$$

This simple analytic approach, of course, does not deliver new results. As shown in Section 5.2, however, it can provide reasonable start values for a numeric simulation procedure. Equation (39) can be simplified in the symmetrical case ($J_{ww} = J_{ss} = J_l$) to

$$J_s t_m H_s = 4\left(1 - \frac{1}{N}\right)(J_l + 2J_q) \tag{40}$$

10.3. Pseudo Spin Valve

The general case with different thicknesses shall be considered in this section. First, the derivation shall be restricted to the case of two magnetic layers; it is an extension of a model earlier used by Zhang [102].

$$\frac{E}{S} = -H_{ext}J_{s1}t_1\cos\varphi_1 - H_{ext}J_{s2}t_2\cos\varphi_2 + J_l\cos(\varphi_1 - \varphi_2) + J_q\cos^2(\varphi_1 - \varphi_2) \tag{41}$$

where 1 indicates the layer with the smaller and 2 with the larger thickness ($t_1 < t_2$). A system of two equations results from this energy by derivation with respect to the unknown angles:

$$\frac{\partial E}{\partial\varphi_1} = 0 = H_{ext}J_{s1}t_1\sin\varphi_1 - \sin(\varphi_1 - \varphi_2)[J_l + 2J_q\cos(\varphi_1 - \varphi_2)]$$
$$\frac{\partial E}{\partial\varphi_2} = 0 = H_{ext}J_{s2}t_2\sin\varphi_2 + \sin(\varphi_1 - \varphi_2)[J_l + 2J_q\cos(\varphi_1 - \varphi_2)]$$

Adding the equations gives an relation between the two angles:

$$\varphi_2 = -\arcsin\left(\frac{J_{s1}t_1}{J_{s2}t_2}\sin\varphi_1\right) \tag{42}$$

In a simple approach, H_{ext} aligns the spins in all the layers to the angles $\pm\varphi(H_{ext})$, only the sign of the angle changes from one layer to the next [72, 103]. This is not valid in general but only in the case of equal thicknesses $t_1 = t_2$, as Eq. (42) shows. The minimum of the energy gives an implicit relation between H_{ext} and $\varphi_{1,2}$:

$$H_{ext} = \frac{\sin(\varphi_1 - \varphi_2)}{J_{s1}t_1\sin\varphi_1}[J_l + 2J_q\cos(\varphi_1 - \varphi_2)] \tag{43}$$

Because of the choice of the coordinates, the magnetization is given by

$$M(\varphi_1, \varphi_2) = \frac{M_{s1} t_1 \cos\varphi_1 + M_{s2} t_2 \cos\varphi_2}{t_1 + t_2} \tag{44}$$

Now we can choose φ_1 within the range $0 \le \varphi_1 \le \pi$, and φ_2 can then simply be calculated in Eq. (42): φ_2 starts increasing with φ_1, but it always stays smaller, reaching a maximum value $-\varphi_2 = \arcsin(t_1/t_2) < \pi/2$ and droping down again to zero. The maximum value is, for example, 30° for $t_2 = 2t_1$, and H_{ext} and M can then be calculated according to Eqs. (43) and (44). The saturation field strength H_s is defined by the condition $\varphi_{1,2} = 0$, but in Eq. (43) the relationship between the two angles must also be considered: $-\varphi_2 = \varphi_1 t_1/t_2$ in the limit $\varphi_{s,t} \ll 1$.

$$H_s = \left[\frac{1}{(J_{s1} t_1)} + \frac{1}{(J_{s2} t_2)}\right](J_t + 2J_q) \tag{45}$$

This equation is responsible for the variation of the saturation field in a large value range by simply varying the thickness of one Co-layer. But that is not yet the whole story. In the limit $\varphi_1 = \pi$ and $\varphi_2 = 0$, Eq. (43) gives another limit value H_p. To get it, in Eq. (43) the relationship between the two angles in the limit φ_2, $\pi - \varphi_1 = \epsilon \ll 1$ must also be considered:

$$\sin\varphi_1 = \epsilon; \quad -\varphi_2 = \frac{t_1 \epsilon}{t_2}; \quad \sin(\varphi_1 - \varphi_2) = \epsilon(1 - t_1/t_2)$$

$$H_p = \left[\frac{1}{(J_{s1} t_1)} - \frac{1}{(J_{s2} t_2)}\right](J_t - 2J_q) \tag{46}$$

In the region $0 < H_{ext} < H_p$ the angles and the related spin-dependent resistance remain constant: $\varphi_1 = \pi$, $\varphi_2 = 0$, and $\Delta R/g_m/R(H_c) = 1$. This implies that a plateau in the GMR curves occurs around zero field. For the detailed behavior at zero field, the anisotropy k_u plays a crucial role. For $k_u = 0$, the two magnetizations will rotate at $H_{ext} = 0$ simultaneously, keeping $\varphi_1 - \varphi_2 = \pi$ (see Fig. 24). For large k_u, the directions of M_1 and M_2 will remain constant even if the field is reversed. This behavior already was discussed by van den Berg et al. [104].

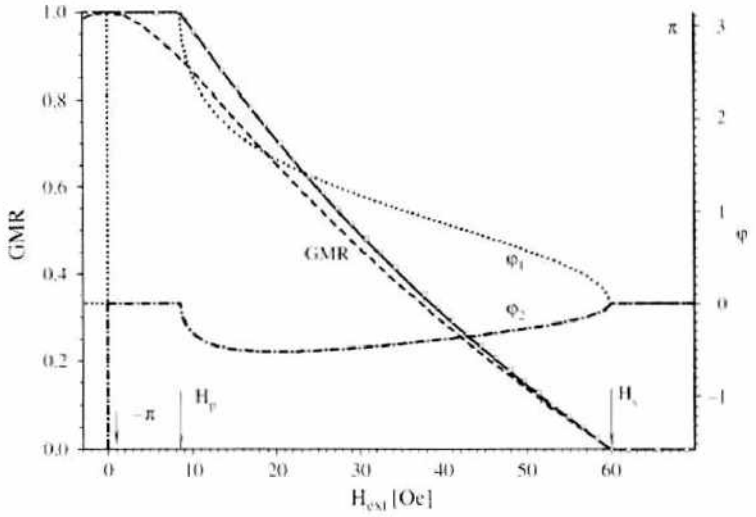


Figure 24. Dependence of the GMR(H_{ext}) on the external field (left axis) for $t_1 = 3$ nm, $t_2 = 6$ nm: $\rightarrow H_p = 8.57$ Oe [Eq. (46)]. Calculation of $H_{ext}(\varphi_1)$ and $\varphi_2(\varphi_1)$ according to Eqs. (43), and (42). GMR(H_{ext}) [solid, Eq. (30)], Eq. (30) and minimization of Eq. (22) (marked with circles), lattice model (marked with squares) [19]. Addition, the two magnetization angles $\varphi_1(H_{ext})$ (dotted) and $\varphi_2(H_{ext})$ [dash-dot, Eq. (42)] are shown, both curves belong to the right axis. $t_1 = t_2 = 4$ nm: $\rightarrow H_p = 0$ [Eq. (46)], the plateau disappears. GMR (H_{ext}) [dashed, Eq. (30)]. Constant parameters for all curves: $M_s = 840$ G; $J_t = 7.210^{-3}$erg/cm^2; J_q/J_t=0.2, $H_s = 60$ Oe [Eq. (45)].

In the region $0 < H_{ext} < H_p$ the angles remain constant, $\varphi_1 = \pi$, $\varphi_2 = 0$, and therefore also the GMR, $\Delta R/g_m/R(H_{ext}) = 1$. In this region, Eq. (44) gives a net moment, which depends on the thickness ratio:

$$M_n = \pm M_s \frac{t_2 - t_1}{t_2 + t_1}; \quad M_{s1} = M_{s2} = M_s \tag{47}$$

The net moment is $M_n = \pm M_s/2$ for $t_2/t_1 = 1.5$. It should be still pointed out that Eq. (43) gives only H_{ext} values in the region $H_p < H_{ext} < H_s$.

Similar considerations lead to the following equation, valid for three magnetic layers:

$$\frac{H_s J_s}{J_l + 2J_q} = \frac{1}{2t_1} + \frac{1}{t_2} + \frac{1}{2t_3} + \sqrt{\frac{1}{4t_1^2} + \frac{1}{t_2^2} + \frac{1}{4t_3^2} - \frac{1}{2t_1 t_3}} \tag{48}$$

Applying Eq. (48) for the case $t_1 = t_2 = t_3 = t$ gives the result $H_s J_s t = 3(J_l + 2J_q)$. The same result can also be obtained by applying Eq. (45) with the symmetry condition ($t_1 = t$, $t_2 = t/2$). There is a very interesting application of Eq. (45). The saturation field H_s can be precisely adjusted in a large region for $t_1 = t_3 > t_2$ only by varying the thickness t_2 of the inner magnetic layer. The other method would be varying the RKKY interlayer coupling constants J_l, J_q, but these material parameters (dependent on the interface spacer/magnetic layer) can only be changed in very rough steps.

10.4. Symmetrical Multilayer

The multilayer $(PyCo)_{t_1}CuCo_{t_2}Cu(PyCo)_{t_1}$ consists of three magnetic layers, but it is a symmetrical system, as the first and third layer are coupled to the second inner one in such a way that the effective fields in these layers are absolutely the same: $\rightarrow \varphi_1 = \varphi_3$. In this case, the energy Eq. (48) can be split into two identical parts with the thicknesses t_1, $t_2/2$. Therefore, this system can be modeled by the simpler system with $n = 2$. Calculations with $n = 3$ have confirmed this assumption.

The normal assumption of alternating signs of the angles φ_i [72, 103] leads in simple models [Eq. (40) in Section 10.2] to the following Equation:

$$h = H_s t J_s/(J_l + 2J_q)/2 = 2(1 - 1/n) \tag{49}$$

Equation (49) just shows the correct values in the cases of one double layer ($h = 1$ for $n = 2$) and the limiting case of a lot of layers ($h = 2$ for $n \rightarrow \infty$); in this case each magnetic layer is af coupled to two neighboring layers, and H_s achieves twice the value of one double layer. The correct saturation fields H_s can be deduced analytically for up to nine magnetic layers, quite similar to Section 10.3 for a multilayer with equal thicknesses $t = t_i$. The necessary calculation is presented here in the case of five magnetic layers. We get for the energy per unit area

$$\frac{E}{S} = -H_s J_s t(\cos\varphi_1 + \cos\varphi_2 + \cos\varphi_3 + \cos\varphi_4 + \cos\varphi_5) + J_l \cos(\varphi_1 - \varphi_2) + J_l \cos(\varphi_2 - \varphi_3) + J_l \cos(\varphi_3 - \varphi_4) + J_l \cos(\varphi_4 - \varphi_5)$$

It is possible to halve the effort with the valid symmetry relations $\varphi_1 = \varphi_5$; $\varphi_2 = \varphi_4$. It is only necessary to calculate three angles instead of five. That can be done by deriving the energy $\partial E/\partial\varphi_i$ for $i = 1 \ldots 3$.

$$0 = 2H_s J_s t \sin\varphi_1 - 2J_l \sin(\varphi_1 - \varphi_2)$$

$$0 = 2H_s J_s t \sin\varphi_2 + 2J_l \sin(\varphi_1 - \varphi_2) - 2J_l \sin(\varphi_2 - \varphi_3)$$

$$0 = H_s J_s t \sin\varphi_3 + 2J_l \sin(\varphi_2 - \varphi_3)$$

In the saturation region the angles are very small; therefore, $\sin\varphi_i$ in the preceding equations can be replaced by the argument $\sin\varphi_i \approx \varphi_i$. This consideration results in a quadratic equation for $H_s J_s t$:

$$0 = (H_s J_s t)^2 - 5H_s J_s t J_f + 5J_f^2 \rightarrow H_s J_s t = \frac{J_f(5+\sqrt{5})}{2}$$

Such simple analytical results exist for up to nine magnetic layers (Table 5). In the general case, $n > 9$, and the further enlargement of H_s can only be calculated numerically with Stoner-Wohlfarth programs as used in Ref. [72]; the results are shown in Fig. 25. The calculated points are marked with squares, and the result of Eq. (40) is dotted as comparison.

Figure 25 gives an answer to the interesting question of how many layers are really necessary to get to the infinite limit. In that limit, each layer is af coupled to two adjacent layers. Therefore, the coupling constants in Eq. (40) must be doubled, as with then increases the saturation value H_s to the doubled value compared to in a double layer. In the case of nine magnetic layers, the saturation value H_s has come up to only 3% below the infinite limit. This means that only nine magnetic layers are necessary to achieve the infinite limit within an error bar of 3%, or only five within 10%.

The two limiting cases ($N = 2$, $N \rightarrow \infty$) are already correctly described by Eq. (40) (dotted line in Fig. 25). That is not valid for the whole curve. We get a prefactor of 1.5 instead of $1.\bar{3}$ [Eq. (40)] for $N = 3$ magnetic layers; in the case of $N = 4$, the prefactor values are $(1 + 1/\sqrt{2} = 1.707)$ instead of 1.5 [Eq. (40)]. As mentioned earlier, the case $N = 3$ can correctly be described with only two layers with the thicknesses $t_1 = t$, $t_2 = t/2$; Eq. (45) produces the correct pre-factor 1.5 (Table 5) for this case.

The maximum GMR value also will be compared as a function of the layer number. The following equation is valid in the limits of the derivation given in Section 6.2: all layers (thickness t_m, conductivity σ_m) are af coupled only in one af maximum (thickness of the nonmagnetic layers $t_n = const$); a buffer layer (thickness t_b) was introduced as well, and g is the GMR effect amplitude:

$$g_{max} = \frac{g_m}{N/(N-1) + t_b/(t_m(N-1)) + \sigma_n t_n/\sigma_m t_m} \tag{50}$$

The GMR maximum g_{max} is also plotted in Fig. 25. The curve increases nearly linearly until the infinite limit ($1/N = 0$, $g_{max} = 1.73$) and shows a completely different behavior compared to the saturation field H_s in Fig. 25. The GMR maximum is obviously achieved later than the saturation field.

10.5. Numeric Solutions in a Multilayer Stack

10.5.1. Infinite Multilayer Stack

In the last section the question was discussed of how many layers are really necessary to get to the infinite limit where each layer is af coupled to two adjacent layers. The previous

Table 5. Saturation field $H_s(N)$ versus the number of magnetic layers in the case of weak anisotropy.

N	$h = H_s(J_s/(J_f + 2J_\varphi)/2$	
2	1	1
3	3/2	1.5
4	$1 + 1/\sqrt{2}$	1.707
5	$(5 + \sqrt{5})/4$	1.809
6	$1 + \sqrt{3}/2$	1.866
7	$\{7 + 2\sqrt{7}\sin[\arctan(\sqrt{3}/9)/3]\}/6$	1.901
8	$1 + \sqrt{(1 + 1/\sqrt{2})/2}$	1.924
9	$1 + [\sqrt{3}\cos(\pi/18) + \sin(\pi/18)]/2$	1.940
∞	2	2

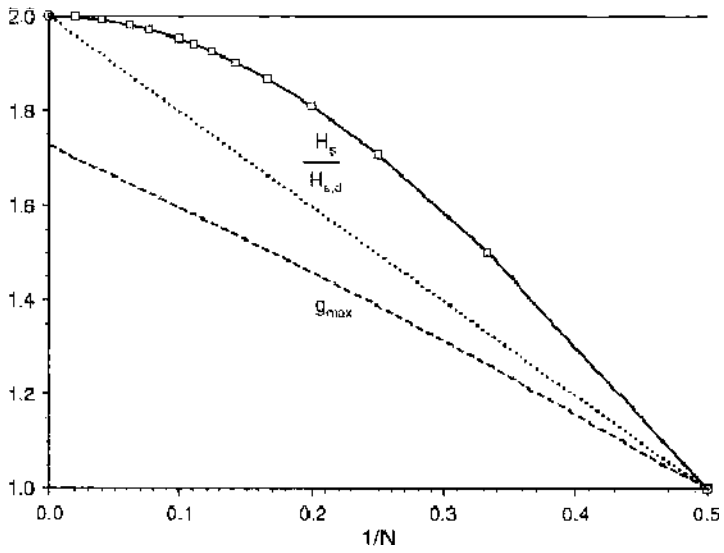


Figure 25. Saturation field $H_s(N)/H_{s,d}$ (solid) and GMR maximum $g_{max}(N)$ according to Eq. (50) (dashed), where N is the number of layers. $H_{s,d} = H_s(0.5)$ corresponds to the double layer case. Calculated points [Table 5, energy minimization of Eq. (25)] are marked with squares; they are fitted by the equation $H_s(N)/H_{s,d} = 1.9987 + 0.0761/N - 5.8173/N^2 + 3.3385/N^3$. Result of Eq. (40) (straight line) is dotted as comparison. $g_{max}(N)$ (dashed) according to Eq. (50). Parameters: $t_F = 2$ nm, $t_m = 1.6$ nm, $t_n = 1$ nm, $t_n = 1$ nm, $\sigma_n/\sigma_m = 5$.

considerations are restricted to the case of neglecting anisotropy. Figure 26 shows for the case of strong anisotropy the development of the GMR curve versus the number N of magnetic layers. The family of curves approaches the curve of the infinite stack (doubled values of J_1) with increasing N; this limit is already nearly obtained with eight layers. There are much larger literature values according to the necessary number of bilayers for the system CoCu (e.g., >20) [Fig. 5 in Ref. [10], simulation] or 50 [Fig. 2 in Ref. [105], experimental]. Figure 26 can be directly compared to Fig. 5 in Ref. [10], as the same model parameters are used. More than 20 bilayers seem to be necessary in this paper, which contradicts Fig. 26. The difference may be caused by the more advanced minimization algorithm (Downhill-Simplex method of Nelder and Mead, [24]) used in our calculation compared to the simpler modified method of deepest descent in Ref. [10]. The GMR maximum is obviously achieved

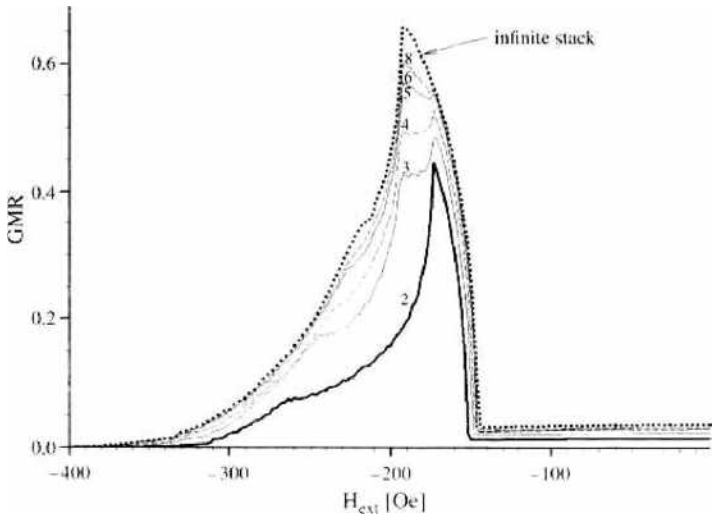


Figure 26. GMR characteristic $\Delta R/R$ (H_{ext}) of a CoCu multilayer for the case of a strong anisotropy. Computer simulation by minimization of the energy equation [Eq. (22)] with the number N (2..8) of magnetic layers as parameter; $N = 2$ corresponds to the bilayer (solid), the dotted curve as comparison to the infinite stack (simulated with doubled J_1 value). Simulation parameters: $M_s = 1300$ emu/cm^3, $J_1 = 6$ merg/cm^2, $J_2 = 0$, $k_u = 2 \cdot 10^5$ erg/cm^3, $t_m = 2$ nm.

later than the saturation field. This statement is completely confirmed by Fig. 26, in which the half width of the GMR curve is already achieved with four magnetic layers.

10.5.2. Thickness Influence in the Case of Three Magnetic Layers

The experimental curves are obtained for the symmetrical system $(PyCo)_{t_1}SpCo_{t_2}Sp(PyCo)_{t_1}$, with Sp is in place of the spacer material. The thickness of the spacer layers has been adjusted to the first af-maximum of the RKKY theory. The anisotropy is reduced by the Py-layer in the bottom and top layers ($6 \rightarrow 1.3$ [10^4 erg/cm^3]). The coupling constants result from fitting the phenomenological numerical results to the experimental curves.

Experimental and theoretical results are obtained for the spacer $Sp=Cu_{85}Ag_{10}Au_5$ in Fig. 27, a small quadratic coupling term ($J_{ll} = 0.17\,J_l$) seems to be obviously typical for this spacer. The plateau is strongly pronounced in agreement with Eq. (46). This is very interesting for the preparation of an AAF (artificial antiferromagnet): the biquadratic coupling has to be minimized for this application. The agreement between the experimental (dotted) and theoretical curves (solid) is very good.

The characteristics in Fig. 28 with Cu as spacer material look quite different. In this case a large quadratic coupling term ($J_{ll} = 0.42J_l$) was found by fitting the numerical to the experimental results. According to the theoretical prediction in Eq. (46), the plateau must vanish in this case ($J_{ll} = J_l/2 \rightarrow H_p = 0$), independent of the thickness ratio.

In general, the agreement between the experimental and theoretical curves is very good: both the plateau width H_p as well as the varying saturation field H_s are reasonably reproduced by the calculations. For finite anisotropy, however, the GMR does not show a real plateau in the region $0 < H_{ext} < H_p$, as is shown in Fig. 27. This was already mentioned [104] and is because of the additional energy terms that modify the condition $\varphi_1 - \varphi_2 = \pi$ and give rise to small deviations of the angles from 0 and π. Numerically, unequal probability for the paths of the system in energy space at the different easy directions is responsible for that effect. As already mentioned, the two magnetizations will keep the condition $\varphi_1 - \varphi_2 \approx 0$ but not $\varphi_1 = \pi$ and $\varphi_2 = 0$ if a strong uniaxial anisotropy is introduced, although this is not directly visible in the GMR.

Analytical [Eq. (46)], numerical, and experimental results (Figs. 27 and 28) have been shown for the magnetization and the GMR of multilayer systems with varying layer thickness, material, and coupling within the layer stack. The general features of the dependence of the GMR on the external field can be reproduced very well by analytical calculations without anisotropy. Both the dependence of the saturation field as well as that of the plateau width

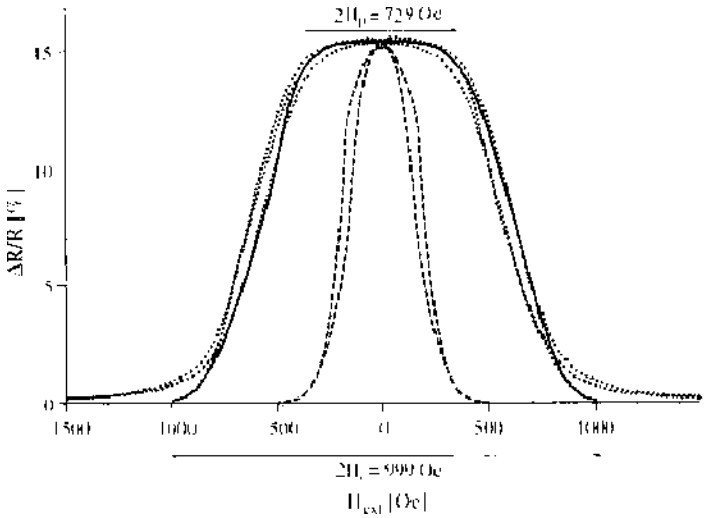


Figure 27. GMR versus H_{ext} for the multilayer $(PyCo)_{15}Cu_{85}Ag_{10}Au_5Co_{15}Cu_{85}Ag_{10}Au_5(PyCo)_{15}$. The experimental curve (dotted line) is compared with theoretical results [Eq. (30) and minimization of Eq. (22), solid line]. Saturation field $H_s = 500$ Oe [Eq. (45)]; plateau $H_p = 364.4$ Oe (Eq. 46). Constant parameters: $J_1 = 18.8 \cdot 10^{-3}$ erg/cm^2, $J_{ll}/J_l = 0.17$, $k_1 = 1.3 \cdot 10^4$ erg/cm^3, $k_2 = 6 \cdot 10^3$ erg/cm^3, $t_2 = 4.5$ Å, $t_1 = 30$ Å. The inner curves (broken line) show theoretical results of Eqs. (30) and (22) for $t_2 = 15$ Å ($H_s = 300$ Oe, $H_p = 88$ Oe).

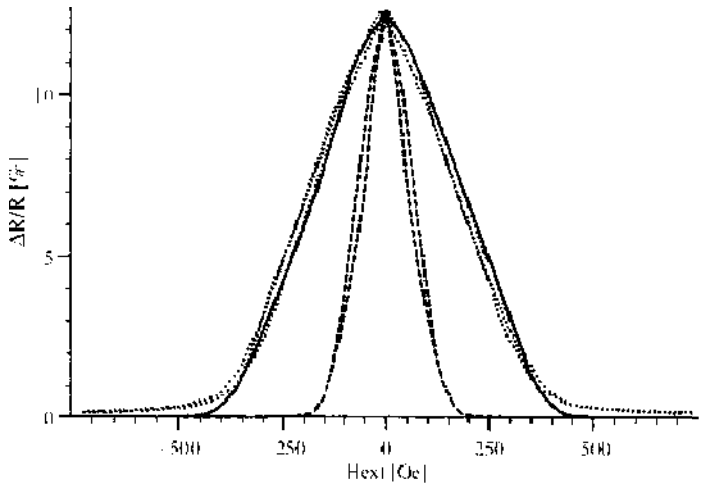


Figure 28. GMR $\Delta R/(g_m R)(H_{ext})$ versus H_{ext} for the multilayer $(PyCo)_{t_1}CuCo_{t_2}Cu(PyCo)_{t_1}$. Experimental results (dotted line) compared with theoretical curves [Eq. (30) and minimization of Eq. (22), solid line]. Simulation parameters: $J_1 = 7 \times 10^{-3}$ erg/cm^2, $J_q/J_1 = 0.42$, $k_1 = 5 \times 10^3$ erg/cm^3, $k_2 = 2 \times 10^5$ erg/cm^3, $t_2 = 5$ Å, $t_1 = 36$ Å $\rightarrow$ $H_s = 394$ Oe [Eq. (45)], $H_p = 30$ Oe [Eq. (46)]. The dotted inner curves show theoretical results of Eqs. (30) and (22) with $t_2 = 15$ Å ($H_s = 148$ Oe, $H_p = 8$ Oe).

on the layer composition can be predicted by analytical results, which is very helpful in designing "good" artificial antiferromagnets. In our calculations, we included the biquadratic coupling, which turned out to play a destructive role in the performance of the AAF. The numerical calculations including small anisotropy terms showed, that the general behavior of the layer system is not strongly affected; that is, both the saturation and the plateau width remain nearly constant.

10.5.3. Magnetization Behavior in a Multilayer Stack

A typical stack in this section consists of up to 50 magnetic layers. An abbreviation to designate a multilayer stack shall be introduced. In the layer $Py_{t_b}/[(Cu_{1.8\,nm}Py_{t_m})_3/(Cu_{0.9\,nm}Py_{t_m})_4]_5$, alternate three weakly coupled with four strongly coupled double layers. The strengths of coupling can be adjusted by the layer thickness (t_{Cu}) of the nonmagnetic spacer. The abbreviating designation $(w_3s_4)_5$ is used for the whole stack; s stands for the first (0.9 nm spacer layer) and w for the second (1.8 nm spacer layer) maximum of the RKKY coupling. The stack always starts with a Py buffer layer ($t_b = 1.9$ nm), and all the other Py layers have the same thickness $t_m = 1.6$ nm.

Figure 29 shows the magnetization directions of two stacks of mixed multilayers for an external field $H_{ext} = 300$ Oe; the field was adjusted to get an angle of $\varphi \approx \pm 45°$. The energy minimization led to the angles represented by arrows in Fig. 29. The angular deviations in comparison to the analytical solution are small in the central region (±0.3°). For the edge spins the angle discrepancies to the simple model are relatively large, particularly for the layer sequence $(ws)_{10}w$ (+42.0°). If one carries out the same calculation for 6 layers $[(ws)_2w]$, the discrepancies turn out to be even larger in the central region (±1.2°). Otherwise the edge differences are considerably smaller (±4.2°) if the layer sequence starts with the strong coupling $[(sw)_{10}s]$.

Figure 30 shows for the example of the sequence $(w_3s_3)_3w_3$ how the spin distribution develops in 22 layers, depending on the field strength H_{ext}. One clearly recognizes the separate behavior in the parcels, consisting of three layers with homogeneous coupling. Already,

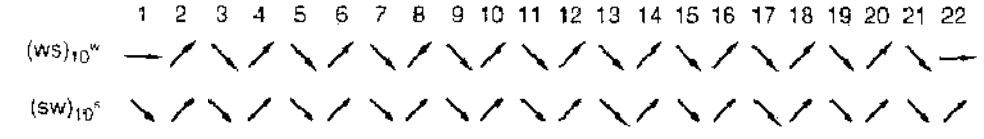


Figure 29. Spin distribution for the sequences $(ws)_{10}w$ and $(sw)_{10}s$ for an external field $H_{ext} = 300$ Oe $\rightarrow \varphi \approx \pm 45°$. Simulation parameters: $M_s = 836$ emu/cm^3, $t_m = 1.6$ nm, $J_s = 27$ merg/cm^2, $J_s/J_{ww} = 21$, $J_q/J_1 = 0.33$.

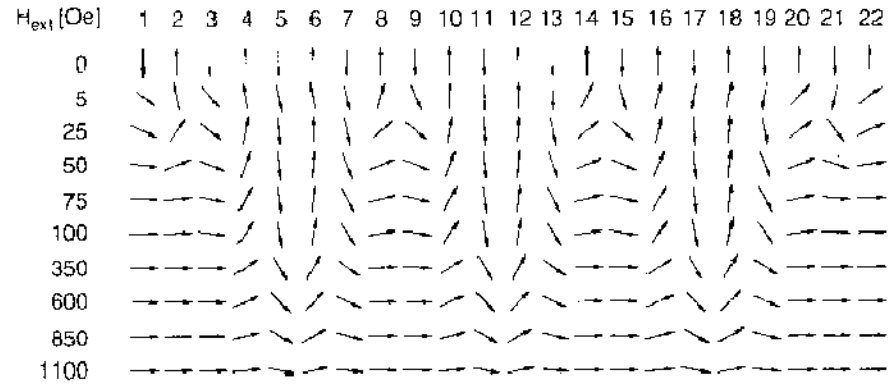


Figure 30. Development of the angle distribution of the spins in the multilayer for the sequence $(w_3s_3)_3w_2$ (even number of layers) as a function of the external field H_{ext}. Simulation parameters as in Fig. 29.

at a fractional part of the saturation field $\approx H_s/16$, the adjustment of the spins with weak coupling is completed. After this the turn of the spins with strong coupling occurs very much slower. In the layers with strong coupling, both spins in the center of the region form a pair with the known behavior that the spin angles only change the sign from one layer to the other. The spins 1 and 4 form a further pair with such a behavior but with different angles. This change in the arrangement of the spins leads to large deviations from the analytical calculation. It becomes clear in Fig. 30 that a simulation program must hold open the degrees of freedom for all magnetizations. The result of the calculation of the inverse sequence $(s_3w_3)_3s_3$ shows a similar behavior.

For zero field, the spins are aligned to $\varphi_i = \pm\pi/2$, and the magnetization is exactly zero $[M(0) = 0]$. In the case of odd number of magnetic layers $[N = 21, (w_3s_3)_3w_2]$, another configuration was obtained $(\varphi_i = 0, \pi)$, as is shown in Fig. 31. In the case $H_{ext} = 0$, the energy depends only on the differences of the angles according to Eq. (22), therefore, the same energy value belongs to this possible configuration. The magnetization, however, according to Eq. (23) is quite different $[M(0) = 1/21 \neq 0]$.

10.5.4. GMR Characteristic in a Multilayer Stack

Figure 32 shows two curves for the different stacks—$(ws)_6$ and $(w_2s_2)_3$—the experimental curves [75] are marked with squares, and the simulated curves have been adjusted to the experiments with the extended Stoner-Wohlfarth model according to Eq. (22). In the simplest case, the layers alternate, being are weakly (J_{ww}) and strongly (J_{ss}) af-coupled, and the layer sequence of the first curve is $(ws)_6$, consisting of 12 magnetic layers. In the second stack two strongly coupled layers follow two weakly coupled layers, and the sequence is $(w_2s_2)_3$. The GMR characteristics look very different in these cases, as Fig. 32 shows. The parameters necessary to fit the experimental and theoretical curves are almost the same. Only the ratios J_q/J_l are slightly different ($0.27 \leftrightarrow 0.33$) to achieve the best fit between experimental and simulated curves.

The comparison with the simple multilayer model is now interesting; Eq. (39) yields with the parameters of Fig. 32 a saturation field $H_s = 548$ Oe for both sequences, in fact, because the number of weak and strong coupled layers is the same. This value agrees approximately with the result of the Stoner-Wohlfarth model in Fig. 32 for the sequence $(ws)_6$ (650 Oe)

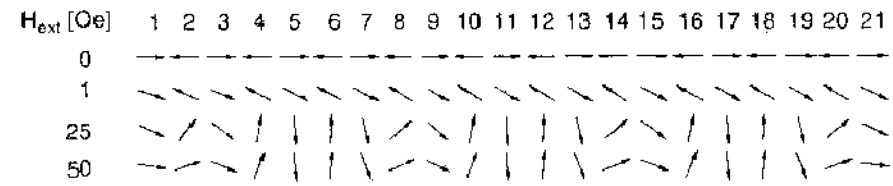


Figure 31. Development of the angle distribution of the spins in the multilayer for the sequence $(w_3s_3)_3w_2$ (odd number of layers) as a function of the external field H_{ext}. Simulation parameters as in Fig. 29.

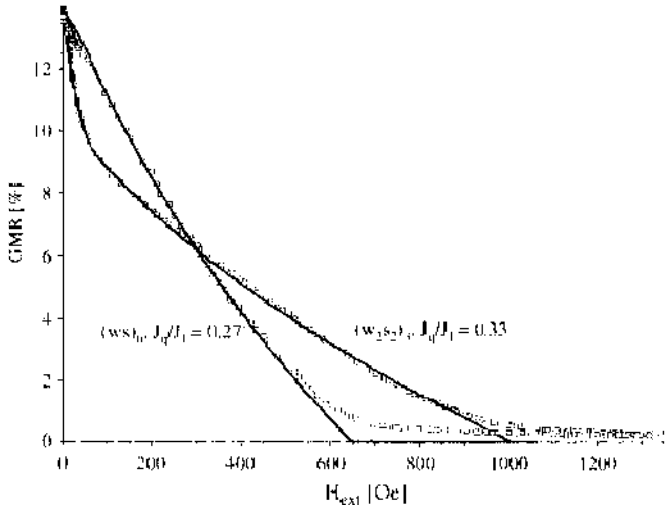


Figure 32. GMR characteristic $\Delta R/R$ (H_{ext}) at mixed multilayer, containing simultaneously the first and the second maximum of af coupling, computer simulation by minimization of the energy equation [Eq. (22)], comparison between experimental (marked with squares) and theoretical curves (solid). Simulation parameters: $M_s = 836$ G, $J_q = 27$ merg/cm^2, $J_q/J_{qc} = 21$, $t_b = 1.9$ nm, $t_m = 1.6$ nm.

but is absolutely wrong for the sequence $(w_2s_2)_3$ (1000 Oe). The relatively small deviation for the sequence $(ws)_6$ can be understood only partially as the difference between the solid and dotted curve in Fig. 25 (563 Oe). The simple model Eq. 39 cannot display the behavior of the sequence $(w_2s_2)_3$ at all. The profound reason for the discussed deviations lies in the fact that the analytical calculation does not allow angular deviations within the stack at all. Therefore, it is very essential to fit the experimental data with a complete Stoner-Wohlfarth model to extract right numbers for the coupling constants J_l, J_q.

Figure 33 shows four curves for the completely different stacks: w_{40}, $(w_4s_4)_5$, $(ws)_{20}$, and s_{40}. The experimental curves [75] are marked with squares, and the simulated curves have been adjusted to the experiments with the extended Stoner-Wohlfarth model according to Eq. (22). In the interesting case of the sequence with alternating coupling $(ws)_{20}$, the overall shape of the experimental curve and the approximate saturation field are reproduced very well by the numeric calculations in Fig. 33. The saturation field achieves approximately half the value of the stack s_{40} with only strongly coupled layers. This result is already confirmed by the simple multilayer model in Eq. (39).

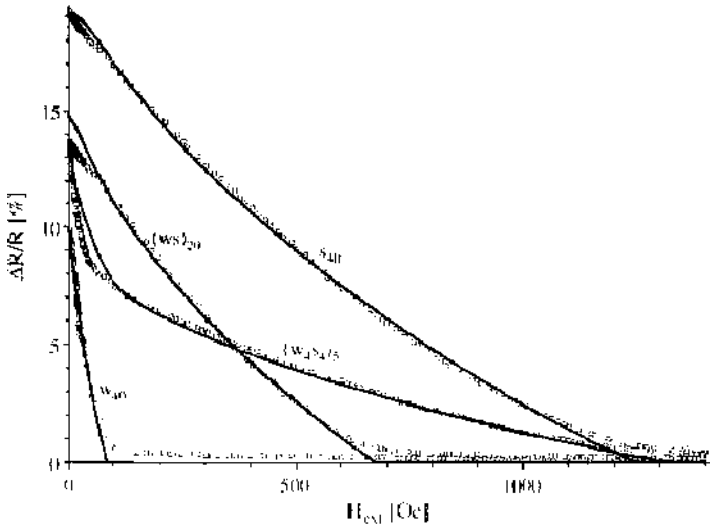


Figure 33. GMR-measurements (marked with squares) and simulations (solid) $\Delta R/R(H_{ext})$ at mixed multilayer structures, each sequence built on the buffer layer $Py_{(80nm)}$, $t_b = 1.9$ nm, $t_m = 1.6$ nm. Simulation (solid) with the parameters: $M_s = 836$ emu/cm^3, $t_m = 1.6$ nm, $J_q = 25$ merg/cm^2, $J_q/J_{qc} = 15$, $J_q/J_l = 0.35$ [s_{40}, w_{40}, $(ws)_{20}$]. 0.5 [$(w_4s_4)_5$].

The family of curves in Fig. 33 could be realized by almost the same parameter set. Only the ratio J_{ij}/J_i must be different ($0.35 \leftrightarrow 0.5$) for the sequence $(w_4 s_4)_5$ to achieve the best fit between experimental and simulated curves.

The agreement between experimental and theoretical curves is very good; as the comparison between the solid curves and measurements (marked with squares) in Figs. 32 and 33 shows, deviations remain in the saturation region. The gradual approximation to the saturation value is obviously not included in the extended Stoner-Wohlfarth model [Eq. (22)]; the grain (lost spins) and domain structure within the layer are responsible for this effect and can only be studied by a micromagnetic simulation.

The behavior of the mixed multilayers with the sequences $(w_2 s_2)_3$ (Fig. 32) and $(w_4 s_4)_5$ (Fig. 33) deviates completely from the other ones [s_{40}, w_{40}, $(ws)_{20}$, $(ws)_6$]: At first the curve decreases fast and then turns extremely slowly into the saturation field. The characteristic behavior with two different gradients is also confirmed experimentally.

10.6. Micromagnetic Results for a GMR Device

Micromagnetic calculations allow us to have a look into the complex spin distributions of the two layers forming a GMR device. The results are summarized in Figs. 34 and 35. The simulation starts with the initial state $H_{ext} = 200$ Oe; an S-state spin pattern as in Fig. 34 with three domains is typical for this situation (upper drawings). The spin pattern in the layer is dominated by a large domain, in which all spins are aligned to the external field H_{ext}; in addition, there are two thin-edge domains at ($x = \pm$ a; see Fig. 1). Differences in the edge domains are remarkably between the two layers. The spins show angles only different in sign: $\pm\varphi(H_{ext})$. We got to know this well-known behavior from af-coupled layers

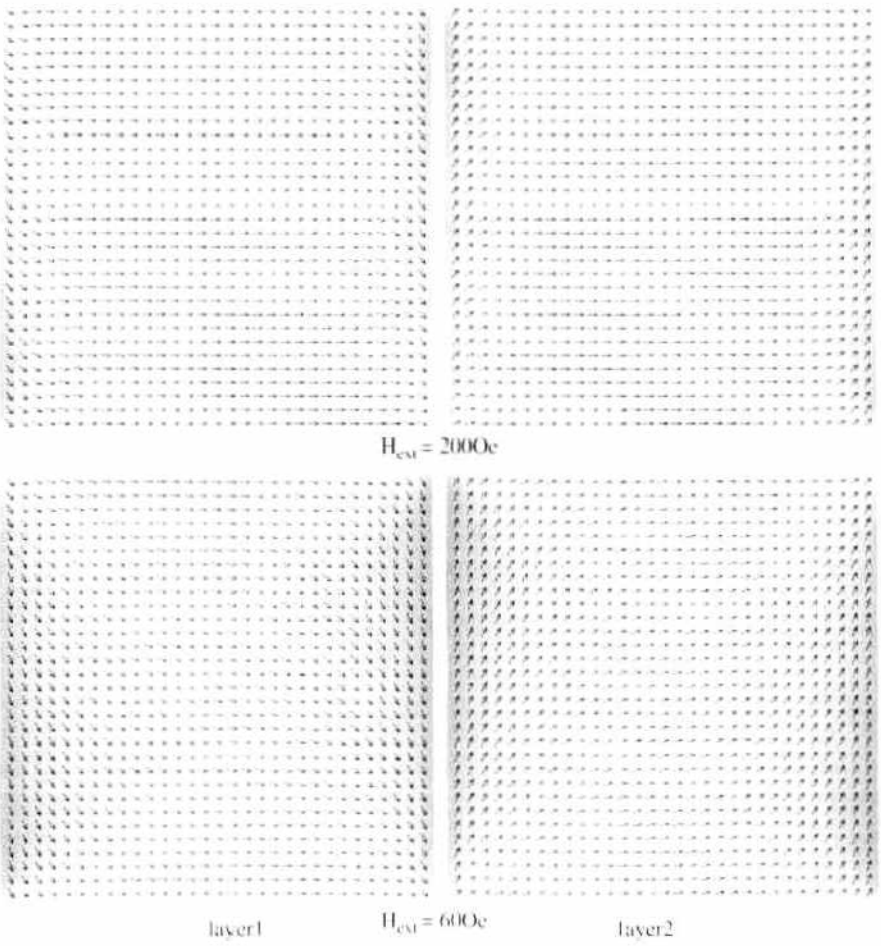


Figure 34. S-state spin configuration (three domains) in a GMR device. $H_{ext} = 200$ Oe (at the top), $H_{ext} = 60$ Oe (below). Simulation parameters: $0.9 * 0.9$ μm^2, $M_s = 900$ emu/cm^3, $A = 1$ μerg/cm, $k_u = 10^5$ erg/cm^3, $t_m = 3$ nm, $t_s = 2$ nm.

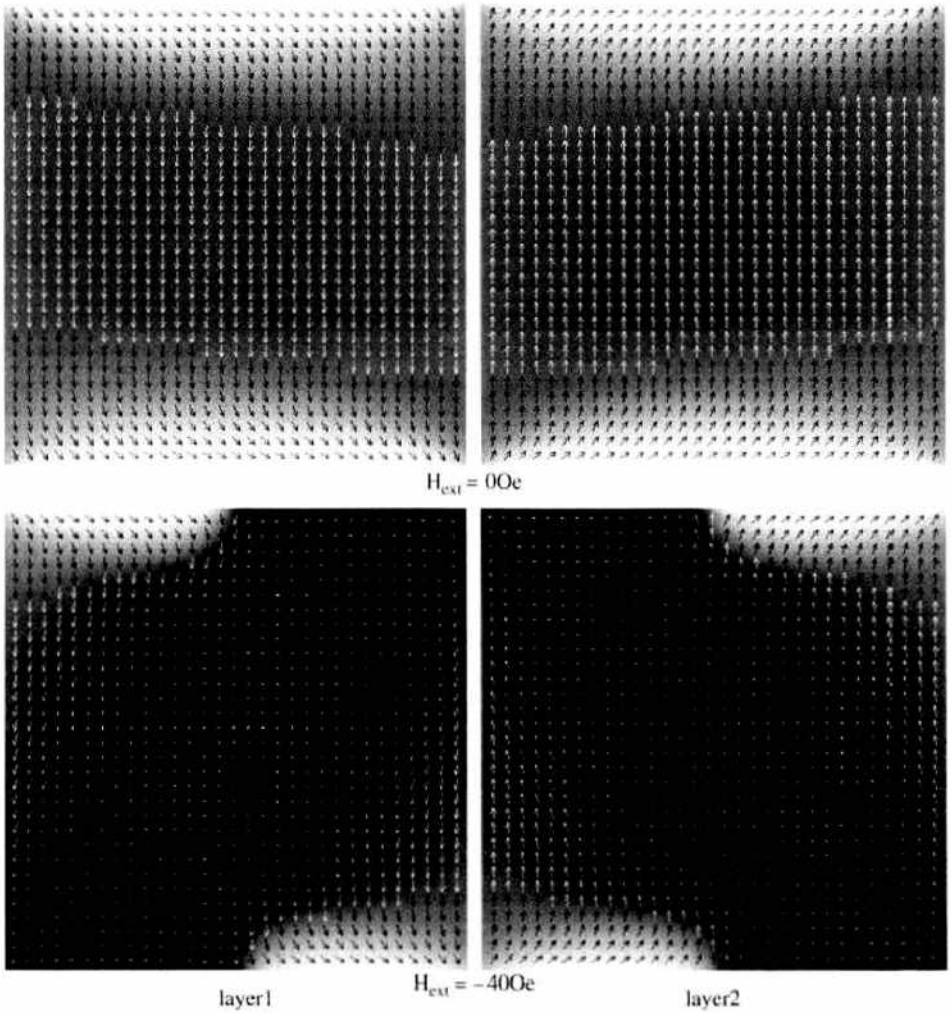


Figure 35. S-state spin configuration (three domains) in a GMR device, $H_{ext} = 0$ Oe (at the top), $H_{ext} = -40$ Oe (below). Simulation parameters as in Fig. 34.

in Section 10.2 or the sketch in Fig. 23. The reason for this behavior is here, of course, the af coupling between the two layers according to their mutual demagnetizing fields.

These edge domains expand with decreasing external fields and show a four-domain structure (lower drawing of Fig. 34) at $H_{ext} = 60$ Oe. The large inner domain already disappeared, and two domains with field-aligned spins have survived at the $(y = \pm b)$ edges. The upper drawing in Fig. 35 shows that the spins are aligned perpendicular $(\pm\pi/2)$ to the initial H_{ext} direction at zero field $H_{ext}=0$, that is the main domain in this state. This behavior already has been predicted in Section 10.2: $\varphi = \pm\pi/2 \rightarrow H_{ext}=0$ in Eq. (37). Two edge domains exist additionally at $(y = \pm b)$, again with spins only different in sign because of the af-coupled layers. Another spin distribution with five domains with spin angles different in sign shows the drawing $H_{ext} = -40$ Oe.

11. STRAY FIELDS IN TMR DEVICES

Magnetic tunnel junctions (MTJs) are key devices for the development in magnetoelectronics [106] both for application as sensors or nonvolatile memory (MRAM) devices [107]. Although large values of the tunneling magnetoresistance (TMR) were demonstrated [107–109], the influence of the geometry of the tunneling elements on the TMR is not yet understood. This is because of the complex interplay between magnetic anisotropies, domain splitting in the layers, new coupling effects across the tunneling barrier [110], and edge [111] or surface effects. One of the most interesting points is the magnetic and related TMR behavior of MTJs with tunneling areas well below 1 μ. Micromagnetic simulations with parameters adapted to the measurements can help in answering the open questions.

CoFe (1.5)/Ru (0.9)/CoFe (2.2)/Al_2O_3 (1.5)/NiFe (6) form the magnetic effective part of a multilayer [42, 104, 112, 113]. The films similar to those discussed earlier [114] are deposited by dc magnetron sputtering on a Ta/Cu buffer, and the stack covered by a Ta(5 nm)/Au(30 nm) top layer to prevent oxidation. The film stack (thickness dimensions in nanometers) was patterned by e-beam lithography and Ar ion etching to rectangular junctions with different sizes and aspect ratios. Using secondary ion mass spectroscopy, the etching was stopped shortly after the Al signal vanished. Thus the patterned part of the films probably exists from the top down until the second layer of the AAF [CoFe (2.2 nm)], whereas the remaining films [Ru (0.9 nm)/CoFe (1.5 nm)] were still continuous. The etch stop turned out to be one of the crucial points of the preparation routine. The junctions were first magnetically saturated to have a well-defined magnetization of the hard layer stack. Then they were transferred to a customized conducting atomic force microscope (C-AFM), which allows topographic imaging and contacting the top electrode with a conducting diamond tip.

An overview of the tunnel magnetic resistance effect is given by Moodera and Mathon [106]. Recently the domain structures in an MTJ element were calculated in a micromagnetic simulation using a hybrid finite element/boundary element method [115]. The equilibrium spin distribution was obtained from the solution of the Gilbert equation, and the anisotropy term in the effective field was neglected. The free layer ($500 * 500$ nm^2, $t = 10$ nm) was placed in the middle of the pinned layer ($1 * 1$ μm^2, $t = 3$ nm); therefore, no stray fields of the aaf can penetrate into the free layer. The numerical calculations show that the energy of different remanent states (C and S state) is rather similar. For a ferromagnet in contact with an antiferromagnet (AFM), a shift of the hysteresis loop along the magnetic field axis can occur, called and exchange bias. Monte Carlo simulations [116] with a heat-bath algorithm and single-spin flip methods for the simulation were used. The calculations show a strong dependence of the exchange bias on the degree of dilution and the thickness of the AFM layer in agreement with experimental observations.

The stray field [H_{dem} in Eq. (14)] component H_x at the layer center ($x = y = z = 0$) can be considerably simplified in the case of thin layers (z_i, $t_i \ll w_i$, l_i):

$$H_x(0,0,0)/M_s \approx \frac{8 t_i w_i}{l_i \sqrt{w_i^2 + l_i^2}} \rightarrow \frac{4\sqrt{2}\, t_i}{l_i} \quad \text{for } w_i = l_i \tag{51}$$

Equation (51) shows that the field increases approximately inversely proportional to the lateral dimensions (e.g., for a square layer the field increases from 2.26% to 21.53% if the edge length decreases from 500 to 50 nm while layer thickness $t = 2$ nm and the distance $z = 5$ nm from the layer center were remaining constant). Equation (51) shows also that the field is approximately independent on the distance z; the field increases, for example, only in the worst case from 21.53% to 22.33% [according to Eq. (14)] for half the distance z ($5 \rightarrow 2.5$ nm). The deviations of Eq. (51) (line 1) are less than 1.3% (column $50 * 50$) and remain much smaller for the larger devices. There are insignificant deviations to Newells formula Eq. 15 (line 4, grid $64 * 64$, respectively, $64 * 32$).

It is remarkable that devices with a larger dimension in the external field direction (aspect ratio $l/w - 2$, $100 * 50$, 7.05%) have smaller stray field values as the quadratic devices ($l/w = 1$, $100 * 100$, 11.17%) and could enable higher storage densities from this point of view.

The field of Eq. (51) is only strictly valid above or below the center of the layer; it increases strongly toward the edge faces, as was shown in Figs. 3–5. Therefore, the mean field value $\bar{H}_x$ was introduced instead of $H_x(x, 0, 0)$; the deviations can be quite big, particularly for the larger devices.

$$\bar{H}_x(z_m) = \frac{1}{twl} \int_{-a}^{a} \int_{-b}^{b} \int_{-c}^{c} H_x(x, y, z_m - z)\, dx\, dy\, dz \tag{52}$$

$\bar{H}_x$ according to Eq. (52) decreases with increasing distance d_{ij} between the layers, especially for the smaller elements.

The phenomenological model of energy minimization includes three layers [aaf (CoFe/Ru/CoFe), sensor Py] and af coupling (linear and quadratic) between the af-coupled layers 1–2 (see Fig. 36 for the geometry). The center line of the hysteresis has been shifted by

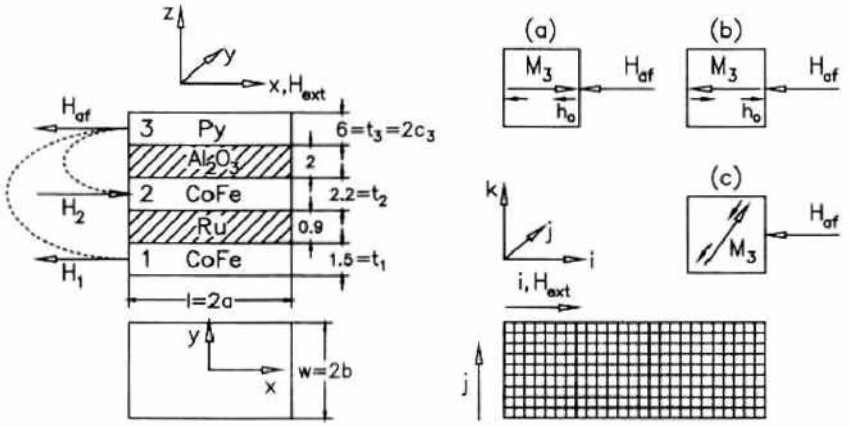


Figure 36. Sketch of the MTJ element, coordinate system, AAF layer 1 + 2, sensor layer 3, grid indices i, j, k and direction of external field H_{ext}. Direction of the stray field lines (dashed) of the AAF for the minor loop, corresponding foreign stray field H_{af} unchanged by the minor loop, own stray field h_o of layer three. Initial state M_3 for the spins in the sensor layer (a), spin direction after magnetization reversal of the sensor layer 3 (b), transition case (c).

10 Oe corresponding to a Néel-type f coupling between layers 2 and 3 in agreement with measurements at the large devices ($600 * 300$ nm^2). The other model parameters (anisotropy constants) were adjusted to the observed major and minor loops of the $600 * 300$ nm^2 device.

The hysteresis curves (minor loops) have been calculated with Eqs. (13), (14), and (22) by the method of energy minimization, including all stray field energies between the three layers; infinity means neglecting dipole coupling. The results in Fig. 37 show that the shift in the minor loop is most pronounced in the case of the smallest device ($50 * 50$ nm^2). The shift decreases continuously toward the larger devices, which can be well understood by the stray field prefactor in Eq. (51) (Table 6). The stray fields in devices with dimensions larger than 1 μ are too small for shift values of 50 Oe, which have been reported earlier as results from micromagnetic calculations [74]. The results on minor loops have shown [43] that a significant improvement of the MRAM cell is attainable if the thickness of the sensor layer 3 is reduced. The appearance of domains will further decrease the stray field influence, particularly for the larger devices.

Figures 38 and 39, show some results for a very small TMR device. Figure 38 contains the micromagnetic parameters; the distance between AAF and soft layer consists of half the extension of the AAF ($1.5 + 0.9 + 2.2$) and the thickness of the insulator layer (2).

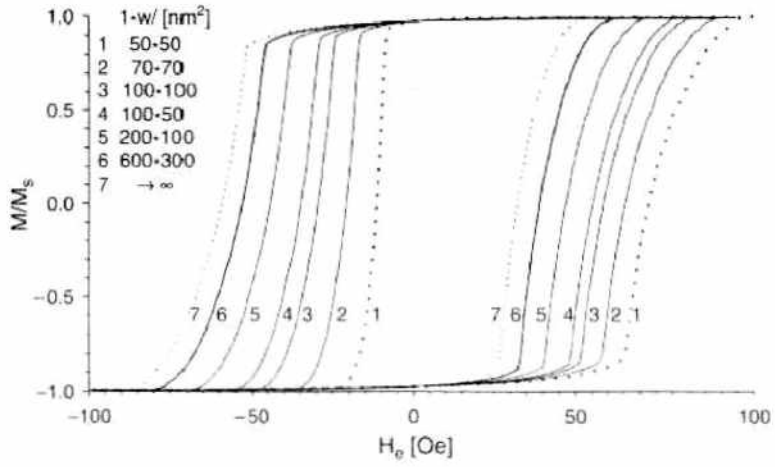


Figure 37. Hysteresis curves $M(H_{ext})$ for different values of length l and width w, calculated by energy minimization resulting from Eq. (22). Curve 7 represents the case without stray fields. Reprinted with permission from [43], W. Schepper et al. in "Analysis of the disturbing influence of stray fields in very small MRAM cells by computer simulation" (D. Shi, B. Aktas, L. Pust, and F. Mikailov, Eds.), Vol. 2593, p. 75, Nanostructured Magnetic Materials and Their Applications, Springer Lecture Notes in Physics, Springer Berlin, 2002, © 2002, Springer.

Table 6. Stray fields $100 * H_z(0, 0, z)/M_s$ of a cube $l * w * t$ in a distance z according to Eq. (14) (lines 2–3), Eq. (51) (line 1), and Newells formula Eq. (15) (line 4, grid $64 * 64$, respectively, $64 * 32$).

	$l * w$ [nm^2]	$50 * 50$	$70 * 70$	$100 * 100$	$100 * 50$	$200 * 200$	$200 * 100$	$600 * 300$
$z = 0$ nm	Eqs. (51)	22.63	16.16	11.31	7.16	5.66	3.58	1.19
2.5	(13), (14)	22.33	16.06	11.28	7.13	5.65	3.57	1.19
5	(13), (14)	21.53	15.75	11.17	7.06	5.64	3.57	1.19
5	(15)	21.52	15.75	11.17	7.05	5.64	3.57	1.19

Constant parameter: $t = 2$ nm.

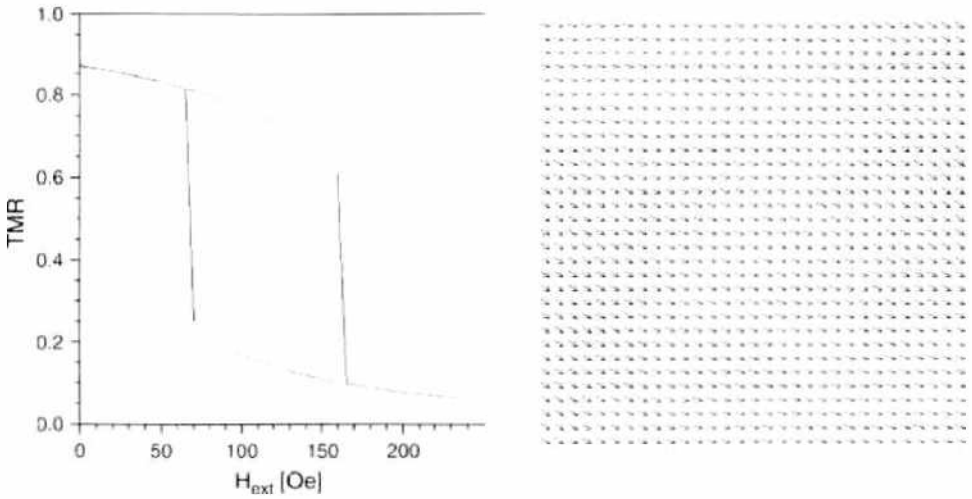


Figure 38. TMR curve (left) and initial state (right) at $H_{ext} = 240$ Oe of a very small TMR device ($50 * 50$ nm^2), hysteresis width: 92.5 Oe and shift: 115 Oe. Simulation parameters: grid $32 * 32$, space between AAF and soft layer: 4.3 nm. Layer parameters Py: $M_s = 800$ emu/cm^3, $A = 1$ μ erg/cm, $k_u = 310^3$ erg/cm^3, $t = 6$ nm, CoFe: $M_s = 1900$emu/cm^3, $A = 3$ μerg/cm, $k_u = 4 \times 10^5$ erg/cm^3, $t = 6$ Å, pinning field $H_{ex} = 3 \times 10^3$ Oe.

Figure 38 shows the expected exchange bias shift of the hysteresis loop along the magnetic field axis because of the demagnetizing fields of the AAF. The dimension of the device is small in comparison to the domain wall width; therefore, the characteristic shows no domain structures at all. Instead of the otherwise pronounced domains, only a slightly bending of the magnetic moments takes place; particularly at the edges, the layer behaves approximately as a single domain particle. The first magnetization reversal transition occurs at $70 \rightarrow 65$ Oe. The characteristics are dominated mainly by the two competing energy terms: external field,

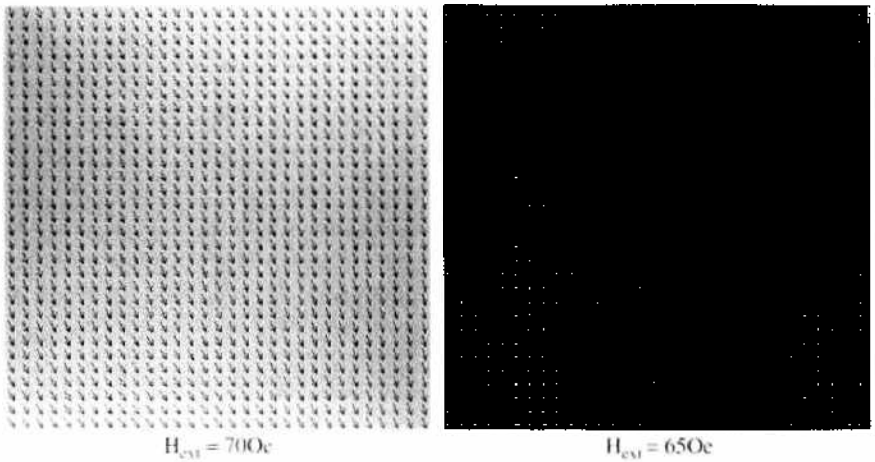


Figure 39. Spin patterns in a TMR device ($50 * 50$ nm^2) just before (left) and after (right) magnetization reversal. Simulation parameters as in Fig. 38.

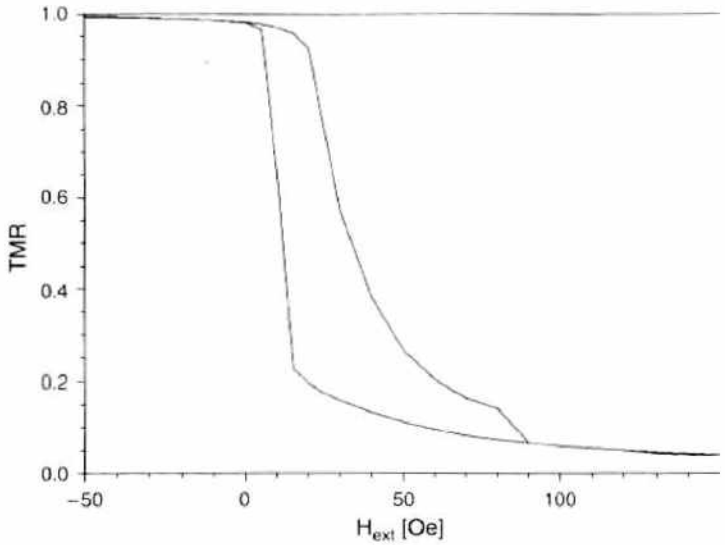


Figure 40. TMR device ($1.6 * 1.6\ \mu m^2$) with a pronounced asymmetry in the characteristic $\Delta R/R$, simulation with two layers, grid $64 * 64$. Layer parameters Py: $M_s = 800$ emu/cm^3, $A = 1$ μerg/cm, $k_u = 500$ erg/cm^3, $t = 5$ nm, CoFe: $M_s = 1900$ emu/cm^3, $A = 3$ μerg/cm, $k_u = 4 \times 10^5$ erg/cm^3, $t = 2$ nm, pinning field $H_{pi} = 500$ Oe.

demagnetizing field. A portion of approximately 10% remains for the exchange energy, and the anisotropy energy can be neglected completely.

Some experimental TMR curves show a significant asymmetry between both branches of the characteristic. This behavior is confirmed by the micromagnetic calculations shown in Fig. 40. The TMR values just before the jump are 0.226 (at 15 Oe in the left branch) and 0.924 (at 20 Oe in the right branch). The difference between saturation and the TMR values just before the jump is 3 [$= 0.226/(1 - 0.924)$] times larger in the left branch in comparison with the right one (case b in Fig. 36).

In the right branch the spins $\mathbf{M}_3$ are still aligned in the negative x-direction just before the transition; $\mathbf{M}_3$ is negative. This alignment is supported by the negative demagnetizing fields H_{af} from the AF; therefore, $\mathbf{M}_3$ and H_{af} are both negative (case b in Fig. 36). The magnetization is very large ($M_3 = -0.888$). In the left branch the spins are still aligned in the positive x-direction just before the transition; therefore, $\mathbf{M}_3$ and H_{af} point into opposite directions (case a in Fig. 36). The positive magnetization ($M_3 = 0.421$) is much smaller (compared to 0.888). The spin pattern Fig. 41 (left drawing) shows quite enormous deviations from the positive direction particularly in both edge domains $x \pm a$, where the magnitudes

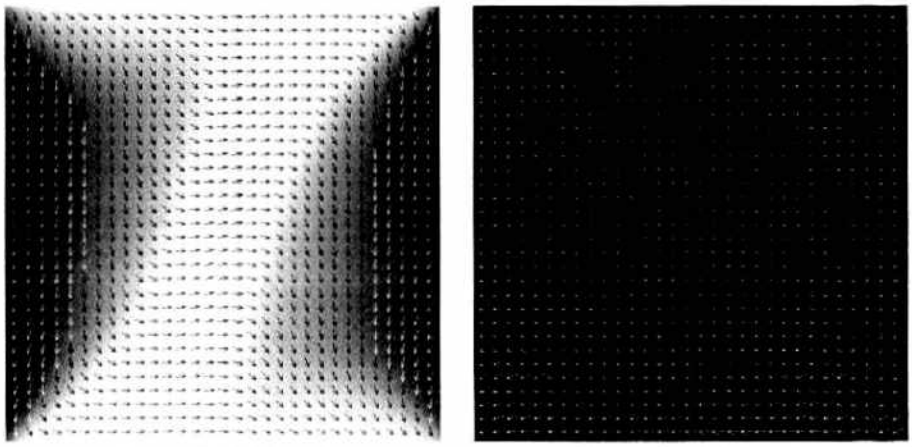


Figure 41. Spin patterns in a TMR device ($1.6 * 1.6\ \mu m^2$) just before magnetization reversal in the left (left drawing) and right (right drawing) branch of the hysteresis curve (see Fig. 40), after transition at 5 Oe the magnetic moments just point into the opposite direction. Simulation parameters as in Fig. 40.

of both stray fields in layer three (own stray field h_{ss} and foreign stray field H_{sf}) are very large; consequently, there must be very large spin bending in this region ($\varphi \approx -120°$). This asymmetry occurs in the micromagnetic calculation only if the contribution of the own stray field of layer three (see h_{ss} in Fig. 36) is taken into account; therefore, this effect cannot occur in the Stoner-Wohlfarth result in Fig. 37.

12. MOLECULE DETECTION WITH BEADS

In the last few years, magnetoresistive sensors have also been used as detection components in biological devices such as high-sensitivity biosensors based on a magnetic labeling technology [117–123]. The control of the movement of magnetically labeled biomolecules by on-chip currents and the use of SP sensors to detect the magnetic labels have been introduced recently and are further advances in the development of biochips [124, 125]. A conductor pair has been placed on both sides of the sensor with the possibility of parallel or series connection. This connector setup makes it possible first to magnetize the beads perpendicular (Fig. 42a) to or in (Fig. 42b) the plane of the sensor [126, 127] and, secondly, to move the beads over the sensor to one of the two conductors, where the bead motion causes the detection signal. This new scheme of combined movement and detection is the first but very essential step on the way to an on-chip laboratory.

A GMR sensor can detect a single paramagnetic bead of any size as long as some size conditions are met. Furthermore, the signal-to-noise ratio has been found to be greater than 5000:1, using reasonable assumptions about the GMR detector and bead properties [128]. A numerical model based on the assumptions of the equivalent average field of magnetic nanoparticles and the coherent magnetization rotation of the free layer in a SV sensor has been presented. Satisfactory signal linearity at low particle number also has been found [129, 130].

The microspheres are produced in a wide diameter range from several micrometers down to the nanometer scale. They provide a vehicle for the placement, detection, and study of other biomolecules and other biomolecular interactions; in particular, single DNA molecule interactions. The particles can behave superparamagnetically or ferromagnetically [84, 85] in dependence on the active volume size of the magnetic material. The beads can serve as markers to control the movement of molecules or as vehicles guided by external magnetic fields to drag the molecule chain. In this case a microarray of sensors is positioned along the channel in which the molecules are driven. A wide range of biotechnological applications has been opened, including miniaturized high-sensitivity biosensors and biochip devices. The sensors can be arranged in microarrays and can significantly improve the throughput, sensitivity, and accuracy of DNA sequencing and of biochemical analysis in general.

Micromagnetic calculations have been done for the sensor types: GMR and TMR. The sensor behavior for the detection of beads is very different in important points; some typical characteristics shall be cleared up in the next sections. As mentioned earlier, it has been

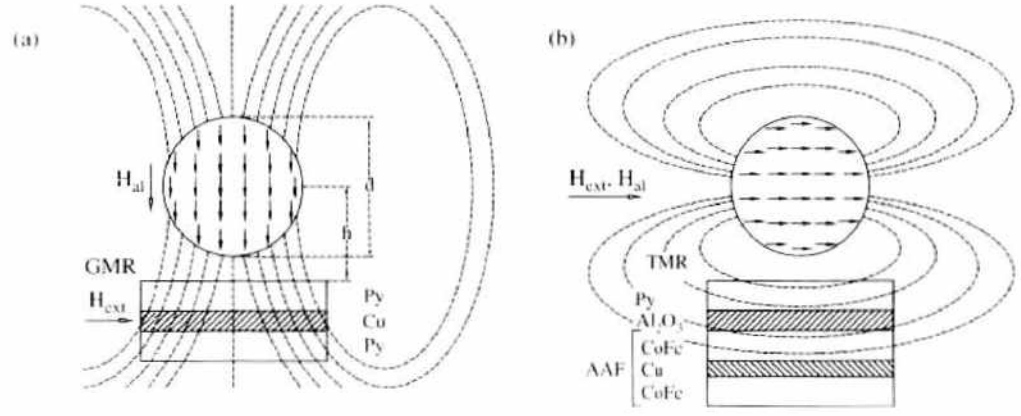


Figure 42. Bead field perpendicular to (a) or in (b) the plane of the sensor (GMR/SV/TMR), measuring field H_{ext} for the sensor signal, aligning field H_{sf} for the induced dipole moment $\mu_0\mu_m \sim H_{sf}$ of the bead material (e.g., magnetite Fe_3O_4), bead diameter d_b, bead radius $R_b = d_b/2$, bead height h_b.

shown [126, 127] that it is nowadays already possible to magnetize superparamagnetic beads perpendicular (Fig. 42a) to or in (Fig. 42b) the plane of the sensor, using a conductor pair placed on both sides of the sensor, and to detect the beads on their movement to one of the conductors. Therefore, it is not necessary to wait for ferromagnetic beads of the next bead generation. The in-plane calculations could already be realized by measurements with the present technology. That is the reason to simulate the behavior of an in-plane magnetic configuration.

Figure 43 shows the bead stray fields; the arrow lengths and angles correspond to the field strengths and directions at the arrow position. Figure 43 represents the fingerprints of the beads, which are very different in the two configurations a and b discussed earlier (see Fig. 42). The completely different field profiles in the out-of-plane (a) and in-plane (b) case were calculated as the magnetic field of a sphere [54] in the sensor area. There is a field with central symmetry in the out-of-plane case (a). In the other case, that of in-plane beads, the fields are in the opposite direction to the aligned bead within a small rectangular area (b). Therefore, averaging the bead fields over the sensor layer supplies a field shift ΔH_{ext} in the opposite direction to H_{ext}.

Single molecules (e.g., in biological experiments) can be detected by labeling them with magnetic beads instead of the otherwise-used dyes and by analyzing the influence of the beads on a GMR or TMR multilayer. The development of suitable sensors has been studied experimentally as well as theoretically by computer simulation to optimize the sensor parameters. A micromagnetic model with the dipole fields of the beads as additional contributions to the effective field $\vec{H}_{ijk}$ already has been used in the simulations very successfully. The bead fields produce vortex-like arrangements within the spin ensemble of a magnetic layer and change the spin relaxation process, and as a result also the GMR/TMR characteristic.

12.1. GMR Sensors

Figure 44 shows some $\Delta R/R(H_{ext})$ characteristics for a GMR sensor (see Fig. 45 for the geometry of the beads). The comparison to the dotted curve (without beads) shows the different behavior when the beads are aligned out of plane (solid) or in plane (broken) of the sensor layers. The geometry of the sensor and the beads (radius, height, number) and all the micromagnetic parameters are the same in both cases.

The broken curve for the out-of-plane case shows that the bead fields decrease mainly the GMR value at zero field. The calculated curve agrees qualitatively with the experimental one [83, 131], as the comparison between Figs. 45 and 46 demonstrates. That relates to the pronounced decrease of GMR at zero field as well as the slight increase at the intermediate region (70 Oe). The half-width of the experimental curve is smaller because of the

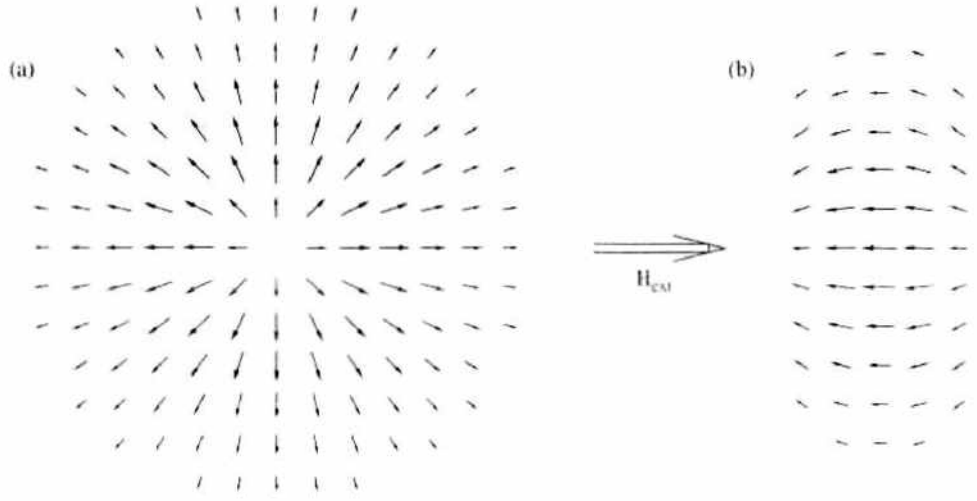


Figure 43. Fingerprints of the bead fields in the area of the sensor layer, the arrow lengths correspond to the field strengths. The completely different field profiles in the out-of-plane (a) and in-plane (b) case were calculated as the magnetic field of a sphere [54] in the sensor area.

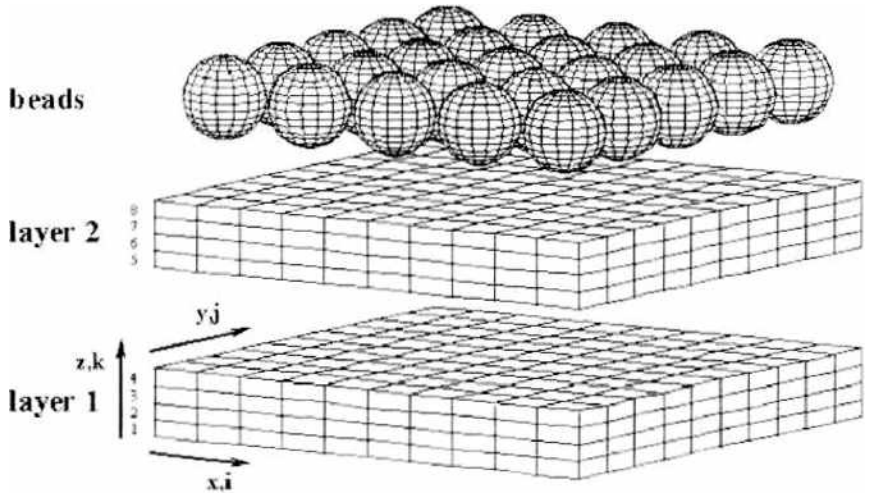


Figure 44. Sketch of the bead and the GMR sensor. 3D-view (left) and sectional view with the most essential dimensions (right). bead: diameter d_b, radius R_b, height h_b, sensor: thickness of magnetic layer t_m and spacer layer t_s. These are the significant values for the micromagnetic calculations as referenced in Figs. 45, 47.

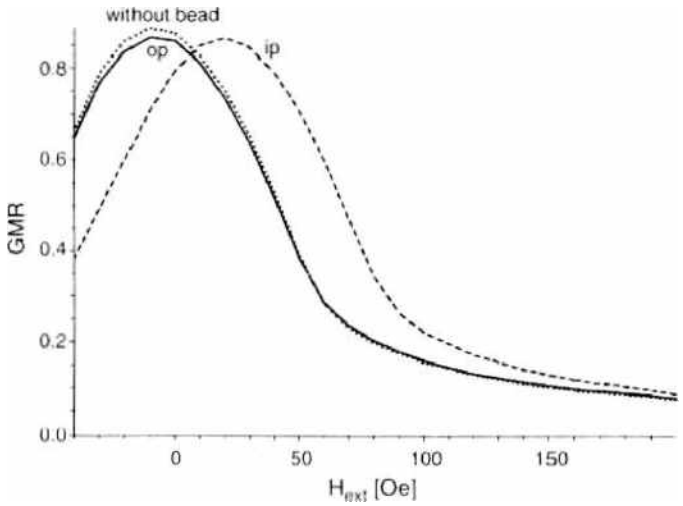


Figure 45. GMR $\Delta R/R\ (H_{\parallel x})$ in the case without (dotted) and with beads, beads aligned out of plane (solid, labeled with op) and in plane (broken, labeled with ip) of the sensor layers, sensor area: $900 * 900\ nm^2$. Simulation parameters: af stack: $M_s = 836\ emu/cm^3$, $J_1 = 3\ merg/cm^2$, $J_0 = 0$, $A = 1\ \mu erg/cm$, beads in a square $4 * 4$ array, bead: radius $R_b = 40$ nm, height $h_b = 150$ nm.

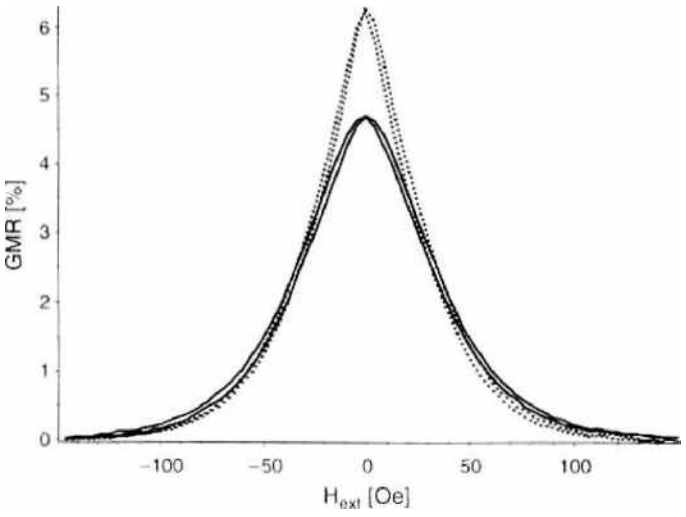


Figure 46. $\Delta R/R(H_{\parallel x})$ experimental GMR curves for the out-of-plane case with (solid) and without (dotted) bead exposure. Reprinted with permission from [83], W. Schepper et al., *J. Biotechnol.* 112, 35 (2004). © 2004, Elsevier.

inhomogeneous aligning field H_{al} [131], and apart from that, the special geometry of the sensor (1-μm-wide lines with a total length of about 2 mm wound into spirals) must still be taken into account in the simulation.

The calculations show for the in-plane case the expected enlargement of sensitivity and also the corresponding change of the GMR curve: a shift into the H_{ext} direction. This surprising offset must agree with the bead field component H_x [54], averaged over the whole sensor area ($4ab$):

$$H_x^{(bead)}(x, y, z) = 4\pi M_s R_b^3 \frac{x^2 - r^2/3}{r^5}; \quad r = \sqrt{x^2 + y^2 + z^2}$$

$$\Delta H_{ext}(h_b) = -\frac{n}{4ab}\int_{-b}^{b}\int_{-a}^{a} H_x(x, y, h_b)\, dx\, dy = \frac{4\pi n M_s R_b^3}{3(h_b^2 + a^2)\sqrt{h_b^2 + a^2 + b^2}} \tag{53}$$

where $r(x, y)$ is the distance between the center of the bead sphere and a lattice cell within the sensor layer, R is the bead radius, and h_b is its height. The integral will give the same result for different bead positions, if the bead fields decrease rapidly enough with distance. ΔH_{ext} is therefore directly proportional to the bead number n. Eq. (53) produces a shift of $\Delta H_{ext} = 24.4$ Oe with the parameters in Fig. 45 ($n = 16$), which agrees very well with the micromagnetic simulation (Fig. 45, broken curve) [132].

Figure 47 shows for the in-plane case a family of curves that shows the dependence on the bead exposure. The already-discussed parallel shift is very pronounced, which means there are no unsensitive parts in the GMR characteristic at all in this case.

It must still be pointed out that in the usual measuring scheme, at constant H_{ext} the resistance ΔR first arises with bead concentration, passes through a maximum value, and then decreases again. The parallel shift, however, seems to be linear up to large values of bead exposure. Therefore, the signal must be derived from the field change ΔH_{ext} at constant resistance ΔR, which can be done by modifying the usual measuring scheme improved by an additional control unit. Therefore, this result establishes a sensor for the bead detection, having an excellent linear characteristic.

The characteristic of an ideal detection system should be linear up to large values of bead exposure. A very interesting question is, can this be ensured in all cases in a real system; the answer has not been clarified completely. It is clear that all sensors work linear for small values of bead exposure, but what happens beyond. GMR sensors show in the out-of-plane case sensitive and unsensitive parts in the GMR characteristic. Moreover, the bead stray fields tend to wipe out within the sensor area, which effect prevents linearity. This can be

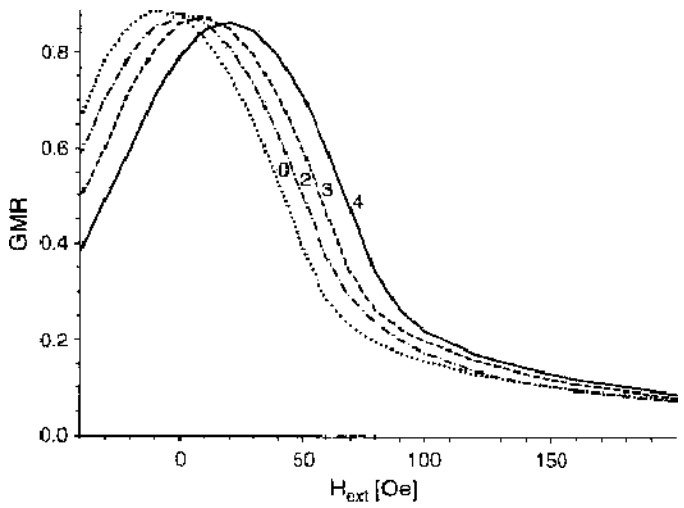


Figure 47. GMR $\Delta R/R$ (H_{ext}) in the case without (cotted) and with beads, beads aligned in plane. The curve parameter is the size of the square bead array $n * n$: $2 * 2$ (dash-dot), $3 * 3$ (broken), $4 * 4$ (solid). Simulation parameters as in Fig. 45. Reprinted with permission from [83], W. Schepper et al., *J. Biotechnol.* 112, 35 (2004). © 2004, Elsevier.

deduced immediately from the fingerprint in the left part of Fig. 43, two beads are positioned next to each other and the drawn field contributions are superimposed. This effect does not occur in the in-plane configuration, as a superposition with the field contribution of Fig. 43b shows. In contrast, TMR and SP sensors can show jumps in the detection characteristic. This can occur if the sensors work in the neighborhood of transition reversal to achieve a large sensitivity. We already have shown this effect in calculations in good agreement with measurements; this cannot happen with GMR sensors. Both sensors are complementary, and large values of sensitiveness can be obtained with the TMR sensor in the neighborhood of transition reversal; the advantage of the GMR sensor is the larger linearity region.

12.2. TMR Sensors

GMR, TMR, and SP devices have been used experimentally as sensors to detect molecules. Some features of the sensor characteristics for the molecule detection as linearity and sensitivity will be different. Therefore, it is necessary to analyze these three different sensor devices to find the best choice.

The special in this section is that the micromagnetic calculations (in plane) are used to look into the bead and sensor areas and to observe what is going on magnetically as a result of the possible interaction between the domains in the magnetic layers. The beads and the sensor are considered as a whole magnetic ensemble. The beads were simply simulated as square or rectangular platelets; there are three active magnetic layers: bead (Py), sensor (Py), and hard magnet (CoFe). The results are shown in the dashed ($8 * 8$ beads) and dotted ($8 * 6$ beads) curve of Fig. 49, compared with the solid curve (without beads; see Fig. 48 for the geometry of the beads).

All the curves in Fig. 49 are shifted in the direction of external field H_{ext} because of the stray field of the hard magnet; this is a well-known effect. The curves in Fig. 49 can be compared with experimental results in Fig. 50, even though the experimental curves belong to the out-of-plane case. This comparison may appear strange. TMR calculations for the out-of-plane case certainly are possible—they look quite similar to Fig. 49, but they do not permit a look at the internal spin structure of the beads at the moment. In contrast, there have been no measurements for the in-plane case until now.

Both families of curves in Figs. 49 and 50 agree qualitatively; the cases without beads show the distinct sharp transition during the magnetization reversal from a positive to a negative field direction (see Fig. 40), and a smooth one in reverse order [43]. In addition, the transition is smoothed out by the bead fields. The half-width of the experimental curve is much smaller because of the inhomogeneous aligning field H_{al} [131].

The difference between the curves with and without beads is more pronounced in the case of the dotted curve ($8 * 6$ rectangular beads) in comparison with the dashed ($8 * 8$ square beads) curve, in spite of the smaller number of beads ($48 \rightarrow 64$).

That is the influence of the larger shape anisotropy originating from magnetostatic properties. Figures 51 shows the corresponding spin patterns. The spins of almost all the beads

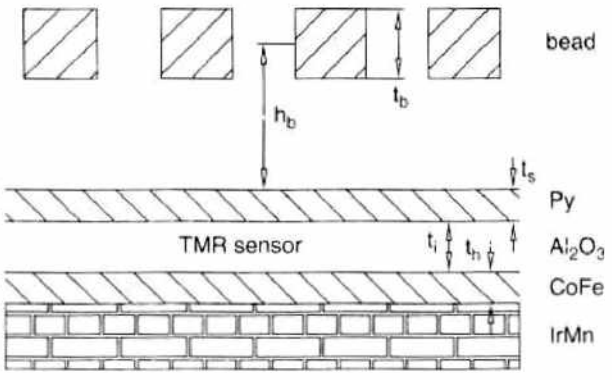


Figure 48. Sketch of the bead and the TMR sensor with the most essential dimensions: bead thickness t_b, bead height h_b, thickness of soft magnetic layer t_s, hard magnetic layer t_h and insulator t_i. These are the significant values for the calculations as referenced in Fig. 49.

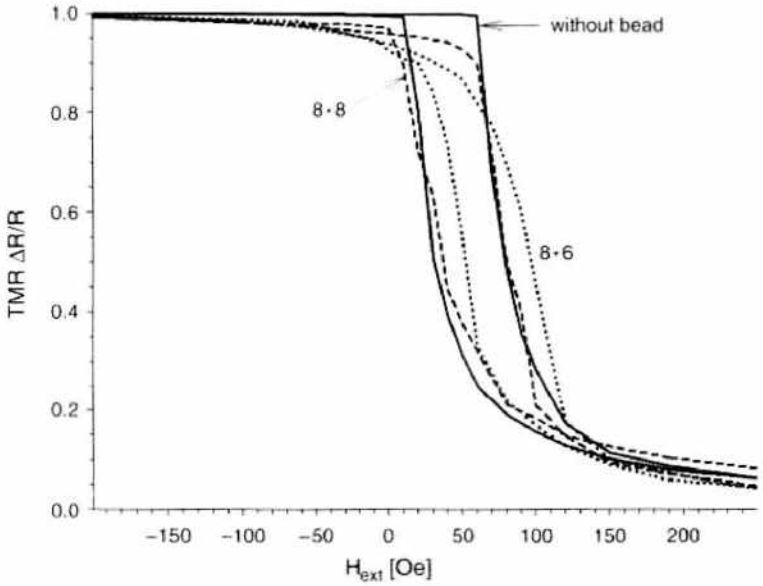


Figure 49. TMR $\Delta R/R$ (H_{ext}) calculation (in plane) without (solid) and with $8 * 8$ (short dash), respectively, $8 * 6$ beads (dotted), beads (one layer) and the sensor (two magnetic layers) are considered as a whole magnetic ensemble. Simulation parameters: $1.6 * 1.6$ μ^2 bead/sensor layer, $100 * 100$ nm^2 bead area, $t_b = t_s = 5$ nm, $t_b = 10$ nm, $h_b = 12.5$ nm, pinning field $H_{pi} = 500$ Oe. Material parameters Py: $M_s = 800$ emu/cm^3, $A = 1$ μerg/cm. CoFe: $M_s = 1900$ emu/cm^3, $A = 3$ μerg/cm. Reprinted with permission from [83], W. Schepper et al., *J. Biotechnol.* 112, 35 (2004). © 2004, Elsevier.

still point to the right (initial direction of external field) in the $6 * 8$ case (lower drawings) in spite of a vanishing external field ($H_{ext} = 0$). The larger anisotropy is responsible for the effect that the beads remain longer in the initial saturation state during transition to the other side of the hysteresis curve ($H_{ext} \to -\infty$). Otherwise, the upper drawings show spin down $\varphi = -90°$ or spin up $\varphi = 90°$ areas ($8 * 8$) instead of areas with the spins aligned to the external field $\varphi = 0$ (lower drawings, $6 * 8$). Because of the smaller shape anisotropy (upper drawings) the spins have been turned already by $\pm 90°$.

The influence of the magnetostatic properties requires two edge domains at the left and right side in Fig. 51. The graph also looks very similar in the saturation state $H_{ext} = 250$ Oe. No change would be expected in a complete layer without division in smaller bead areas, as far as only this feature is concerned.

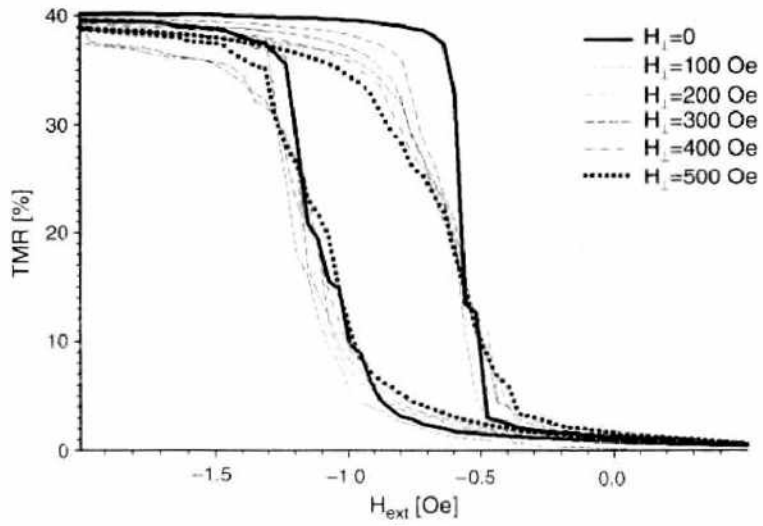


Figure 50. Experimental TMR $\Delta R/R$ (H_{ext}) curves for the out of plane case with the perpendicular field $H_{\perp}$ (Fig. 42) as parameter. Reprinted with permission from [83], W. Schepper et al., *J. Biotechnol.* 112, 35 (2004). © 2004, Elsevier.

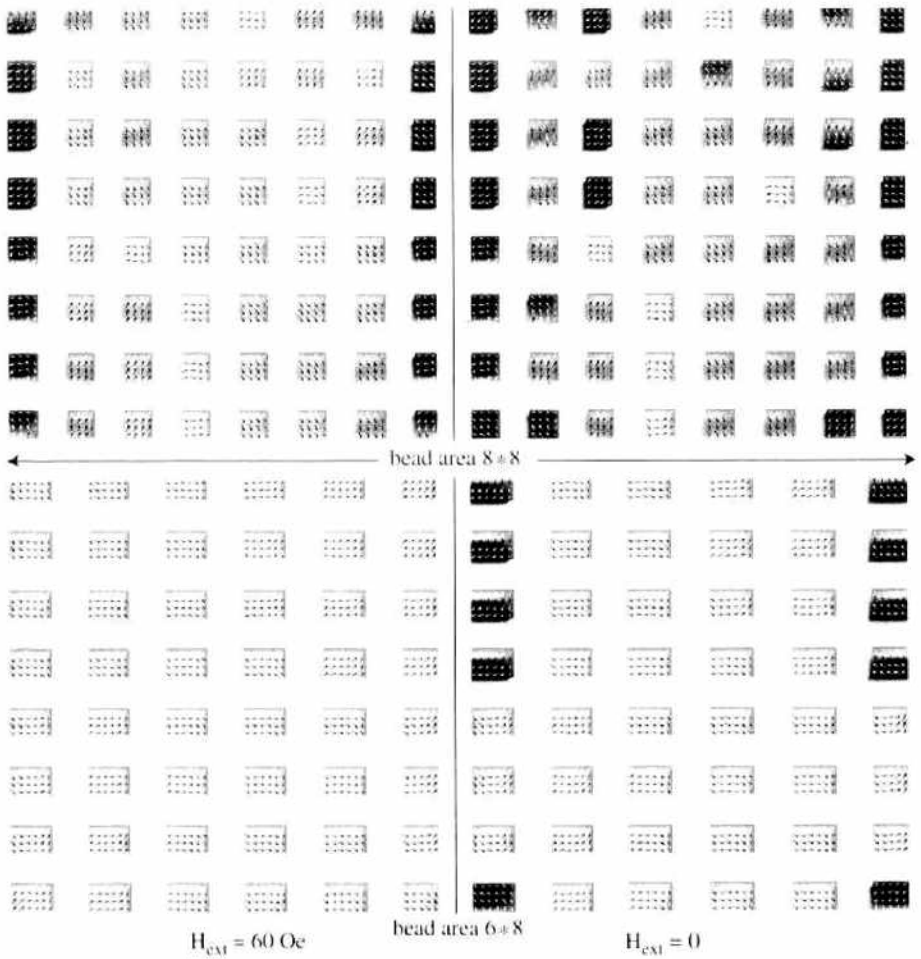


Figure 51. Spin patterns of the bead area for the dashed (bead area $8 * 8$) and dotted curve (bead area $6 * 8$) in Fig. 49; initial state $H_{ext} = 250$ Oe. Reprinted with permission from [83], W. Schepper et al., *J. Biotechnol.* 112, 35 (2004). © 2004, Elsevier.

The spin pattern in Fig. 52 shows the sensor area at $H_{ext} = 60$ Oe for the dotted curve in Fig. 49: the fingerprints of the beads in the sensor area are to be recognized clearly. The stripes perpendicular to the external field are a direct consequence of the rectangular field contribution in Fig. 43b. In comparison, the spin pattern in Figs. 52 shows three clearly structured domains in the so-called S- state. It is understandable that such a difference

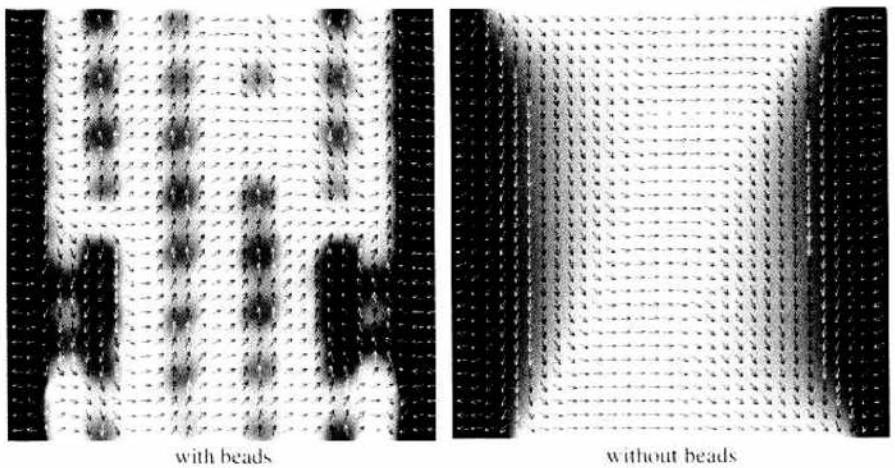


Figure 52. Spin patterns of the sensor area for the dotted (with beads) and solid curve (without beads) in Fig. 49 at $H_{ext} = 60$ Oe still before magnetization reversal, initial state $H_{ext} = 250$ Oe. Reprinted with permission from [83], W. Schepper et al., *J. Biotechnol.* 112, 35 (2004). © 2004, Elsevier.

between the spin distributions (Fig. 52) will give rise to corresponding modifications in the TMR characteristic (Fig. 49).

13. CONCLUSION

13.1. Sensor Design

We have shown that both magnetization as well as GMR can be numerically modeled for complex layer sequences, where different types of coupling are present within one film system. The numeric simulations can produce reliable results, as compared with experimental GMR-curves. Moreover, it is possible to tailor numerically the dependence of the resistance on the magnetic field by constructing film stacks with different sequences of strong and weak coupling. This procedure can be therefore very useful for the understanding of the magnetic and GMR properties of such a complex film system and in designing optimized sensors.

13.2. Magnetic Molecule Detection

The micromagnetic calculations show for the in-plane case the expected enlargement of sensitivity, and also the corresponding change of the GMR curve: a shift into the H_{ext} direction. This surprising offset must agree with the bead field component in the H_{ext} direction, averaged over the whole sensor area. The term ΔH_{ext} is proportional to the bead volume V_b and saturation magnetization M_s, and therefore to the bead surface concentration n_b. The integral will give the same result for different bead positions if the bead fields decrease rapidly enough with distance. Therefore, this result establishes an ideal sensor for the bead detection on account of the linearity of the detection characteristic. To achieve this advantage the signal must be derived from the field change ΔH_{ext} at constant resistance ΔR. That can be done by modifying the usual measuring scheme improved by an additional control unit.

In addition, we have shown the results of the magnetization behavior of GMR/TMR sensors, considering also the interaction between the domains in the magnetic layers of the sensor and the bead area. The calculations show a pronounced domain structure also in the bead area and suggest that the bead particles (signal source) and the magnetic layers (sensor device) should be considered as a whole magnetic ensemble in further micromagnetic calculations. Therefore, it is not sufficient to study only the sensor under the influence of the bead fields.

ACKNOWLEDGMENTS

First I thank my colleagues Günter Reiss, Andreas Hütten, Hubert Brückl, and Jan Schmalhorst from the Thin films and Nanostructures research group of the University of Bielefeld for many inspiring discussions about sensor devices and their modeling aspects, particularly concerning the Stoner-Wohlfarth model. I extend my thanks also to the whole research group for the access to corresponding experimental results, and above all to Sonja Heitmann, Jörg Schotter, and Tobias Hempel. Last but not least I would like to thank Riccardo Hertel at the Max-Planck-Institute of Microstructure Physics in Halle for many stimulating and helpful discussions about micromagnetics.

Our micromagnetic activities have been supported by the Deutsche Forschungsgemeinschaft through the "Collaborative Research Centre 613, Physics of single molecule processes and of molecular recognition in organic systems," project K4.

REFERENCES

1. P. Grünberg, R. Schreiber, Y. Pang, M. B. Brodsky, and H. Sowers, *Phys. Rev. Lett.* 57, 2442 (1986).
2. M. N. Baibich, J. M. Broto, A. Fert, F. Nguyen van Dau, F. Petroff, P. Etienne, G. Creuzet, A. Friederich, and J. Chazelas, *Phys. Rev. Lett.* 61, 2472 (1988).
3. S. S. P. Parkin, N. More, and K. P. Roche, *Phys. Rev. Lett.* 64, 2304 (1990).
4. P. Bruno and C. Chappert, *Phys. Rev. B* 46, 261 (1992).
5. J. Coehoorn, *Phys. Rev. B* 44, 44 (1991).
6. J. Kudrnovsky et al., *Phys. Rev. B* 53, 5125 (1996).

7. J. Mathon et al., *Phys. Rev. B* 56, 11797 (1997).
8. D. Edwards et al., *Phys. Rev. Lett.* 67, 493 (1991).
9. J. Mathon et al., *Phys. Rev. Lett.* 74, 3696 (1995).
10. H. Holloway and D. J. Kubinski, *J. Appl. Phys.* 83, 2705 (1998).
11. W. F. Brown, Jr., "Micromagnetics," John Wiley and Sons, 1963.
12. E. Della Torre, *Physica B* 343, 1 (2004).
13. S. M. Rubens, *IEEE Trans. Magn.* 15, 1192 (1979).
14. W. F. Brown, Jr., *IEEE Trans. Magn.* 20, 112 (1984).
15. U. Nowak, R. Chantrell, and E. C. Kennedy, *Phys. Rev. Lett.* 84, 163 (2000).
16. M. R. Gibbons, *J. Magn. Magn. Mat.* 186, 389 (1998).
17. A. Hubert and R. Schäfer, "Magnetic Domains. The Analysis of Magnetic Microstructures," Springer, 1998.
18. D. B. Fulghum and R. E. Camley, *Phys. Rev. B* 52, 13436 (1995).
19. W. Schepper, K. Diplas, and G. Reiss, *J. Appl. Phys.* 87, 6597 (2000).
20. K. Diplas, Ph.D Thesis, Department of Physics, University of Bielefeld, 2003.
21. X.-P. Wang, C. J. Garcia-Cervera, and E. Weinan, *J. Comp. Phys.* 171, 357 (2001).
22. C. J. Garcia-Cervera, Z. Gimbutas, and E. Weinan, *J. Comp. Phys.* 184, 37 (2003).
23. C. J. Garcia-Cervera and E. Weinan, *IEEE Trans. Magn. Mag.* 39, 1766 (2003).
24. W. Press, B. Flannery, S. Teukolsky, and W. Vetterling, "Numerical Recipes: The Art of Scientific Computing," Cambridge University Press, 1994.
25. K. Ramstöck, Ph.D Thesis, Faculty of Engineering, University of Erlangen-Nürnberg, 1997.
26. D. R. Fredkin and T. R. Koehler, *IEEE Trans. Magn.* 23, 3385 (1987).
27. D. R. Fredkin and T. R. Koehler, *J. Appl. Phys.* 63, 3179 (1988).
28. D. R. Fredkin and T. R. Koehler, *IEEE Trans. Magn.* 24, 2362 (1988).
29. D. R. Fredkin and T. R. Koehler, *IEEE Trans. Magn.* 25, 3473 (1989).
30. D. R. Fredkin and T. R. Koehler, *IEEE Trans. Magn.* 26, 415 (1990).
31. D. R. Fredkin and T. R. Koehler, *J. Appl. Phys.* 67, 5544 (1990).
32. D. R. Fredkin and T. R. Koehler, *IEEE Trans. Magn.* 26, 1518 (1990).
33. T. R. Koehler and D. R. Fredkin, *IEEE Trans. Magn.* 287, 4763 (1991).
34. T. R. Koehler and D. R. Fredkin, *IEEE Trans. Magn.* 28, 1239 (1992).
35. W. Chen, T. R. Koehler, and D. R. Fredkin, *IEEE Trans. Magn.* 29, 2124 (1993).
36. T. Schrefl, H. F. Schmidts, J. Fidler, and H. Kronmüller, *J. Magn. Magn. Mat.* 124, 251 (1993).
37. T. Schrefl, H. Kronmüller, and J. Fidler, *J. Magn. Magn. Mat.* 127, L273 (1993).
38. H. Kronmüller and T. Schrefl, *J. Magn. Magn. Mat.* 129, 66 (1994).
39. T. Schrefl, J. Fidler, and H. Kronmüller, *Phys. Rev. B* 49, 6100 (1994).
40. T. Schrefl, J. Fidler, and H. Kronmüller, *J. Magn. Magn. Mat.* 138, 15 (1994).
41. C. W. Gardiner, "Handbook of Stochastic Methods," Springer Berlin, 1985.
42. H. Kubota, G. Reiss, H. Brückl, W. Schepper, J. Wecker, and G. Gieres, *Jap. J. Appl. Phys.* 41, L180 (2002).
43. W. Schepper, J. Kubota, and G. Reiss, in "Analysis of the disturbing influence of stray fields in very small MRAM cells by computer simulation" (D. Shi, B. Aktas, L. Pust, and F. Mikailov, Eds.), Vol. 2593, p. 75, Nanostructured Magnetic Materials and Their Applications. Springer Lecture Notes in Physics, Springer Berlin, (2002).
44. OOMMF Vers. 1.2, http://math.nist.gov/oommf
45. LLG Vers. 2.56, http://llgmicro.home.mindspring.com/
46. MicroMagus Vers. 5.15, http://www.micromagus.de/
47. R. Dittrich, T. Schrefl, A. Thiaville, J. Miltat, V. Tsiantosa, and J. Fidler, *J. Magn. Magn. Mat.* 272–76, 747 (2004).
48. R. Hertel and J. Kirschner, *Phys. B Condens. Matter* 343, 206 (2004).
49. J. M. García, A. Thiaville, J. Miltat, K. J. Kirk, J. N. Chapman, and F. Alouges, *Appl. Phys. Lett.* 79, 656 (2001).
50. P. H. W. Ridley, G. W. Roberts, and R. W. Chantrell, *J. Appl. Phys.* 87, 1069 (2002).
51. D. V. Berkov, K. Ramstöck, and A. Hubert, *Phys. Stat. Sol. (a)* 137, 207 (1993).
52. D. V. Berkov and N. L. Gorn, *Phys. Rev. B* 57, 332 (1998).
53. A. J. Newell, W. Williams, and D. J. Dunlop, *J. Geophys. Res.* 98, 9551 (1993).
54. A. Aharoni, "Introduction to the Theory of Ferromagnetism," Oxford Science Publications, 2000.
55. A. J. Newell, D. J. Dunlop, and W. Williams, *J. Geophys. Res.* 98, 9533 (1993).
56. R. I. Joseph and E. Schlömann, *J. Appl. Phys.* 36, 1579 (1965).
57. M. E. Schabes and A. Aharoni, *IEEE Trans. Magn.* MAG-23, 3882 (1987).
58. M. E. Schabes and H. N. Bertram, *J. Appl. Phys.* 64, 1347 (1988).
59. M. E. Schabes and H. N. Bertram, *J. Appl. Phys.* 64, 5832 (1988).
60. A. Aharoni, *J. Appl. Phys.* 83, 3432 (1998).
61. R. Hertel, *J. Appl. Phys.* 90, 5752 (2001).
62. R. Hertel, *Z. Metallkd.* 93, 957 (2002).
63. H. Kronmüller and M. Fähnle, "Introduction to the Theory of Ferromagnetism," Cambridge University Press, 2003.
64. J. Fidler and T. Schrefl, *J. Phys. D* 33, R135 (2000).
65. J. Fidler, T. Schrefl, W. Scholz, D. Suess, R. Dittrich, and M. Kirschner, *J. Magn. Magn. Mat.* 272–76, 641 (2004).

66. R. Wieser, U. Nowak, and K. D. Usadel, *Phys. Rev. B* 69, 064401 (2004).
67. D. G. Porter and M. J. Donahue, *J. Appl. Phys.* 95, 6729 (2004).
68. R. Hertel and J. Kirschner, *J. Magn. Magn. Mat.* 278, L291 (2004).
69. A. Thiaville, J. M. García, R. Dittrich, J. Miltat, and T. Schrefl, *Phys. Rev. B* 67, 094410 (2003).
70. T. Okuno, K. Shigeto, T. Ono, K. Mibu, and T. Shinjo, *J. Magn. Magn. Mat.* 240, 1 (2002).
71. E. C. Stoner and E. P. Wohlfarth, *Phil. Trans. Roy. Soc. A* 240, 599 (1948). Reprint: *IEEE Trans. Magn.* 27, 3475 (1991).
72. W. Schepper, A. Hütten, and G. Reiss, *J. Appl. Phys.* 88, 993 (2000).
73. M. Jullière, *Phys. Lett. A* 54, 225 (1975).
74. J. O. Oti and E. Russek, *IEEE Trans. Magn.* 33, 3298 (1997).
75. A. Hütten, S. Mrozek, S. Heitmann, T. Hempel, H. Brückl, and G. Reiss, *acta mat.* 47, 4245 (1999).
76. S. Heitmann, Ph.D Thesis, Department of Physics, University of Bielefeld, 2004.
77. S. S. P. Parkin, "Ultrathin Magnetic Structures II" (B. Heinrich and J. A. C. Bland, Eds.), p. 148. Springer Berlin, 1994.
78. T. Luciński, F. Stobiecki, D. Elefant, D. Eckert, and G. Reiss, *J. Magn. Soc. Japan* 23, 126 (1999).
79. K. Pettit, S. Gider, S. S. P. Parkin, and M. B. Salamon, *Phys. Rev. B* 56, 7819 (1997).
80. M. Ye. Zhuravlev, W. Schepper, S. Heitmann, H. Vinzelberg, P. Zahn, I. Mertig, H. O. Lutz, A. V. Vedyayev, G. Reiss, and A. Hütten, *Phys. Rev. B* 65, 144428 (2002).
81. M. Ye. Zhuravlev, W. Schepper, S. Heitmann, H. O. Lutz, A. V. Vedyayev, G. Reiss, and A. Hütten, in "Model Calculation of the Giant Magnetoresistance in Multilayers with an Arbitrary Number of Layers" (D. Shi, B. Aktas, L. Pust, F. Mikailov, Eds.), 2593, pp. 43. Nanostructured Magnetic Materials and Their Applications. Springer Lecture Notes in Physics, Springer Berlin, (2002).
82. W. Schepper, J. Schotter, H. Brückl, and G. Reiss, *J. Magn. Magn. Mat.* 272–276S, E1695 (2004).
83. W. Schepper, J. Schotter, H. Brückl, and G. Reiss, *J. Biotechnol.* 112, 35 (2004).
84. D. Sudfeld, K. Wojczykowski, W. Hachmann, S. Heitmann, K. Rott, T. Hempel, S. Kämmerer, R. Jutzi, A. Hütten, and G. Reiss, *IEEE Trans. Magn.* 38, 2601 (2002).
85. D. Sudfeld, K. Wojczykowski, W. Hachmann, P. Jutzi, G. Reiss, and A. Hütten, *J. Appl. Phys.* 93, 7328 (2003).
86. A. Wachowiak, J. Wiebe, M. Bode, O. Pietzsch, M. Morgenstern, and R. Wiesendanger, *Science* 298, 577 (2002).
87. E. Feldtkeller and H. Thomas, *Phys. Kondens. Mat.* 4, 8 (1965).
88. D. Goll, G. Schütz, and H. Kronmüller, *Phys. Rev. B* 67, 094414 (2003).
89. H. Hoffmann and F. Steinbauer, *J. Appl. Phys.* 92, 5463 (2002).
90. M. R. Scheinfein and A. S. Arrott *J. Appl. Phys.* 93, 6802 (2003).
91. A. S. Arrott, *Z. Metallkd.* 93, 963 (2002).
92. H. Koop, H. Brückl, D. Meyners, and G. Reiss, *J. Magn. Magn. Mat.* 272–276S, E1475 (2004).
93. H. Koop, Ph.D Thesis, Department of Physics, University of Bielefeld, 2004.
94. C. A. Ross, M. Farhoud, M. Hwang, H. I. Smith, M. Redjdal, and F. B. Humphrey, *J. Appl. Phys.* 89, 1310 (2001).
95. C. A. Ross, M. Hwang, M. Shima, J. Y. Cheng, M. Farhoud, T. A. Savas, H. I. Smith, W. Schwarzacher, F. M. Ross, M. Redjdal, and F. B. Humphrey, *Phys. Rev. B* 65, 144417 (2002).
96. Y. Xu, Z. G. Sun, Y. Qiang, and D. J. Sellmyer, *J. Magn. Magn. Mat.* 266, 164 (2003).
97. B. Rellinghaus, S. Stappert, M. Acet, and E. F. Wassermann, *J. Magn. Magn. Mat.* 266, 142 (2003).
98. W. Scholz, K. Y. Guslienko, V. Novosad, D. Suess, T. Schrefl, R. W. Chantrell, and J. Fidler, *J. Magn. Magn. Mat.* 266, 155 (2003).
99. W. Folkerts, *J. Magn. Magn. Mat.* 94, 302 (1991).
100. H. Fujiwara and M. R. Parker, *J. Magn. Magn. Mat.* 135, L23 (1994).
101. J. García-Otero, M. Porto, J. Rivas, and A. Bunde, *J. Appl. Phys.* 85, 2287 (1999).
102. Z. Zhang, L. Zhou, and P. E. Wigen, *Phys. Rev. B* 50, 6094 (1994).
103. H. Fujiwara and M. R. Parker, *J. Magn. Magn. Mater.* 135, L23 (1994).
104. H. A. M. van den Berg, W. Clemens, G. Gieres, G. Rupp, M. Vieth, J. Wecker, and S. Zoll, *J. Magn. Magn. Mat.* 165, 524 (1997).
105. M. Perez, C. H. Marrows, and B. J. Hickey, *J. Appl. Phys.* 89, 7116 (2001).
106. J. S. Moodera and G. Mathon, *J. Magn. Magn. Mat.* 200, 248 (1999).
107. S. S. P. Parkin, K. P. Roche, M. G. Samant, P. M. Rice, R. B. Beyers, R. E. Scheuerlein, E. J. O'Sullivan, S. L. Brown, J. Bucchigano, D. W. Abraham, Y. Lu, M. Rooks, P. L. Trouilloud, R. A. Wanner, and W. J. Gallagher, *J. Appl. Phys.* 85, 5828 (1999).
108. Z. Z. Zhang, S. Cardoso, P. P. Freitas, P. Wei, N. Barradas, and J. C. Soares, *Appl. Phys. Lett.* 78, 2911 (2001).
109. Y. Ando, H. Kubota, M. Hayashi, M. Kamijo, K. Yaoita, A. Chak Chung Yu, Xiu-Feng Han, and T. Miyazaki, *Jap. J. Appl. Phys.* 39, 5832 (2000).
110. M. R. McCartney, R. E. Dunin-Borkowski, M. R. Scheinfein, D. J. Smith, S. Gider, and S. S. P. Parkin, *Science* 28, 1337 (1999).
111. A. Anguelouch, B. D. Schrag, G. Xiao, Y. Lu, P. L. Trouilloud, R. A. Wanner, W. J. Gallagher, and S. S. P. Parkin, *Appl. Phys. Lett.* 76, 622 (2000).
112. J. Schmalhorst, Ph.D Thesis, Department of Physics, University of Bielefeld, 2001.
113. A. Thomas, Ph.D Thesis, Department of Physics, University of Bielefeld, 2003.
114. C. Tiusan, T. Dimopoulos, K. Ounadjela, and M. Hehn, H. A. M. van den Berg, V. da Costa, and Y. Henry, *Phys. Rev. B* 61, 580 (2000).

115. T. Schrefl, J. Fidler, J. N. Chapman, and K. J. Kirk, *J. Appl. Phys.* 89, 7000 (2001).
116. U. Nowak, A. Misra, and K. D. Usadel, *J. Appl. Phys.* 89, 7269 (2001).
117. D. L. Graham, H. Ferreira, J. Bernardo, P. P. Freitas, and J. M. S. Cabral, *J. Appl. Phys.* 91, 7786 (2002).
118. J. Schotter, P. B. Kamp, A. Becker, A. Pühler, D. Brinkmann, W. Schepper, H. Brückl, and G. Reiss, *IEEE Trans. Magn.* 38, 3365 (2002).
119. J. Schotter, P. B. Kamp, A. Becker, A. Pühler, G. Reiss, and H. Brückl, *Biosensors & Bioelectronics* 19, 1149 (2004).
120. M. Panhorst, P. Kamp, G. Reiss, H. Brückl, *Biosensors & Bioelectronics* 20, 1685 (2004).
121. H. Brückl, M. Brzeska, D. Brinkmann, J. Schotter, G. Reiss, W. Schepper, P.-B. Kamp, and A. Becker, *J. Magn. Magn. Mat.* 282, 219 (2004).
122. M. Brzeska, M. Panhorst, P. B. Kamp, J. Schotter, G. Reiss, A. Pühler, A. Becker, and H. Brückl, *J. Biotechnol.* 112/1–2, 25 (2004).
123. M. Brzeska, M. Justus, J. Schotter, K. Rott, G. Reiss, and H. Brückl, *Molecular Phys. Rep.* 39, 32 (2004).
124. H. A. Ferreira, D. L. Graham, P. P. Freitas, and J. M. S. Cabral, *J. Appl. Phys.* 93, 7281 (2003).
125. D. L. Graham, H. A. Ferreira, P. P. Freitas, J. M. S. Cabral, *Biosensors & Bioelectronics* 18, 483 (2003).
126. L. Lagae, R. Wirix-Speetjens, J. Das, D. Graham, H. Ferreira, P. P. Freitas, G. Borghs, and J. De Boeck, *J. Appl. Phys.* 91, 7445 (2002).
127. R. Wirix-Speetjens, "On-chip Magnetic Particle Manipulation and Detection for Biochip Applications," contribution 4SY4, ICM (2003).
128. M. Tondra, M. Porter, and R. J. Lipert, *J. Vac. Sci. Technol. A* 18, 1125 (2002).
129. G. Li and S. X. Wang, *IEEE Trans. Magn.* 39, 3313 (2003).
130. G. Li, S. X. Wang, and S. Sun, *IEEE Trans. Magn.* 40, 3000 (2004).
131. J. Schotter, Ph.D Thesis, Department of Physics, University of Bielefeld, 2004.
132. W. Schepper, J. Schotter, H. Brückl, and G. Reiss, *Physica B* 372, 337 (2006).
133. W. F. Brown and A. E. LaBonte, *J. Appl. Phys.* 36, 1380 (1965).
134. A. E. LaBonte, *J. Appl. Phys.* 40, 2450 (1969).
135. A. Holz, *Z. Angew. Physik* 23, 170 (1967).
136. A. Holz, *Phys. Stat. Sol.* 25, 567 (1968).
137. A. Holz, *Phys. Stat. Sol.* 26, 751 (1968).
138. A. Hubert, *Phys. Stat. Sol.* 32, 519 (1969).
139. A. Holz and A. Hubert, *Z. Angew. Physik* 26, 145 (1969).
140. L. Landau and E. Lifshitz, *Physikal. Z. Sowjetunion* 8, 153 (1935).
141. T. L. Gilbert, *Phys. Rev.* 100, 1243 (1955).
142. M. R. Scheinfein and J. L. Blue, *J. Appl. Phys.* 69, 7740 (1991).

CHAPTER 11

Computational Micromagnetics

Josef Fidler, Thomas Schrefl, Werner Scholz
Vienna University of Technology, Institute for Solid State Physics, Vienna, Austria

CONTENTS

1. INTRODUCTION

The effects of magnetism and magnetic materials have been exploited since the invention of the magnetic compass for navigation in the 10th century in China. Micromagnetics is a phenomenological theory and its concept is based on the work of Landau and Lifshitz [1] as well as Brown [2, 3]. It is a continuum theory, which means that the discretized magnetic moments at the positions of the atoms are replaced by the continuum function of the magnetization. Therefore this theory is a suitable framework to investigate the magnetic phenomena in the nanometer regime. Computational micromagnetics leads to a deeper understanding of hysteresis effects at an intermediate length scale between magnetic domains and atomic distances by visualization of the magnetization reversal process of hard magnets, soft magnets, and materials for magnetic recording. At present, many processes accomplished in sensor technology (e.g., automotive technology) or data storage (hard disk drives) are based on discoveries of novel properties of magnetic films, multilayers, and micro- or nanostructures.

ISBN: 1-58883-049-7

Handbook of Theoretical and Computational Nanotechnology
Edited by Michael Rieth and Wolfram Schommers
Volume 7: Pages (623–675)

All these results are in the field of nanomagnetism and spin electronics composed of ferromagnetic, antiferromagnetic, and nonmagnetic, either metallic or insulating regions. Nanofabrication and the trend toward nanocrystalline magnetic materials open new opportunities for engineering innovative magnetic materials. The need for a better understanding of the magnetization reversal processes and the improved availability of large-scale computer power are the main reasons micromagnetic modeling has been developing extremely rapidly.

In 1956 IBM introduced the 305 RAMAC (random access method for accounting and control) with a capacity of 5 MB. It was the first magnetic storage device, which stored digital information by writing magnetization patterns on a "hard disk" using a thin film of granular magnetic material instead of magnetic tapes. One of the major advantages was that any position on the disk could be directly accessed by the read-write heads and it was not necessary to wind a tape any more. Since then many new discoveries and developments have improved computer hard disks. The areal density determines the amount of information, that can be stored on a given area of a hard disk. This figure of merit measures the performance of hard disk media in a similar way such as the energy density product for permanent magnetic materials. And it has shown a similar development during the past 20 years. In 1985 the typical areal density in mass production was 20 Mbits/in^2. The industry trend showed a typical increase of about 27% per year. In 1992 it exceeded 100 Mbits/in^2 and the annual growth rate jumped to approximately 50% per year. Starting with the introduction of giant magnetoresistive read-write heads in 1997, the areal density has been doubling every year [4]. Present-day drives have an areal density of about 30 Mbits/in^2, and read-write heads based on the extraordinary magnetoresistance effect will allow data densities beyond 100 Mbits/in^2 [5, 6]. As an alternative to these thin-film granular magnetic storage media, ferromagnetic nanostructures are considered for the basic information storage elements in magnetic random access memories (MRAM), high-density magnetic storage media, and magnetic sensors. These structures can be produced using well-established techniques for semiconductors, but they have several advantages over today's semiconductor-based materials, including nonvolatility, nondestructive readout, radiation hardness, low voltage, and unlimited read and write endurance [7, 8]. The magnetic properties, switching behavior, and switching dynamics of magnetic nanostructures (i.e., of individual grains and particles, thin films, and bulk nanocrystalline materials) are of great interest, and the results of numerical micromagnetic simulations are presented in Section 3.1–3.5.

The progress in the field of permanent magnets is very well illustrated by the maximum energy density product $(BH)_{max}$. By the end of the 19th century, magnetic steels with a $(BH)_{max} \approx 2$ kJ/m^3 were available. AlNiCo precipitation hardened magnets, which were discovered by Mishima in 1931, led to energy density products as high as 90 kJ/m^3 by 1955. In the late 1940s, hardferrites (ceramic oxides) were developed, and they are still commonly used because of the great abundance of their raw materials and low price. A major breakthrough was the discovery of magnetocrystalline anisotropy in rare-earth intermetallic compounds in the 1960s. Strnat and coworkers [9] found that the combination of the high-magnetic moment of iron and cobalt together with the high-magnetocrystalline anisotropy caused by rare-earth elements gives permanent magnetic materials with excellent properties and energy density products of greater than 90 kJ/m^3. The highest energy density products so far have been obtained with rare-earth iron-based permanent magnets. In 1984 Sagawa et al. [10], Croat et al. [11], and Hadjipanayis [12] exceeded energy density products of 300 kJ/m^3 for a material based on $Nd_2Fe_{14}B$. In the following years, continuing improvement of the production route has resulted in energy density products in excess of 420 kJ/m^3 [13–15]. Especially SmCo-based materials retain a high magnetic-ordering temperature, which makes them suitable for high-temperature applications. For the development and optimization of permanent magnetic materials for application temperatures of up to 500 °C, concentrated on SmCo-based precipitation hardened materials. The nucleation of reversed domains at grain boundaries and defects and the expansion of domains and their interaction with various phases determine the coercivity of modern rare-earth permanent magnets. Micromagnetic simulations have been carried out, and the results of the interaction between magnetic domain walls and a complex precipitation structure determine the coercivity of the magnet material [16].

Both research areas, permanent magnets and magnetic storage media, have made tremendous progress during the past decades, and they have become key technologies in today's information society. In order to keep the pace of these developments and push the limits, many research projects are carried out worldwide. The properties of modern magnetic materials are strongly influenced by their microstructure. The continuing improvement of the properties of SmCo-based magnets has been made possible by additives like Cu and Zr and a refined processing route and heat treatment, which has a great impact on the resulting cellular precipitation structure [17]. The typical size of the cells is in the order of 100 nm with an intercellular phase of around 10 nm. The particle size in magnetic recording tapes is of the same order of magnitude. The typical grain size in current hard-disk storage media is about 8 nm with a Cr-enriched intergranular region of about 2 nm for exchange decoupling of the grains. These structures are so small that quantum mechanical effects–like exchange have to be taken into account. However, they are too large for a pure quantum mechanical description, which would exceed the capabilities of today's *ab-initio* computational models. On this intermediate level between the macroscopic world and a description with atomic resolution, micromagnetic models have proved to be a useful tool [18]. These computational models provide great freedom in the choice of experiment conditions and in the variation of material parameters. In addition to measurements of the remanent magnetization and the coercive field, it is possible to study the details of the magnetization distribution and the magnetization reversal processes, which are difficult to investigate experimentally.

Computational micromagnetics leads to a deeper understanding of hysteresis effects at an intermediate length scale between magnetic domains and atomic distances by visualization of the magnetization reversal process. By the switching of small particles, thin-film elements, and nanowires become increasingly important in magnetic storage and magnetoelectronic devices. The magnetization reversal processes are studied using a three-dimensional hybrid finite element/boundary element micromagnetic model. Transient magnetization states during switching are investigated numerically in thin NiFe, Fe, and Co nanoelements of various shapes [19]. Switching dynamics are calculated for different external field profiles and frequencies [20]. The numerical solution of Brown's equations can be effectively performed using finite-element and related methods that easily handle complex microstructures and take into account the long-range magnetostatic interactions and short-range exchange coupling between the grains in granular thin films. Dynamic finite-element simulations successfully predict the influence of microstructural features such as grain size, particle shape, intergranular phases, and surface irregularities on the magnetic properties. Theoretical limits for remanence, coercive field, switching behavior at a short time scale of less than 1 ns, and other properties have successfully been calculated for a large number of materials. Topics of active research include the switching dynamics of patterned mesoscopic and nanoscopic elements, including the thermal activation process; the remanence enhancement in exchange-coupled, nanocrystalline magnets; the nucleation field of highest-energy-density $Nd_2Fe_{14}B$ magnets; and the domain wall pinning in $SmCo_5/Sm_2Co_{17}$–based magnets for high-temperature applications. The characteristic length scale, the domain wall width, is in the order of 5 to 10 nm. The size of the basic structural units of magnetic devices, such as sensor or storage elements, may extend toward micrometers. The time scale ranges from the subnanosecond regime for fast precessional switching to years for thermally activated magnetization reversal. The wide range of time and length scales involved in micromagnetic simulations is a challenge for effective computational tools.

In this work, temperature-dependent effects have not been considered explicitly for simplicity, but they are included in the temperature-dependent material parameters. Also eddy current effects, which should be taken into account in materials with high conductivity and in high-speed switching experiments, are only implicitly included in the Gilbert damping constant. In various fields of computer-aided engineering, such as structural analysis, fluid dynamics, and electromagnetic field computation, as well as micromagnetics [21], the finite-element method has been successfully applied. Its flexibility in modeling arbitrary geometries especially has made it very popular. In the light of the importance of the microstructure of magnetic materials, the finite element method has been chosen for the implementation of a micromagnetic model. There are several commercial and open-source micromagnetics

packages available; however, all of them use the finite difference method:

Commercial Codes

- LLG Micromagnetics Simulator, developed by M. R. Scheinfein, http://llgmicro.home.mindspring.com/ available for MS Windows (dual CPU supported).
- MagFEM3D, developed by K. Ramstöck, http://www.ramstock.de/ available on Unix, MS Windows platforms.
- Magsimus, Euxine Technologies, http://www.euxine.com/, available for MS Windows platforms.
- MicroMagus, developed by D. V. Berkov, N. L. Gorn, http://www.micromagus.de/ available for MS Windows platforms.

Free Open Source Codes

- PC Micromagnetic Simulator (SimulMag), developed by John Oti, http://math.nist.gov/oommf/contrib/simulmag/ available for MS Windows platforms.
- General Dynamic Micromagnetics (GDM2), developed by Bo Yang, http://physics.ucsd.edu/~drf/pub/
- Object Oriented MicroMagnetic Framework (OOMMF), developed by Mike Donahue and Don Porter, http://math.nist.gov/oommf/ available on Unix, MS Windows platforms.
- Parallel Finite Element Micromagnetics Package MAGPAR, developed by Werner Scholz, http://magnet.atp.tuwien.ac.at/scholz/MAGPAR/ available on Unix, MS Windows platforms.

In addition, the work on large systems required static energy minimization methods such as for the study of bulk SmCo permanent magnets as well as dynamic time integration methods for the investigation of the magnetization dynamics in magnetic nanoparticles. Therefore, a finite element micromagnetics package MAGPAR (http://magnet.atp.tuwien.ac.at/scholz/magpar/) has been implemented that combines several unique features. It is

- entirely based on portable, free, open source software packages,
- highly portable to different hardware platforms, which range from simple PCs to massively parallel supercomputers,
- highly optimized and scalable,
- well integrated, combining static energy minimization, dynamic time integration, and nudged elastic-band methods.

An introduction to the finite element method is given in Section 2.1. The basic micromagnetic equations and their discretization in the context of the finite element method are outlined in Section 2.2, while the appropriate solution methods are described in Section 2.3.

2. NUMERICAL FINITE ELEMENT MICROMAGNETIC TECHNIQUE

A parallel finite element micromagnetics package has been developed, which is highly scalable, easily portable, and combines different solvers for the micromagnetic equations. The micromagnetic technique is based on the standard Galerkin discretization on tetrahedral meshes with linear basis functions. A static energy minimization, a dynamic time integration, and the nudged elastic-band method have been implemented. In the light of the importance of the granular microstructure of magnetic materials, the finite element method has been chosen. For the calculation of the demagnetizing field, a hybrid finite element/boundary element method is used.

The resolution of the finite element mesh—the maximum size of the finite elements—is determined by the smallest features that might occur in the solution of the partial differential equation (PDE). In the most general case, the maximum cell size of the finite element mesh has to be smaller than the minimum of the three typical critical lengths [22], such as the width of the Bloch-type magnetic domain wall (δ_{BW}), typically found in hard bulk magnets, the width of Néel walls (δ_{NW}), typically found in soft magnetic materials and thin films, and

the thermal exchange length (l_{thex}) with the thermal field H_{th} [23]:

- width of Bloch-type magnetic domain walls,

$$l_{bex} = \sqrt{\frac{A}{K_1}}, \quad \delta_{BW} = \pi \cdot l_{bex} \tag{1}$$

  A is the exchange constant and K_1 is the magnetocrystalline anisotropy constant.
- width of Néel walls,

$$l_{nex} = \sqrt{\frac{2\mu_0 A}{J_s^2}}, \quad \delta_{NW} = \pi \cdot l_{nex} \tag{2}$$

  J_s is the saturation polarization.
- thermal exchange length (δ_{thex}) [23],

$$l_{thex} = \sqrt{\frac{A}{J_s \cdot H_{th}}}, \quad H_{th} = \sqrt{\frac{2\alpha k_B T}{\Delta t \gamma J_s l^3}} \tag{3}$$

The thermal field H_{th} depends on the Gilbert damping constant α, the temperature T, the time step of the numerical time integration scheme Δt, the gyromagnetic ratio γ, and the spatial correlation length l, which is equal to the cell size.

The thermal fluctuation field depends on the cell volume. The thermal exchange length depends on the fluctuation field and therefore it is cell-size dependent [23]. In most cases, the finite element size is in the order of 2–5 nm.

The total energy of a micromagnetic system is given by the Gibbs free energy, which depends on the magnetic polarization, the external field, and temperature-dependent material parameters. It includes macroscopic contributions, such as the Zeeman energy and the magnetostatic energy, as well as microscopic contributions, such as the magnetocrystalline anisotropy energy and the exchange energy. The minimization of the total energy yields an equilibrium magnetization distribution. However, the energy landscape of micromagnetic systems is complex and contains many local maxima, minima, and saddle points. Therefore, the choice of the initial magnetization distribution has a strong influence on the result. A more physical and realistic approach of the system to its equilibrium in a local minimum is provided by a dynamic description of the path through the energy landscape. The motion of a magnetic moment in a magnetic field is mainly governed by its Larmor precession around the local magnetic field. The damping of the precession causes the relaxation to equilibrium. There are many processes that contribute to the damping in a magnetic solid like magnon-magnon and magnon-phonon interactions, interactions between localized and itinerant electrons, and eddy currents, for example.

The Gilbert equation describes the precession of the magnetic polarization toward equilibrium and combines all damping effects in a phenomenological damping term with a single damping constant α [24]:

$$\frac{d\vec{J}}{dt} = -|\gamma|\vec{J} \times \vec{H}_{eff} + \frac{\alpha}{J_s}\vec{J} \times \frac{d\vec{J}}{dt} \tag{4}$$

At each time step, the effective field H_{eff} includes the applied field, the exchange field, the magnetocrystalline anisotropy field, and the demagnetizing field. To solve the Gilbert equation numerically, the magnetic volume is divided in finite elements.

2.1. The Finite Element Method

For the calculation of the demagnetizing field, a hybrid finite element/boundary element method is used, as explained in Section 7. This method requires the solution of a Poisson and a Laplace equation. Therefore the former is used in this chapter as an example for a short introduction to the finite element method [25].

2.1.1. Poisson Problem

We want to calculate a numerical solution U, which approximates the true solution u of the Poisson (boundary value) problem (P) in the solution domain $\Omega \subset R^3$ with closed boundary Γ. Dirichlet boundary conditions apply on $\Gamma_D \subset \Gamma$, and Neumann boundary conditions apply on $\Gamma_N := \Gamma \backslash \Gamma_D$.

The Poisson problem (P) is defined as follows: Given $f \in L^2(\Omega)$, $u_D \in H^1(\Omega)$, and $g \in L^2(\Gamma_N)$, we are searching for the solution $u \in H^1(\Omega)$, which satisfies the Poisson equation

$$-\Delta u = f \quad \text{in } \Omega \tag{5}$$

with Dirichlet boundary conditions

$$u = u_D \quad \text{on } \Gamma_D \tag{6}$$

and Neumann boundary conditions

$$\frac{\partial u}{\partial n} = g \quad \text{on } \Gamma_N \tag{7}$$

2.1.2. The Weak Formulation

The weak formulation of the boundary value problem (P) is then obtained by the multiplication of Eq. (5) with $w \in H_D^1(\Omega) := \{w \in H^1(\Omega) | w = 0 \text{ on } \Gamma_D\}$ and integration over Ω:

$$-\int_\Omega \Delta u \cdot w dv = \int_\Omega f . w dv \tag{8}$$

Integration by parts gives

$$\int_\Omega \nabla u \cdot \nabla w dv - \int_{\Gamma_N} \partial u \, \partial n \cdot w da = \int_\Omega f \cdot w dv \tag{9}$$

and substitution of the boundary conditions and rearrangement leads to

$$\int_\Omega \nabla u \cdot \nabla w dv = \int_\Omega f . w dv + \int_{\Gamma_N} g . w da \tag{10}$$

Now we incorporate the (possibly inhomogeneous) Dirichlet boundary conditions

$$\int_\Omega \nabla u \cdot \nabla w dv - \int_\Omega \nabla u_D \cdot \nabla w dv = \int_\Omega f \cdot w dv + \int_{\Gamma_N} g \cdot w da - \int_\Omega \nabla u_D \cdot \nabla w dv \tag{11}$$

and substitute the homogeneous solution $v \in H_D^1(\Omega)$, which is given by $v = u - u_D$ and satisfies $v = 0$ on Γ_D. This gives us the weak formulation of the Poisson problem P, which reads: Find $v \in H_D^1(\Omega)$ such that

$$\int_\Omega \nabla v \cdot \nabla w dv = \int_\Omega f . w dv + \int_{\Gamma_N} g . w da - \int_\Omega \nabla u_D \cdot \nabla w dv \tag{12}$$

2.1.3. Galerkin Discretization

In order to solve the Poisson problem numerically, we have to discretize the weak formulation of the Poisson Eq. (12) and restrict the solution space of the numerical solution U to a finite dimensional subspace S of $H^1(\Omega)$. Accordingly, $U_D \in S_D := S \cap H_D^1$ approximates u_D on Γ_D. The discretized problem P_S can then be written as follows: Find $V \in S_D$ such that

$$\int_\Omega \nabla V \cdot \nabla W dv = \int_\Omega f \cdot W dv + \int_{\Gamma_N} g \cdot W da - \int_\Gamma \nabla U_D \cdot \nabla W dv \tag{13}$$

with $W \in S_D$.

If we assume that $(\eta_1, \ldots, \eta_N)$ is a basis of the N-dimensional space S and $S_D := S \cap H_D^1$, an M-dimensional subspace, then we can rewrite Eq. (13) as

$$\int_\Omega \nabla V \cdot \nabla \eta_i \, dv = \int_\Omega f \cdot \eta_i \, dv + \int_{\Gamma_N} g \cdot \eta_i \, da - \int_\Omega \nabla U_D \cdot \nabla \eta_i \, dv \; (\eta_i \in S_D) \tag{14}$$

If we now make a series expansion of V and U_D in terms of η_k,

$$V = \sum_{k=1}^{M} x_k \eta_k \quad (\eta_k \in S_D) \quad \text{and} \quad U_D = \sum_{k=1}^{N} U_k \eta_k \quad (\eta_k \in S) \tag{15}$$

then we obtain

$$\int_\Omega \nabla \sum_k x_k \eta_k \cdot \nabla \eta_i \, dv = \int_\Omega f \cdot \eta_i \, dv + \int_{\Gamma_N} g \cdot \eta_i \, da - \int_\Omega \nabla \sum_{k=1}^{N} U_k \eta_k \cdot \nabla \eta_i \, dv \tag{16}$$

which can be rewritten as

$$\sum_k x_k \int_\Omega \nabla \eta_k \cdot \nabla \eta_i \, dv = \int_\Omega f \cdot \eta_i \, dv + \int_{\Gamma_N} g \cdot \eta_i \, da - \sum_{k=1}^{N} U_k \int_\Omega \nabla \eta_k \cdot \nabla \eta_i \, dv \tag{17}$$

and finally, when simplified to a system of linear equations as

$$Ax = b \tag{18}$$

where the "stiffness matrix" is given by

$$A_{ik} = \int_\Omega \nabla \eta_i \cdot \nabla \eta_k \, dv \tag{19}$$

and the right-hand side by

$$b_i = \int_\Omega f \cdot \eta_i \, dv + \int_{\Gamma_N} g \cdot \eta_i \, da - \sum_{k=1}^{N} U_k \int_\Omega \nabla \eta_k \cdot \nabla \eta_i \, dv \tag{20}$$

The stiffness matrix is sparse, symmetric, and positive definite. Thus, Eq. (18) has exactly one solution, $x \in R^M$, which gives the Galerkin solution

$$U = U_D + V = \sum_{i=1}^{N} U_i \eta_i + \sum_{k=1}^{M} x_k \eta_k \tag{21}$$

2.1.4. Mesh Generation

The finite element method requires the discretization of the spatial domain Ω with "finite elements"—a regular triangulation T. For two-dimensional problems, triangles and rectangles are used, and in three dimensions, tetrahedral (Fig. 1) and hexahedral elements are

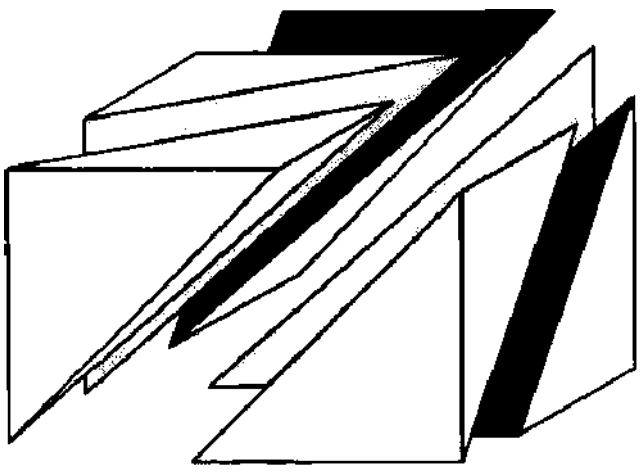


Figure 1. Kuhn triangulation of a cube into six tetrahedral finite elements (exploded view).

commonly used [26]. Also a mixture of different types of elements is possible, but after the evaluation of various other implementations and for simplicity, a tetrahedral discretization has been implemented.

As compared with finite difference methods [27], the finite element mesh may be entirely unstructured, which makes the modeling of complicated geometries and irregular microstructures more convenient, until, a "regular" triangulation of a three-dimensional body of arbitrary shape with tetrahedral elements is required to meet several conditions, defined by Ciarlet [28]:

- The nodes of the mesh lie on the vertices of the tetrahedra.
- The elements of the triangulation do not overlap.
- No node lies on an edge of a tetrahedron.
- Each face on the surface of the body belongs either to Γ_D or Γ_N.

If we assume that the domain Ω has a surface Γ of flat polygons it is possible to find a triangulation T with (tetrahedral) finite elements T, which cover Ω:

$$\Omega = \bigcup_{T \in T} T \tag{22}$$

However, the creation of the geometrical model and its triangulation are still very demanding tasks, which require sophisticated (commercial) tools. PATRAN by MSC Software (http://www.mscsoftware.com/products/) and GID [29] are suitable software packages for this purpose.

As an example, the micrographs of Fig. 2 show a granular thin-film model consisting of 420 randomly oriented, columnar grains with a grain size of 8 nm and a thickness of 15 nm before and after discretization into 15,412 triangular surface elements and 52,884 tetrahedral volume elements.

In the most general case, the maximum cell size of the finite element mesh has to be smaller than the minimum of the three lengths defined [30]. However, if the structure of the solution is roughly known, it is possible to use an "adapted" mesh for the simulations. It has a high resolution (small elements) in areas with very small features (domain walls) and a low resolution (large elements) in other areas where the solution (magnetization) is very uniform. As a result, the number of nodes (and therefore the number of unknowns) is reduced, and the time required for the simulation can be greatly reduced.

For dynamic time-dependent problems a rigid adapted mesh is often not suitable because the solution changes over time and the smallest features of the solution move through the mesh. If they leave the high-resolution mesh and reach areas with larger finite elements, artificial pinning on the mesh will occur and give wrong results. Thus, an adaptive mesh refinement method is required, which changes the structure of the mesh during the simulation and adapts the resolution of the mesh to the solution. It can be shown that these methods lead to (almost) optimal complexity and give most accurate results with the smallest numerical (computational) effort [31]. Adaptive mesh refinement methods are still a very active research area, and they have been successfully applied also in numerical micromagnetics [32–36].

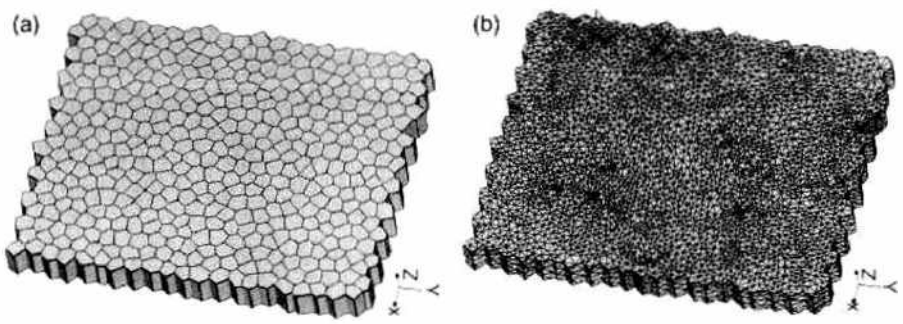


Figure 2. Finite element model of a granular thin film consisting of 420 randomly oriented grains: (a) grain structure; (b) after discretization into finite elements.

2.1.5. Stiffness Matrix and Right-Hand Side

When a regular triangulation T has been generated for the domain Ω, the space S of the numerical solution U has to be defined. A common choice of basis functions for the spline spaces S and S_0 are "hat functions" (Fig. 3), which are defined for every node (x_j, y_j, z_j) of the finite element mesh as

$$\eta_j(x_k, y_k, z_k) = \delta_{jk} \quad (j, k = 1, \ldots, N) \tag{23}$$

We are using isoparametric elements; that is, we use the same polynomials (linear basis functions) for the approximation of the geometry and the solution. If we define

$$S_0 := span\{\eta_j | (x_j, y_j, z_j) \notin \Gamma_D\} \tag{24}$$

then

$$U_D \in S_D$$

and we can calculate the stiffness matrix A_{jk} (Eq. [19]) and the right-hand side b_j (Eq. [20]) as a sum over all elements T and surface triangles E on Γ_N:

$$A_{jk} = \sum_{T \in T} \int_T \nabla \eta_j \cdot \nabla \eta_k \, dv \tag{25}$$

and

$$b_j = \sum_{T \in T} \int_T f \cdot \eta_j \, dv + \sum_{E \in \Gamma_N} \int_E g \cdot \eta_j \, da - \sum_{k=1}^{N} U_k \int_\Omega \nabla \eta_j \cdot \nabla \eta_k \, dv \tag{26}$$

It is most convenient to calculate the stiffness matrix on an element-by-element basis (local or element matrices) and finally assemble the contributions from the local matrices to the global stiffness matrix.

If we assume that the four vertices of a tetrahedral element T are given by (x_j, y_j, z_j) with $j = 1, \ldots, 4$, then the volume $|T|$ of the element is given by

$$|T| = \frac{1}{6} \det \begin{pmatrix} 1 & x_j & y_j & z_j \\ 1 & x_{j+1} & y_{j+1} & z_{j+1} \\ 1 & x_{j+2} & y_{j+2} & z_{j+2} \\ 1 & x_{j+3} & y_{j+3} & z_{j+3} \end{pmatrix} \tag{27}$$

where the local numbering $j = 1, 2, 3, 4$ is chosen in such a way, that the right-hand side of Eq. (23) is positive.

The corresponding basis functions are given by

$$\eta_j(x_k, y_k, z_k) = \delta_{jk}, \quad j, k = 1, \ldots, 4$$

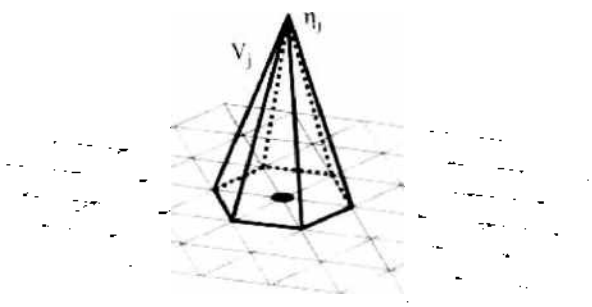


Figure 3. Hat function (linear basis function) for a triangulation in two dimensions.

Thus, η_j can also be written as

$$\eta_j(x, y, z) = \det \begin{pmatrix} 1 & x & y & z \\ 1 & x_{j+1} & y_{j+1} & z_{j+1} \\ 1 & x_{j+2} & y_{j+2} & z_{j+2} \\ 1 & x_{j+3} & y_{j+3} & z_{j+3} \end{pmatrix} \Big/ \det \begin{pmatrix} 1 & x_j & y_j & z_j \\ 1 & x_{j+1} & y_{j+1} & z_{j+1} \\ 1 & x_{j+2} & y_{j+2} & z_{j+2} \\ 1 & x_{j+3} & y_{j+3} & z_{j+3} \end{pmatrix} \tag{28}$$

and

$$\nabla\eta_j(x, y, z) = \frac{1}{6|T|} \cdot \begin{pmatrix} y_{j+2}z_{j+1} - y_{j+3}z_{j+1} - y_{j+1}z_{j+2} + y_{j+3}z_{j+2} + y_{j+1}z_{j+3} - y_{j+2}z_{j+3} \\ -x_{j+2}z_{j+1} + x_{j+3}z_{j+1} + x_{j+1}z_{j+2} - x_{j+3}z_{j+2} - x_{j+1}z_{j+3} + x_{j+2}z_{j+3} \\ x_{j+2}y_{j+1} - x_{j+3}y_{j+1} - x_{j+1}y_{j+2} + x_{j+3}y_{j+2} + x_{j+1}y_{j+3} - x_{j+2}y_{j+3} \end{pmatrix} \tag{29}$$

where all indices are understood modulo 4.

As a result, we can easily calculate the stiffness matrix entries:

$$A_{jk} = \sum_{T \in t} \int_T \nabla\eta_j (\nabla\eta_k)^T \, dv \tag{30}$$

For the right-hand side of Eq. (20), we need to evaluate $\int_T f \cdot \eta_j \, dv$. If we use the value of f in the center of gravity (x_S, y_S, z_S) of T, we can make the approximation

$$\int_T f \cdot \eta_j \, dv \approx \frac{|T|}{4} f(x_S, y_S, z_S) \tag{31}$$

The second term of the right-hand side (Eq. [20]) can be evaluated in a similar way. However, Neumann boundary conditions will not be required in the following.

Finally, the Dirichlet boundary conditions have to be incorporated. One straightforward and easy to implement method is to replace all rows of the stiffness matrix A_{jk}, which correspond to Dirichlet boundary nodes, with zero and a single one in the main diagonal. On the right-hand side, the entries of the Dirichlet nodes are replaced with their boundary values.

2.2. Finite Element Micromagnetics

The total energy of a micromagnetic system is given by the Gibbs free energy E_{tot}, which depends on the magnetic polarization, the external field, and some (temperature-dependent) material parameters. It includes macroscopic contributions, such as the Zeeman energy and the magnetostatic energy, as well as microscopic contributions, such as the magnetocrystalline anisotropy energy and the exchange energy. This highlights the intermediate level of micromagnetics as a continuum theory again, which bridges the gap between the macroscopic world and microstructural and quantum mechanical effects.

The external field is independent of the magnetization distribution and the exchange and anisotropy energy are short-range interactions that depend only on the local magnetization distribution. Thus, they can be computed very efficiently. However, the magnetostatic field is a long-range interaction, which is the most expensive part in terms of memory requirements and computation time. Its calculation is usually based on a magnetic vector [37] or scalar potential (cf. Section 2.2.7). In addition, it is an open boundary problem, for which various methods have been developed [38, 39].

2.2.1. Gibbs Free Energy

The total Gibbs free energy is given by [3, 40]

$$\begin{aligned} E_{\text{tot}} &= \int_\Omega (w_{\text{exch}} + w_{\text{ani}} + w_{\text{ext}} + w_{\text{demag}}) dv \\ &= \int_\Omega (A((\nabla\vec{u}_x)^2 + (\nabla\vec{u}_y)^2 + (\nabla\vec{u}_z)^2) + K_1(1-(\vec{a}\cdot\vec{u})^2) - \vec{J}\cdot\vec{H}_{\text{ext}} - \frac{1}{2}\vec{J}\cdot\vec{H}_{\text{demag}}) dv \end{aligned} \tag{32}$$

where

$$\vec{J}(\vec{x}, t) = J_s(\vec{x}) \cdot \vec{u}(\vec{x}, t), \quad |\vec{u}| = 1 \tag{33}$$

describes the magnetic polarization as a function of space and time, A is the exchange constant, K_1 is the first magnetocrystalline anisotropy constant and $\vec{a}$ the unit vector parallel to the easy axis, H_{ext} the external field, and H_{demag} the demagnetizing field.

In thermodynamic equilibrium, a micromagnetic system tries to reach a state with minimum total energy. The aim of micromagnetic theory is to find the magnetic polarization in equilibrium. Brown proposed a variational method [3], which is based on the calculation of the variational derivative of the total energy with respect to the magnetic polarization. In equilibrium (in an energy minimum), the coefficients of the linear term vanish for any variation $\delta\vec{u}$:

$$\frac{\delta E_{tot}}{\delta \vec{u}} = 0 \tag{34}$$

This leads to Brown's equations

$$\vec{u} \times (2A\Delta\vec{u} + 2K_1\vec{a}(\vec{u} \cdot \vec{a}) + \vec{H}_{ext} + \vec{H}_{demag}) = 0 \tag{35}$$

Thus, in equilibrium the magnetic polarization $\vec{J}$ is parallel to an "effective field"

$$\vec{H}_{eff} = \frac{2A}{J_s}\Delta\vec{u} + \frac{2K_1}{J_s}\vec{a}(\vec{u} \cdot \vec{a}) + \vec{H}_{ext} + \vec{H}_{demag} \tag{36}$$

and the torque that acts on the polarization vanishes:

$$\vec{J} \times \vec{H}_{eff} = 0 \tag{37}$$

Since any contribution parallel to the polarization $\vec{J}$ does not add to the torque, it does not make any difference if the magnetic field $\vec{H}$ or the magnetic induction $\vec{B} = \mu_0\vec{H} + \vec{J}$ is used for the effective field.

2.2.2. Gilbert Equation of Motion

The minimization of Eq. (32) can find an equilibrium magnetization distribution. However, the energy landscape of micromagnetic systems is usually very complicated and contains many local maxima, minima, and saddle points. Therefore, the choice of the initial magnetization distribution has a strong influence on the result. A more physical approach to the problem and a more realistic approach of the system to its equilibrium in a local minimum is provided by a dynamic description of the path through the energy landscape.

The motion of a magnetic moment in a magnetic field is mainly governed by its Larmor precession around the local magnetic field. The damping of the precession causes the relaxation to equilibrium. There are many processes that contribute to the damping in a magnetic solid, such as magnon-magnon and magnon-phonon interactions, interactions between localized and itinerant electrons, and eddy currents [41–43].

The Gilbert equation [24, 44] describes the precession and combines all damping effects in a phenomenological damping term with a single damping constant α:

$$\frac{d\vec{J}}{dt} = -|\gamma|\vec{J} \times \vec{H} + \frac{\alpha}{J_s}\vec{J} \times \frac{d\vec{J}}{dt} \tag{38}$$

where $\gamma = 2.2101735 \times 10^5$ m/As is the gyromagnetic ratio.

This formulation is equivalent to the Landau-Lifshitz-Gilbert (LLG) equation,

$$\frac{d\vec{J}}{dt} = -\gamma'\vec{J} \times \vec{H} - \frac{\alpha\gamma'}{J_s}\vec{J} \times (\vec{J} \times \vec{H}) \tag{39}$$

with

$$\gamma' = \frac{\gamma}{1 + \alpha^2} \tag{40}$$

The intrinsic time scale is determined by the Larmor frequency $\omega = \gamma \vec{H}_{\text{eff}}$, which is usually in the order of gigahertz. Thus, the precession time is smaller than a nanosecond, which requires time steps in the order of picoseconds or even less. This limits the maximum simulated time to about 100 ns.

2.2.3. Discretization

In the following sections, we will discretize the contributions to the total energy with the finite element method as shown in Section 2.1. For static energy minimization methods as well as for the calculation of the effective field Eq. (45), we have to calculate the derivative of the total energy with respect to the local magnetic polarization $\vec{J}$. In the following sections, we will also derive these gradients.

First we have to define the discrete approximation of the magnetic polarization $\vec{J}(\vec{x})$ by

$$\vec{J}(\vec{x}) \approx J_s(\vec{x}) \sum_i \vec{u}_i \eta_i \approx \sum_i J_{s,i} \vec{u}_i \eta_i = \sum_i \vec{J}_i \tag{41}$$

where η_i denotes the basis function (hat function) at node i of the finite element mesh. The material parameters A, K, and J_s are defined element by element and they are assumed to be constant within each element. However, the magnetic polarization that depends on the saturation polarization J_s is defined on the nodes. Thus, we have to introduce the node-based discrete approximation J_s of the saturation polarization $J_s(\vec{x})$ as

$$J_{s,i} = \frac{\int_\Omega J_s(\vec{x}) \eta_i \, dv}{\int_\Omega \eta_i \, dv} = \frac{1}{4V_i} \sum_{T \in T | i \in T} J_{s,i} |T| \tag{42}$$

where V_i denotes the volume that is assigned to node i of the mesh. It is given by

$$V_i = \int_\Omega \eta_i \, dv \frac{1}{4} \sum_{T \in T | i \in T} |T| \tag{43}$$

Since $\vec{J}$ is a vector with three Cartesian components, we have three times the number of nodes unknowns to calculate.

For a given basis η_i, the total energy can be expanded as

$$E_{\text{tot}} = \int_\Omega w_{\text{tot}}(\vec{J}) dv \tag{44}$$

and we get for the effective field using the box scheme [45]

$$\vec{H}_{i,\text{eff}} = -\left(\frac{\delta E_{\text{tot}}}{\delta \vec{J}}\right)_i \approx -\frac{1}{V_i} \frac{\partial E_{\text{tot}}}{\partial \vec{J}_i} = -\frac{1}{V_i J_{s,i}} \frac{\partial E_{\text{tot}}}{\partial \vec{u}_i} \tag{45}$$

2.2.4. Exchange Energy

The exchange energy for one Cartesian component is given by

$$E_{\text{exch}} = \int_\Omega \sum_j A (\nabla u_i \eta_j)^2 dv \tag{46}$$

For the gradient, we obtain

$$\frac{\partial E_{\text{exch}}}{\partial u_i} = \int_\Omega \sum_j A \frac{\partial}{\partial u_i} (\nabla u_i \eta_j)^2 dv \tag{47}$$

$$\begin{aligned} \frac{\partial}{\partial u_i} (\nabla u_i \eta_j)^2 &= 2 u_i \nabla \eta_i \cdot \frac{\partial u_i \nabla \eta_i}{\partial u_i} \\ &= 2 u_i \nabla \eta_i \cdot \nabla \eta_j \delta_{ij} \\ &= 2 u_i \nabla \eta_j \cdot \nabla \eta_i \end{aligned} \tag{48}$$

Finally, the gradient of the exchange energy is given by

$$\frac{\partial E_{\text{exch}}}{\partial u_i} = 2A \int_\Omega \sum_j u_j \nabla \eta_j \cdot \nabla \eta_i Dv \tag{49}$$

which can be written as a linear system of equations with the coefficient matrix

$$G_{\text{exch},ij} = 2A \int_\Omega \nabla \eta_j \cdot \nabla \eta_i dv \tag{50}$$

The gradient can then be simply calculated as

$$\vec{g}_{\text{exch}} = G_{\text{exch}} \cdot \vec{u} \tag{51}$$

The expressions for the x, y, and z component are identical, and there are no mixed terms. This exchange energy matrix is proportional to the stiffness matrix of the Laplacian operator Eq. (19), which is also obvious from Brown's equation Eq. (35) and the effective field Eq. (36).

2.2.5. Magnetocrystalline Anisotropy Energy

The magnetocrystalline anisotropy energy for uniaxial anisotropy is given by

$$E_{\text{ani}} = \int_\Omega \sum_j K_1 (1 - (\vec{a} \cdot \vec{u}_j \eta_j)^2) dv \tag{52}$$

The gradient is given by

$$\frac{\partial E_{\text{ani}}}{\partial u_{i,l}} = \int_\Omega \sum_j K_1 \frac{\partial}{\partial u_{i,l}} \left(1 - \left(\sum_k^{\{x,y,z\}} (a_k \cdot u_{j,k} \eta_j) \right)^2 \right) dv \tag{53}$$

$$\begin{aligned} \frac{\partial}{\partial u_{i,l}} \left(\sum_k^{\{x,y,z\}} (a_k \cdot u_{j,k} \eta_j) \right)^2 &= 2 \sum_k^{\{x,y,z\}} (a_k \cdot u_{j,k} \eta_j) \cdot \sum_m^{\{x,y,z\}} (a_m \delta_{ij} \delta_{lm} \eta_j) \\ &= 2 \sum_k^{\{x,y,z\}} (a_k \cdot u_{i,k} \eta_i) \cdot a_l \eta_i \end{aligned} \tag{54}$$

and we get the result

$$\frac{\partial E_{\text{ani}}}{\partial u_{i,l}} = -2K_1 a_l \int_\Omega \sum_j \sum_k^{\{x,y,z\}} a_k u_{i,k} \eta_j \cdot \eta_i \, dv \tag{55}$$

This can be rewritten in matrix notation as

$$\vec{g}_{\text{ani}} = G_{\text{ani}} \cdot \vec{u} \tag{56}$$

with

$$G_{\text{ani},ij} = -2K_1 a_l \int_\Omega \sum_k^{\{x,y,z\}} a_k \eta_j \cdot \eta_i \, dv \tag{57}$$

2.2.6. Zeeman Energy

The Zeeman energy of a magnetic body $\vec{J}(\vec{x})$ in an external field $\vec{H}_{\text{ext}}(\vec{x})$ is simply given by

$$E_{\text{ext}} = \int_\Omega w_{\text{ext}}(\vec{J}) \, dv = \int_\Omega (-\vec{J} \cdot \vec{H}_{\text{ext}}) dv = \int_\Omega \sum_i J_{i,l} \sum_k^{\{x,y,z\}} u_{i,k} \eta_i H_{\text{ext},k} \, dv \tag{58}$$

For the gradient, we find

$$\begin{aligned}\frac{\partial E_{\text{ext}}}{\partial u_{i,l}} &= -\int_{\Omega} \sum_j J_{\text{s},j} \frac{\partial}{\partial u_{i,l}} \sum_k^{\{x,y,z\}} u_{j,k}\eta_j H_{\text{ext},k}\, dv \\ &= \int_{\Omega} \sum_j -J_{\text{s},j} \sum_k^{\{x,y,z\}} \delta_{ij}\delta_{lk}\eta_j H_{\text{ext},k}\, dv \\ &= -J_{\text{s},i} \int_{\Omega} \eta_i H_{\text{ext},l}\, dv \end{aligned} \tag{59}$$

Since we know the external field explicitly, we can just simply add it to the other contributions to the effective field.

2.2.7. Demagnetizing Field and Magnetostatic Energy

The demagnetizing field is a little more complicated to handle because it is an "open boundary problem" with one of its boundary conditions at infinity. To overcome this problem, Fredkin and Koehler [21, 46, 47] proposed a hybrid finite element/boundary element method, which requires no finite elements outside the magnetic domain Ω.

Since we assume no free currents in our system, we can calculate the demagnetizing field using a magnetic scalar potential $\phi(\vec{x})$. It has to satisfy

$$\begin{aligned}\Delta\varphi &= \nabla \cdot \vec{J}(\vec{x}) \quad \text{for } \vec{x} \in \Omega \\ \Delta\varphi &= 0 \quad \text{for } \vec{x} \notin \Omega\end{aligned} \tag{60}$$

with the boundary conditions at the boundary Γ of Ω,

$$\text{Div}\varphi = 0 \tag{61}$$

and

$$\text{Div}\frac{\partial\varphi}{\partial\vec{n}} = -\vec{n}\cdot\vec{J} \tag{62}$$

In addition, it is required that $\varphi \to 0$ for $|\vec{x}| \to \infty$. The weak formulation of $\nabla \cdot \vec{J}$ is simply given by

$$\int_{\Omega} \nabla\cdot\vec{J}\, dv = \int_{\Omega} \sum_j \sum_k^{\{x,y,z\}} \nabla_k u_{j,k}\eta_j = \int_{\Omega} \sum_j \sum_k^{\{x,y,z\}} u_{j,k}\nabla_k\eta_j \tag{63}$$

which can again be written in matrix-vector format as

$$\vec{d} = D\cdot\vec{u} \tag{64}$$

with

$$D_{i,3j-k} = \int_{\Omega} u_{j,k}\nabla_k\eta_j \tag{65}$$

where k stands for the three Cartesian components $\{x, y, z\}$.

The main idea now is to split the magnetic scalar potential φ into φ_1 and φ_2. Then the problem can be reformulated for these potentials as

$$\Delta\varphi_1 = \nabla\cdot\vec{J} \tag{66}$$

with the boundary condition

$$\frac{\partial\varphi_1}{\partial\vec{n}} = \vec{n}\cdot\vec{J} \tag{67}$$

In addition $\varphi_1(\vec{x}) = 0$ for $\vec{x} \notin \Omega$.

As a result, we find for φ_2

$$\Delta\varphi_2 = 0 \tag{68}$$

with

$$\mathrm{Div}\varphi_2 = \varphi_1 \tag{69}$$

and

$$\mathrm{Div}\frac{\partial\varphi_2}{\partial\vec{n}} = 0 \tag{70}$$

It is required that $\varphi_2 \to 0$ for $|\vec{x}| \to \infty$.

Potential theory tells us that

$$\varphi_2(\vec{x}) = \frac{1}{4\pi}\int_\Gamma \varphi_1(\vec{y})\frac{\partial G(\vec{x},\vec{y})}{\partial\vec{n}(\vec{y})}da \tag{71}$$

where $G(\vec{x},\vec{y}) = 1/|\vec{x}-\vec{y}|$ is the Green function, and φ_1 is easily be calculated using the standard finite element method.

The (numerically expensive) evaluation of Eq. (71) in all Ω can be avoided by just calculating the boundary values of φ_2 on Γ and then solving the Dirichlet problem Eq. (68) with the given boundary values. For $\bar{x} \to \Gamma$, Eq. (71) is given by

$$\varphi_2(\vec{x}) = \frac{1}{4\pi}\int_\Gamma \varphi_1(\vec{y})\frac{\partial G(\vec{x},\vec{y})}{\partial\vec{n}(\vec{y})}da + (S(\vec{x})4\pi - 1)\varphi_1(\vec{x}) \tag{72}$$

where $S(\vec{x})$ denotes the solid angle subtended by Γ at $\vec{x}$. Upon triangulation of the surface Γ of the domain Ω with triangular elements (which we naturally get from a triangulation of Ω with tetrahedral elements) and discretization of φ_1 and φ_2, we can rewrite Eq. (72) as

$$\varphi_2 = B\varphi_1 \tag{73}$$

with the boundary matrix B, which is a dense matrix with a size of $n_b \times n_b$ elements, where n_b is the number of nodes on the surface Γ.

The discretization of the scalar double-layer operator in Eq. (72) has been derived by Lindholm [48]:

$$\int_\Gamma \varphi_1(\vec{y})\frac{\partial G(\vec{x},\vec{y})}{\partial\vec{n}(\vec{y})}da \approx \sum_{t\in\Gamma}\sum_{i=1}^{3} L_{t,i}\varphi_{1,t,i} \tag{74}$$

where t runs over all triangles on the surface Γ of the domain Ω and i runs over the three nodes of each triangle.

In order to calculate the matrix entries of B element by element (rather triangle by triangle) we use the local coordinates defined in Fig. 4 [48].

$$L_{t,i} = \frac{s_{i+1}}{8\pi|t|}\left(\eta_{i+1}S_i - \zeta\sum_{j=1}^{3}\gamma_{ij}P_j\right) \tag{75}$$

$$\vec{\rho}_i = \vec{r}_i - \vec{r} \tag{76}$$

$$s_i = |\vec{\rho}_{i+1} - \vec{\rho}_i| \tag{77}$$

$$\eta_i = \hat{\eta}_i \cdot \vec{\rho}_i \tag{78}$$

$$\zeta = \hat{\xi}\cdot\vec{\rho}_i \tag{79}$$

$$\gamma_{ij} = \hat{\xi}_{i+1}\cdot\hat{\xi}_j \tag{80}$$

$$P_i = \ln\frac{\rho_i + \rho_{i+1} + s_i}{\rho_i + \rho_{i+1} - s_i} \tag{81}$$

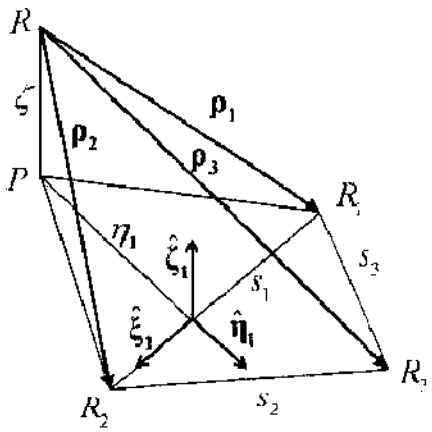


Figure 4. Local coordinate system and various vectors required for the discretization of the boundary integral Eq. (72).

$|t|$denotes the area of triangle t and S_i the solid angle subtended by triangle t at the "observation point" $\vec{r}$, which is given by

$$S_i = 2 \cdot sgn(\zeta) \cdot \arccos\left(\frac{\rho_1\rho_2\rho_3 + \rho_1\vec{\rho}_2 \cdot \vec{\rho}_3 + \rho_2\vec{\rho}_3 \cdot \vec{\rho}_1 + \rho_3\vec{\rho}_1 \cdot \vec{\rho}_2}{\sqrt{2(\rho_2\rho_3 + \vec{\rho}_2 \cdot \vec{\rho}_3)(\rho_3\rho_1 + \vec{\rho}_3 \cdot \vec{\rho}_1)(\rho_1\rho_2 + \vec{\rho}_1 \cdot \vec{\rho}_2)}} \right) \tag{82}$$

To calculate the demagnetizing field, we have to perform the following steps:

Initialization

- Discretize Eq. (66).
- Calculate the boundary matrix in Eq. (73).

Solution

- Solve Eq. (66) for a given magnetization distribution $\vec{J}$ using the standard finite element method.
- Calculate φ_2 on the boundary Γ using Eq. (73) to get the values for the Dirichlet boundary conditions.
- Calculate φ_2 in the whole domain Ω using Eq. (68) with Dirichlet boundary values.
- Calculate $\vec{H}_{\text{demag}} = -\nabla(\varphi_1 + \varphi_2)$.

2.2.8. Effective Field

Finally, we can collect all contributions to the effective field and calculate it by simple matrix-vector multiplications.

$$\vec{H}_{i,\text{eff}} \approx -\frac{1}{V_i J_{s,i}} \frac{\partial E_{\text{tot}}}{\partial \vec{u}_i} = \left(\vec{H}_{\text{ext}} + \vec{H}_{\text{demag}}\right)_i - \frac{1}{V_i J_{s,i}} \left((G_{\text{exch}} + G_{\text{ani}}) \cdot \vec{u}\right)_i \tag{83}$$

Since the matrices for the exchange and anisotropy energies depend only on the (time-independent) material parameters and the geometry, they need to be calculated only once at the beginning. In order to save time and memory, they can also be assembled into a single combined matrix, if the energies and fields are not required separately.

2.3. Solution of the Micromagnetic Equations

2.3.1. Energy Minimization

In Section 2.2, we introduced the Gibbs free energy of a micromagnetic system, discretized various contributions to the total energy, and derived the matrix-vector formulation. Some complication was introduced by the demagnetizing field, but the hybrid finite element/boundary element method provides an elegant way to solve the problem accurately with a finite element mesh, which is restricted to the magnetic bodies. This enables us to

implement a simple energy minimization scheme to find the equilibrium magnetization distribution. If the magnetic polarization $\vec{J}$ is defined in Cartesian coordinates,

$$\vec{J}(\vec{x}) \approx \sum_i J_{s,i} \vec{u}_i \eta_i = \sum_i J_{s,i} \begin{pmatrix} u_{i,x} \\ u_{i,y} \\ u_{i,z} \end{pmatrix} \eta_i \tag{84}$$

one has to use a constrained solver, which ensures that the norm of $\vec{J}$ is preserved: $|\vec{J}| = J_s$. However, it has been shown [49, 50] that the use of spherical coordinates has several advantages:

- The number of unknowns is reduced by one-third, which speeds up the solver.
- The norm is automatically preserved, which allows the use of an unconstrained solver.
- Convergence problems with a Lagrange multiplier approach for the constraint $|\vec{J}| = J_s$ are avoided.

However, the direct calculation of the energy gradient in spherical coordinates causes various problems because of the periodicity of the polar and azimuth angles. Therefore, the magnetic polarization for the minimizer is given in spherical coordinates ($\theta \in [0; \pi[$, $\varphi \in [0; 2\pi[$):

$$\begin{pmatrix} u_{i,x} \\ u_{i,y} \\ u_{i,z} \end{pmatrix} = \begin{pmatrix} \sin\theta_i \cos\varphi_i \\ \sin\theta_i \sin\varphi_i \\ \cos\theta_i \end{pmatrix}, \quad \begin{pmatrix} \theta_i \\ \varphi_i \end{pmatrix} = \begin{pmatrix} \arccos(u_{i,z}) \\ \arctan(u_{i,y}/u_{i,x}) \end{pmatrix} \tag{85}$$

Then it is converted to Cartesian coordinates. The energy gradient is calculated in Cartesian coordinates, converted back to spherical coordinates

$$\begin{aligned} \frac{\partial E}{\partial \theta_i} &= \frac{\partial E}{\partial u_{i,x}} \frac{\partial u_{i,x}}{\partial \theta_i} + \frac{\partial E}{\partial u_{i,y}} \frac{\partial u_{i,y}}{\partial \theta_i} + \frac{\partial E}{\partial u_{i,z}} \frac{\partial u_{i,z}}{\partial \theta_i} \\ &= \frac{\partial E}{\partial u_{i,x}} \cos\theta_i \cos\varphi_i + \frac{\partial E}{\partial u_{i,y}} \cos\theta_i \sin\varphi_i - \frac{\partial E}{\partial u_{i,z}} \sin\theta_i \end{aligned} \tag{86}$$

$$\begin{aligned} \frac{\partial E}{\partial \varphi_i} &= \frac{\partial E}{\partial u_{i,x}} \frac{\partial u_{i,x}}{\partial \varphi_i} + \frac{\partial E}{\partial u_{i,y}} \frac{\partial u_{i,y}}{\partial \varphi_i} + \frac{\partial E}{\partial u_{i,z}} \frac{\partial u_{i,z}}{\partial \varphi_i} \\ &= \frac{\partial E}{\partial u_{i,x}} (-\sin\theta_i \sin\varphi_i) + \frac{\partial E}{\partial u_{i,y}} \sin\theta_i \cos\varphi_i \end{aligned} \tag{87}$$

and returned to the minimizer.

For the minimizer itself, the limited memory variable metric (LMVM) algorithm—a quasi-Newton method—of the TAO package [46, 47, 51, 52] has been selected because it requires only the function values and the gradient of the total energy. In replacement for the Hessian (which is not available due to the demagnetizing field), the second-order information is approximated by a limited history of previous points and gradients. A similar method has been used in [49] and showed better convergence rates than Newton or Gauß-Seidel methods.

2.3.2. The Dynamic Equation

The Landau-Lifshitz-Gilbert equation (39) is a system of ordinary differential equations (ODEs) that can be written in a general form as

$$\frac{d\vec{y}}{dt} = \vec{f}(t, \vec{y}), \quad \vec{y} \in R^N \tag{88}$$

with the initial condition

$$\vec{y}(t_0) = \vec{y}_0 \tag{89}$$

The PVODE package [53, 54] is a general purpose solver for initial-value problems for stiff and non stiff ODEs of the form of Eq. (88). It is based on CVODE [55, 56] and uses MPI for parallelization and portability. Two methods are available for the numerical solution of Eq. (88): The backward differentiation formula (BDF), which is recommended for stiff problems, and the Adams-Moulton formula for nonstiff problems, both of which feature a variable step size and variable order. Both formulas can be written as the linear multistep formula

$$\sum_{i=0}^{K_1} \alpha_{n,i} y_{n-i} + h_n \sum_{i=0}^{K_2} \beta_{n,i} \frac{dy_{n-i}}{dt} = 0 \tag{90}$$

The step size is $h_n = t_n - t_{n-1}$, and the order is q. The Adams-Moulton formula is obtained with $K_1 = 1$ and $K_2 = q - 1$ with $1 \leq q \leq 12$. The BDF formula is represented by Eq. (90) with $K_1 = q$ and $K_2 = 0$ with $1 \leq q \leq 5$. The numerical integration is started with $q = 1$ and then varied automatically and dynamically.

If we insert Eq. (88) in Eq. (90), we get an implicit nonlinear system of equations for $\vec{y}$

$$\vec{G}(\vec{y}_n) := \vec{y}_n - h_n \beta_{n,0} \vec{f}(t_n, \vec{y}_n) - \vec{a}_n = 0 \tag{91}$$

which has to be solved at each time step. $\beta_{n,0}$ and $\vec{a}_n$ depend on the method, the integration order, and the previous time steps. An efficient method for nonstiff problems is functional iteration because it does not require the solution of a linear system of equations. However, for stiff problems it is better solved by Newton iteration, which does involve the solution of a linear system of equations. PVODE uses a Krylov subspace method—the iterative scaled preconditioned generalized minimal residual method (SPGMR) [57], whose performance can be considerably improved with suitable preconditioners.

2.3.3. Preconditioning

Preconditioning of the system of linear equations involved in the Newton iteration of the Krylov subspace method can considerably speed up its solution [58]. In addition, this method leads to fewer function evaluations of the Landau-Lifshitz-Gilbert equation and allows larger time steps, which gives an excellent performance of the numerical time integration.

In order to find the root of Eq. (91), the Newton method requires the calculation of the intermediate corrections $\Delta\vec{y} = \vec{y}_m - \vec{y}_{m-1}$, which follow from

$$\frac{\partial \vec{G}}{\partial \vec{y}} \Delta\vec{y} = -\vec{G}(\vec{y}_{m-1}) \tag{92}$$

The matrix $\partial\vec{G}/\partial\vec{y}$ in this linear system of equations is approximated by

$$\frac{\partial \vec{G}}{\partial \vec{y}} \approx I - h_n \beta_{n,0} \frac{\partial \vec{f}}{\partial \vec{y}} \tag{93}$$

The calculation of the Jacobian of $\vec{f}$

$$\frac{\partial \vec{f}}{\partial \vec{y}} = \frac{\partial}{\partial \vec{J}} \left(-\gamma' \vec{J} \times \vec{H} - \frac{\alpha\gamma'}{J_s} \vec{J} \times (\vec{J} \times \vec{H}) \right) \tag{94}$$

requires the calculation of the Jacobian of the total energy with respect to the magnetization.

We have derived the expressions for the gradient of the total energy in Section 2.2.3. Since the total energy is a simple sum of exchange, magnetocrystalline anisotropy, Zeeman, and magnetostatic energy, we have calculated their gradients individually. For the first two contributions, we found that the gradient of the energy is a linear function of magnetization and ended up with a matrix-vector formulation. The external field is explicitly given anyway, but

the magnetostatic field had to be calculated with a hybrid finite element/boundary element method.

Now we can analyze their contributions to the Jacobian of the total energy. The external field does not contribute at all because it is independent of the magnetization and its first derivative with respect to the magnetization is zero. The first derivative of the demagnetizing field would contribute. However, it is not considered for the calculation of the Jacobian for two reasons. First, its calculation would be very expensive in terms of computational effort, and because of its long-range nature, it would lead to a full matrix for the Jacobian. This results in huge memory requirements and a lot of communication between the processors in a parallel program. Moreover, we do not need the exact Jacobian, but a sensible approximation, which still speeds up the Newton iterations. Thus, it is sensible to consider only the contributions from the exchange and magnetocrystalline anisotropy energy.

The calculation of the Jacobian of these two energy terms is finally very easy. We have already calculated their gradient with respect to the magnetization in order to calculate their contributions to the local field. We found the energy gradients to be linear with respect to the magnetization and came up with a matrix-vector formulation. Because of this linearity, their Jacobians are just simply given by these matrices Eq. (49) and Eq. (57), and we just have to add them up to get the approximate Jacobian for the total energy.

Finally, instead of calculating Eq. (92) with Eq. (93) directly, the preconditioning technique is applied [59]. The linear system

$$A\vec{x} = \vec{b} \tag{95}$$

is rewritten as

$$(AP^{-1})(P\vec{x}) = \vec{b} \tag{96}$$

and

$$A'\vec{x}' = \vec{b} \tag{97}$$

with $A' = AP^{-1}$ and $\vec{x}' = P\vec{x}$. If P is a good approximation to A, then A' is close to the identity matrix and Eq. (4.16) can be solved very efficiently.

2.3.4. The Nudged Elastic Band Method

To find possible paths of a micromagnetic system through its energy landscape to a local minimum of the total energy, we have implemented the static energy minimization method (cf. Section 2.3.1) and the time integration of the Landau-Lifshitz-Gilbert equation (cf. Section 2.3.2). We have applied it to nucleation of reversed domains and domain wall pinning problems as well as investigations of dynamic magnetization reversal processes. However, the investigation of thermal stability, which is an important topic especially in the area of magnetic storage devices, requires the calculation of transition rates between stable equilibrium states of the system. The transition rate between two stable equilibria is determined by the lowest energy barrier (saddle point), which separates them. Henkelman and Jónsson proposed the nudged elastic-band method to calculate these minimum energy paths [60]. This method has been successfully applied to complex micromagnetic systems [61, 62] and it is especially suitable for parallelization.

The path is represented by a sequence of "images" (magnetization distributions) that connect the two given stable equilibrium states $\vec{M}_i$ and $\vec{M}_f$. These equilibria may be obtained using the static energy minimization method, for example. The initial path is given by the initial magnetization distribution $\vec{M}_i$ and the final magnetization distribution $\vec{M}_f$ and a number of images $\vec{M}_k$ in between, which can be obtained by simple linear interpolation. Then, an optimization algorithm is applied, which moves the "elastic band of images" through the energy landscape toward the optimal path, which is defined by

$$(\nabla E(\vec{M}_k) \cdot \vec{t})\vec{t} = \nabla E(\vec{M}_k) \tag{98}$$

where $\nabla E(\vec{M}_k)$ denotes the gradient of the total energy at image $\vec{M}_k$ and $\vec{t}$ is the unit tangent vector along the path. In order to avoid kinks in the path the tangent is calculated using forward, backward, or second-order central differences. Equation (98) requires the component of the gradient parallel to the tangent to be equal to the gradient. In other words, the optimal path is characterized by the fact that the gradient of the total energy is parallel to the tangent for any image $\vec{M}_k$.

Starting from the initial path an iterative optimization scheme is applied that moves the images $\vec{M}_k$ in a direction $\vec{D}(\vec{M}_k)$, which is given by

$$\vec{D}(\vec{M}_k) = -(\nabla E(\vec{M}_k) - (\nabla E(\vec{M}_k) \cdot \vec{t})\vec{t}) \tag{99}$$

The interpretation of this expression is obvious: The images are moved along the negative gradient of the total energy perpendicular to the tangent. The negative gradient determines the direction toward lower energy, while the distance between the images is preserved by taking only the component perpendicular to the tangent. Thus, an ordinary differential equation can be formulated

$$\frac{\partial \vec{M}_k}{\partial t} = \vec{D}(\vec{M}_k) \tag{100}$$

where t denotes some artificial time parameter, which is integrated using any ODE solver.

After the discussion of the static energy minimization method and the time integration, the Landau-Lifshitz-Gilbert equation, the implementation of the nudged elastic-band method has become very simple, because we can reuse parts of both of them. The static energy minimization provides us with the gradient of the total energy, and for the integration of Eq. (100), we can use the same methods explained in Section 2.3.2 by just replacing the right-hand side of Eq. (88) with Eq. (99).

The parallelization of this method can be done by distributing the images across the processors. Thus, every processors needs the full set of matrices (but only one copy independent of the number of images), which are required for the calculation of the local fields (and gradients). However, there is no need to partition the finite element mesh anymore because every processor has to do the full calculation for the images. This has the advantage that no communication is required during the calculation of the gradient of the total energy. Only for the calculation of the tangents the magnetization of some images has to be copied to "neighboring" processors.

3. MAGNETIZATION REVERSAL IN NANOSTRUCTURED MAGNETIC MATERIALS

Small magnetic structures play a crucial role in modern storage devices and sensors. Nanotechnology and nanoscience in modern magnetics also lead to an increased demand for the better understanding of the magnetic reversal or switching processes of magnetic nanoparticles, thin films, and even bulk magnetic materials with nanosized precipitates and intergranular regions. In the following, typical examples are shown, where numerical, three-dimensional micromagnetic simulations of granular materials have revealed magnetic hysteresis and switching properties in good agreement with experimental findings. Such examples are schematically shown in Fig. 5, where various types of nanostructured materials are built by assembling of individual grains in the order of several tens of nanometer (a) to rod-shaped wires, (b) layered-shaped thin films and multilayers, and equiaxed nanocrystalline bulk materials (d) and (e). In reality the simplified microstructures have to be extended to complex grain shapes and considering distributions of grain sizes and material parameters (Fig. 5[e]). For this reason, finite element micromagnetic models are advantageous compared with finite difference models because of the possibility to take into account realistic granular microstructures.

In particular, the role of the physical microstructure on hysteresis properties will be discussed, and numerical techniques to treat microstructural effects will be proposed. Tremendous progress has been made in preparing nanosized magnetic particles and thin films for magnetic recording media. The reduction of grain sizes and film thickness to less

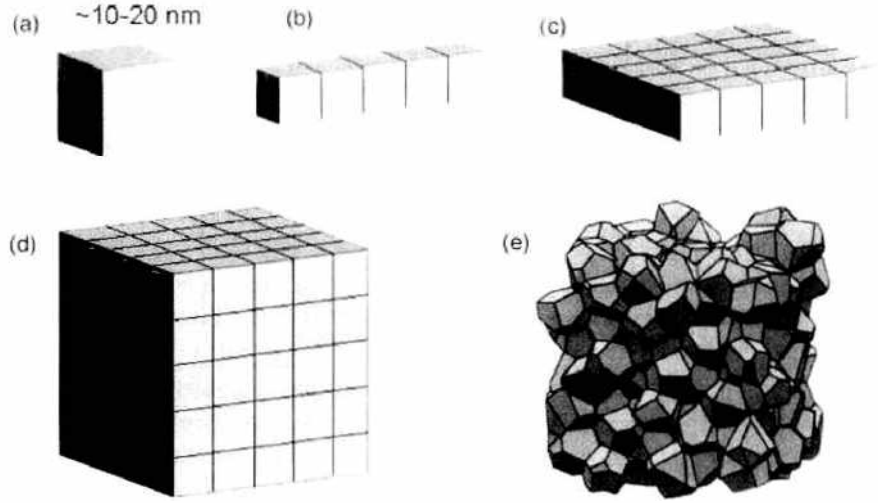


Figure 5. Schematic classification of nanostructured materials by assembling of individual grains in the order of several tens of nanometer (a) to rod-shaped wires (b), to layered-shaped thin films, and multilayers and equiaxed nanocrystalline bulk materials (d) and (e).

than 10 nm lead to the improvement of the magnetic storage data density to values larger than 100 Gbits/in^2 [63]. Co/Pt, FePt, and CoPd multilayers and small particles are of great interest for high-density storage magnetic recording associated with large perpendicular magnetocrystalline anisotropy.

3.1. Particles, Grains

3.1.1. Stoner-Wohlfarth Behavior

The magnetization reversal of a single-domain particle or grain is simply described by the Stoner-Wohlfarth model of magnetization reversal in which the switching field is controlled by the precession of the magnetic polarization vector toward the effective or external field [64, 65]. A rigid magnetic moment is an approximation for a very small magnetic particle with strong exchange interaction. If it is small enough, the exchange interaction will keep the magnetization uniform, which leads to coherent rotation. In this reversal mode, the constituent spins rotate in unison. The exchange interaction gives a constant contribution to the Landau free energy in this approximation and therefore does not influence the motion of the polarization vector. These are the common assumptions in the Stoner-Wohlfarth model of magnetization reversal.

The equilibrium direction of the magnetic moment is determined by the magnetocrystalline anisotropy axis and the direction of the external field. It can be readily obtained by considering the total free energy of the magnetic moment, which is in the case of uniaxial anisotropy given by

$$E_{\text{tot}} = -K_1 V \cos^2(\theta - \phi) - J_s V H_{\text{ext}} \cos\phi \tag{101}$$

K_1 is the first magnetocrystalline anisotropy constant, V the volume of the magnetic particle, J_s its saturation polarization, and H_{ext} the external field. If an external field is applied at an angle θ to the easy axis of the uniaxial anisotropy of the particle, the magnetization vector will reach its equilibrium position at an angle ϕ from the field direction, where the total free energy has a local minimum. Since there is not only one minimum, the equilibrium direction is also influenced by the history of the magnetization vector. This effect is called *hysteresis*.

It should be noted that the minimum switching field that differs from the value of the coercive field is obtained at an angle of 45° (Fig. 6). Another important aspect can be explained using the Stoner-Wohlfarth model—the difference between reversible and irreversible magnetization processes. Reversible processes are those in which the magnetization returns to its initial position after the perturbation has been removed again. For example, when a weak external bias field is applied, the magnetization will change slightly, but upon removal of the bias field, it returns to its initial position. However, if the bias field is strong enough, the

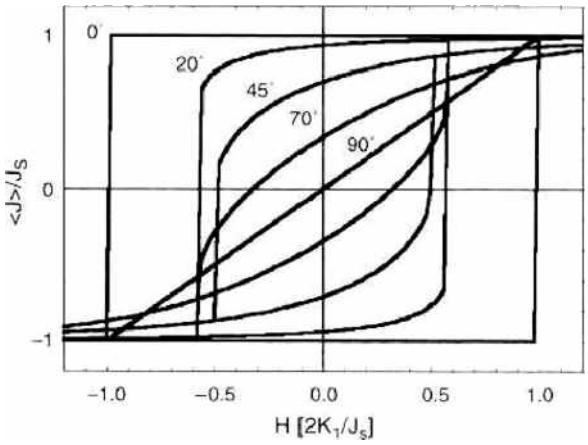


Figure 6. Hysteresis curves of a spherical single domain particle for different angles θ between anisotropy axis and external field. Magnetization reversal is originated by rotational processes only.

magnetization can switch irreversibly and remain in a different energy minimum after the bias field has been switched off. Of course, this effect is used to switch the magnetization between different directions and store information thereby.

The undamped equation of motion ($\alpha = 0$) describes the continuous precession of the polarization vector around the direction of the effective magnetic field.

$$\frac{d\vec{J}}{dt} = -|\gamma|\vec{J} \times \vec{H}_{\text{eff}} \tag{102}$$

where $\gamma = 2.2101735 \times 10^5$ m/As is the gyromagnetic ratio.

However, changes of the magnetization are known from experiments to decay in finite time. Thus, damping is introduced by a phenomenological term. The commonly used Gilbert (38) or Landau-Lifshitz equation of motion (39) results in a spiraling movement of the polarization vector toward its equilibrium direction. The analytical solution [66]

$$\mathbf{J} = J_s \begin{pmatrix} \operatorname{sech}(\alpha\gamma Ht)\cos(\alpha\gamma Ht) \\ \operatorname{sech}(\alpha\gamma Ht)\sin(\alpha\gamma Ht) \\ \tanh(\alpha\gamma Ht) \end{pmatrix} \tag{103}$$

is given as a projection into the x-y plane in Fig. 7 for two different values of the damping parameter. If the damping parameter is rather low ($\alpha = 0.1$), the polarization vector

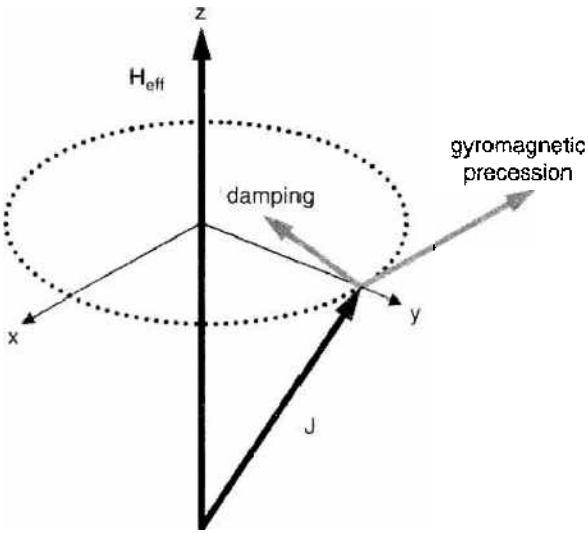


Figure 7. Gyromagnetic precession and damping of the polarization vector during the magnetization reversal toward the effective field.

precesses many times before it reaches its equilibrium direction. If ($\alpha = 1$), the precession is critically damped, and the polarization turns directly into the direction of the effective field (Fig. 8).

Figure 9 shows how the switching time of a single polarization vector depends on the value of the damping parameter. Initially, the magnetization vector points in a direction opposite to the effective field. A small deflection starts the reversal process. We can measure the switching time, which is the time until the component of the polarization parallel to the effective field has reached a certain value (e.g., zero). For $\alpha < 1.0$, the motion of the magnetization vector is undercritically damped. It makes many precessions around the direction of the effective field. On the contrary, for $\alpha > 1.0$ the motion is overcritically damped. The minimum of the switching time is found for $\alpha = 1.0$, which is the case of critical damping [67]. The damping parameters are in the range between 0.01 and 1.0 for magnetic materials that are commonly used in magnetic recording media [66, 68]. These values are obtained by ferromagnetic resonance experiments: A very strong external DC field keeps the magnetization of the sample homogeneous and parallel to its direction. A small AC field perpendicular to the DC field excites a periodic motion of the magnetization with a small amplitude. As the frequency of the AC field is varied, the absorbed energy varies, and at the resonance frequency it reaches a maximum. From the width of the absorption spectrum, the damping constant can be derived.

Finite element micromagnetic simulations of a CoPtCr particle with the materials parameters of $J_s = 0.37$ T, $A = 1 \times 10^{-11}$ J/m, $K_1 = 2 \times 10^5$ J/m^3, and $\alpha = 0.02$ reveal for a particle diameter of 16 nm a homogeneous precession toward the external field, whereas for 100 nm in diameter, the reversal process is dominated by the nucleation of reversed domains and by the expansion of domain walls. Figure 10 compares the transient magnetization states at $H_{ext} = H_c$ for the particle diameters of 16 nm and 100 nm. The calculations were started after saturation. The field was reduced in steps of $\Delta H = 20$ kA/m.

From the simulations, it is obvious that the switching time depends on the value of the damping parameter and how the external field is applied (i.e., field strength, field profile, and sweep rate). Figure 11 shows the time evolution of the polarization during the magnetization reversal of the CoCrPt particle described in Fig. 10. The precessional switching of the small particle ($D = 16$ nm) leads to an oscillation of the total polarization parallel to the external field. The nucleation of domain at corners and the expansion of domain walls lead to a smooth decay of the total polarization and to a smaller switching time t at $\langle J_y \rangle = 0$.

In summary, when a magnetic particle is small enough that any nonuniform distribution of its magnetization becomes energetically unfavorable, its magnetization reversal is expected to take place in a uniform way. This situation was described more than 50 years ago by Stoner and Wohlfarth and Néel [69] in a theory known as the *uniform rotation model*. The Stoner-Wohlfarth model has been extensively used for describing properties of assemblies of particles. Only in the past few years have experiments on individual nanoparticles become possible thanks to the development of new, very sensitive magnetometers [70–74].

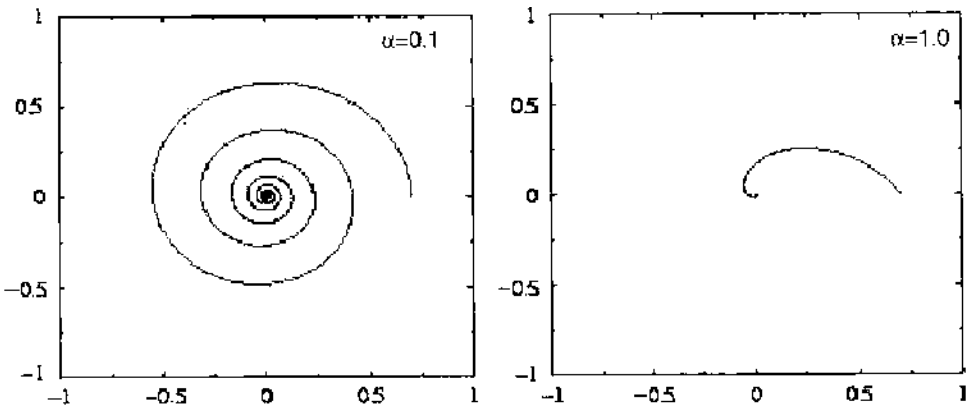


Figure 8. Trajectory of the polarization vector of a single polarization vector in the x-y plane.

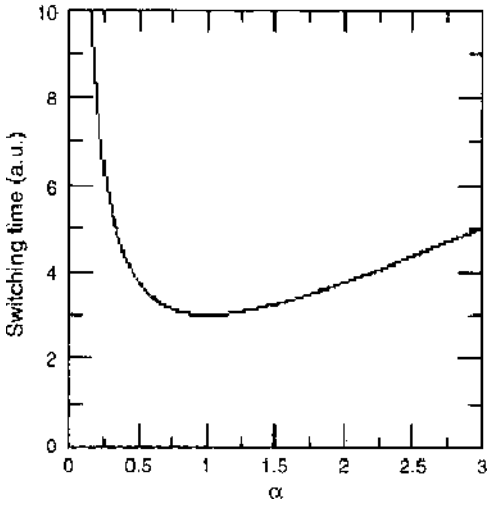


Figure 9. Dependence of switching time of a single polarization vector on the damping constant.

3.1.2. Nucleation Processes in Fe-Pt Particles with Multiple Easy Axes

High-density magnetic storage media require tight control of the grain size, grain size distribution, chemical composition, and microstructure to ensure the thermal stability of the bits and to keep the media noise low. However, as the areal density increases, the grain size and the magnetic switching volume decrease. In order to maintain the stability, materials with higher uniaxial anisotropy than the common CoCrPt alloys are required. FePt thin films and self-assembled nanoparticles are promising candidates for high-density magnetic storage media. Their magnetocrystalline anisotropy is 50–100 times larger than in CoPtCr media alloys that may allow areal densities in the Tbit/in^2 regime [63].

Controlled self-assembly of magnetic nanocrystals has been widely used as fabrication technique to obtain regular deposition of magnetic particles in the form of superlattices with high symmetry [75]. In the case of FePt, the phase transformation from the disordered fcc structure into the ordered $L1_0$ phase occurs at 530 °C. Among the magnetic nanoparticles, FePt or CoPt ordered alloy nanoparticles are taken as promising candidates. These alloys have a $L1_0$-type ordered structure with K_1 as high as 10^6 to 10^7 J/m^3. In the case of FePt nanocrystals with an average size of 12 nm, a coercive field of about 3 kOe has been observed after a heat treatment above 600 °C. Transition electron microscope (TEM) investigations have revealed that the crystallographic three-variant domain structures of the tetragonal $L1_0$-FePt phase coexisted even in small nanoparticles or coalesce to form larger grains after the annealing [76]. Both mechanisms drastically deteriorate the coercive field. The classical theory assuming a coherent magnetization rotation by a reversed external field and developed

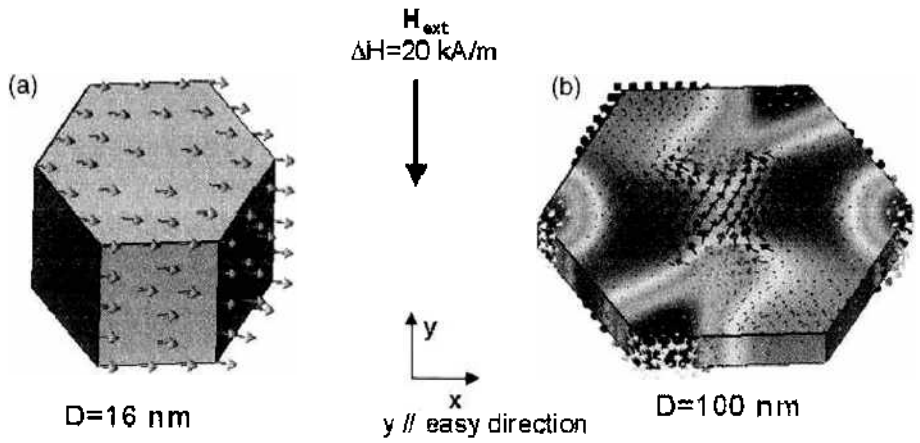


Figure 10. Transient magnetization states at $H_{ext} = H_c$ for the particle diameters D of 16 nm and 100 nm, respectively, obtained by "quasistatic" finite element micromagnetic simulations.

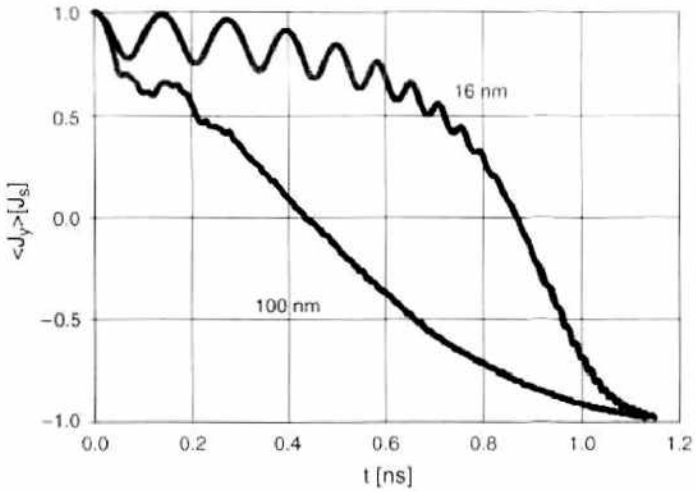


Figure 11. Time evolution of the polarization during switching of a CoCrPt particle with $J_s = 0.37$ T, $A = 1 \times 10^{-11}$ J/m, $K_1 = 2 \times 10^5$ J/m^3, and $\alpha = 0.02$.

by Stoner and Wohlfarth for noninteracting, single-domain particles predicts a coercive field in the order of 10.8 MA/m (135 kOe). This discrepancy between experimental and theoretical coercive field values can be explained by micromagnetic simulations. The basic geometry of the FePt nanocrystal with several regions with different anisotropy regions is shown in Fig. 12. Based on Fig. 12, the micromagnetic model has been developed assuming various regions, or grains, with different easy magnetization directions within one nanoparticle. The hexahedral particle is split into six parts of equal volume. In each part, the magnetocrystalline anisotropy axis is uniform, but the axes in the different parts have been varied. Typical values for the material parameters of FePt ($L1_0$) thin films and nanoparticles have been measured and published in various papers [63, 77].

The influence of a distribution of easy axes within the particle has been studied. The easy axis in the six parts of our model has been varied, and the coercivity has been calculated for an external field applied parallel to the z-axis. The following material parameters [63] have been used for the numerical simulations: $J_s = 1.43$ T, $A = 1.0 \times 10^{-11}$ J/m, $K_1 = 7.7$ MJ/m^3, $\alpha = 0.1$. The exchange length of FePt is about 1.2 nm and therefore the resulting classical domain wall width is about 3.5 nm. The results are summarized in Table 1. The left column indicates, how many of the six parts of our model have their easy axis parallel to the z-, y-, and x-axes, respectively. The results show that the coercivity is decreased by a factor of three as compared with the nucleation field. However, the different distributions of easy

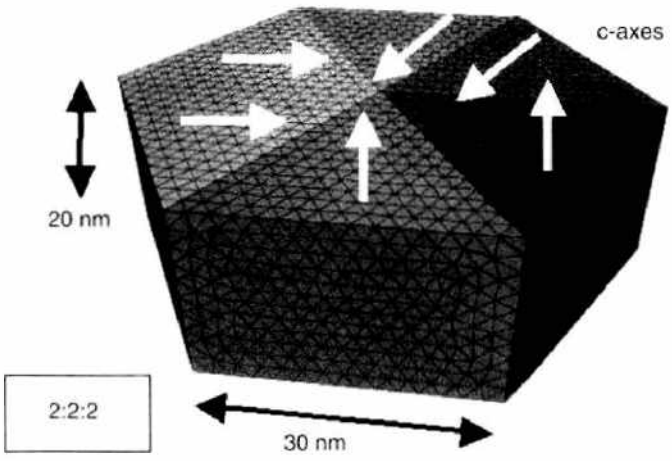


Figure 12. Geometrical model of an FePt nanoparticle used in the micromagnetic simulations. The finite element mesh consists of 10,340 nodes and 55,076 elements, which gives a discretization length of 1.2 nm, if the edge length is 30 nm. The model is split into six parts of equal volume in which the anisotropy axis are varied (2 : 2 : 2 configuration shown).

Table 1. Coercivity as a function of the easy axis distribution. The first, second, and third number of a triplet in the first column indicates in how many of the six parts of the finite element model the easy axes are parallel to the z-, x-, and y-axis, respectively.

$x:y:z$	5 : 1 : 0	4 : 2 : 0	4 : 1 : 1	3 : 3 : 0	3 : 2 : 1	2 : 2 : 2[a]	2 : 2 : 2[b]
H_c [kA/m]	3330	3140	3310	360	3420	3430	2040

[a]Three pairs of neighbouring parts with equal anisotropy axes.
[b]Each pair of neighbouring parts has perpendicular easy axes.

axes do not show any significant influence on the coercivity. This behavior indicates that the 90° domain wall at the interface between two misaligned parts of the particle determines the coercivity. Thus, already a single misaligned part is sufficient to reduce the coercivity by a factor of three. Finally, we have reduced the size of the particles and studied their coercivity. The shape and aspect ratio remained the same; the model has just been rescaled to the desired size. As a result, the properties of very small particles are modified because of the increasing importance of the exchange interactions. The results of our simulations are summarized in Fig. 13. We have used the (2 : 2 : 2(b)) distribution of easy axes, where each pair of neighbouring parts in our model has perpendicular easy axes. For this distribution, we find a further reduced coercivity of 2040 kA/m, which drops to 600 kA/m, if the particle size is reduced to 3.75 nm.

As a result, the properties of very small particles are modified due to the increasing importance of the exchange interactions. The results of the micromagnetic simulations show that the coexistence of crystallographic three-variant domain structures are responsible for the drop in coercive field and that a minimum particle size of about 6 nm is necessary. The influence of the magnetostatic field has been shown to be negligible as compared with the anisotropy field [78] and has been omitted in all simulations, because the nucleation field (i.e., the coercive field of these particles) is reduced by less than 5%.

The magnetic properties of the surface of FePt nanoparticles can be altered by oxidation. Fe_3O_4 and other oxidation states have been found by near edge X-ray absorption fine structure spectroscopy [79], which indicates oxide shells of about 0.4 nm around FePt nanoparticles measuring 4–6 nm in diameter. Therefore, the properties of γ-Fe_3O_4 [80] have been assumed for the surface shell surrounding the FePt core: $K_1 = 0.011$ MJ/m^3, $J_s = 0.5$ T, $A = 13.2$ pJ/m.

In addition to the micromagnetic model of Fig. 12, where the particle has been split into six different parts (according to its hexagonal shape) and in which the orientation of anisotropy axes has been varied, a surface shell of variable thickness surrounding the core of the particle has been defined to introduce surface anisotropy effects. Figure 14 shows an exploded view of the finite element micromagnetic model.

Figure 15 shows the coercive field of a uniaxial particle with s2 surface configuration (cf. Table 2) with in-plane anisotropy in each facet. The external field is applied at an angle of 3° with respect to the anisotropy axis (which is parallel to the sixfold symmetry axis). As the

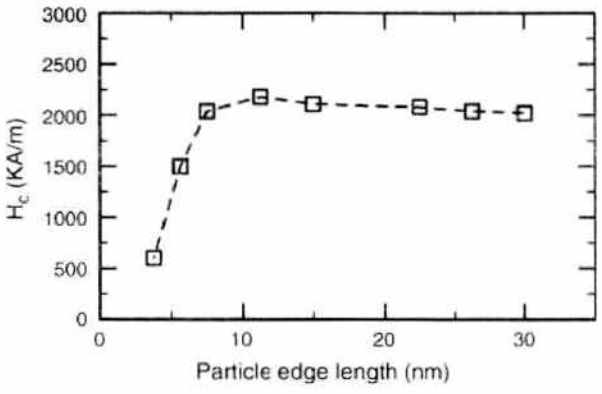


Figure 13. Coercivity as a function of the edge length of the nanoparticle. The easy axis distribution is shown in the top figure for the (2 : 2 : 2(b)) distribution, where all neighboring pairs have perpendicular easy axes.

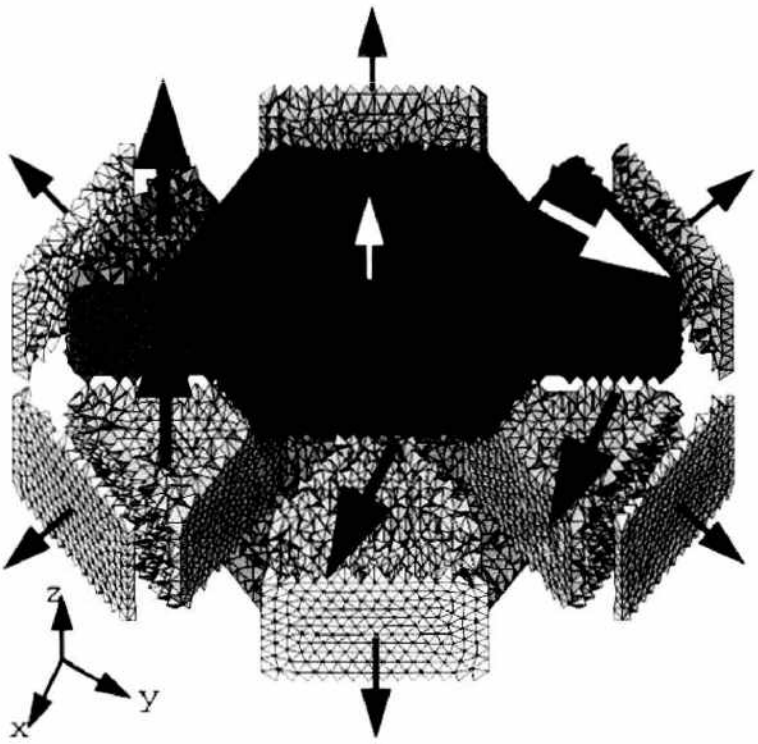


Figure 14. Exploded view of the finite element model of a FePt nanoparticle with hexagonal basis, six inner parts, and eight surface parts. The big arrows indicate the anisotropy axes of the inner parts in the "2 : 2 : 2" configuration, and the smaller arrows indicate those for the surface shell (perpendicular to the surface). The roughness on the interior interfaces is caused by the structure of the finite element mesh.

particle size is reduced, the thickness of the soft oxide shell is scaled down and its influence diminishes, which makes the coercivity converge to the Stoner-Wohlfarth limit. The oxide on the surface makes the nucleation process easier due to its reduced anisotropy. As a result, the coercivity is considerably reduced as compared with the Stoner-Wohlfarth result. For very small particles, the surface shell is very thin, and it cannot support the nucleation process effectively.

As the size of the particle is scaled up, the thickness of its surface shell increases in our model, too. This makes the nucleation process easier and reduces the coercivity. If misaligned anisotropy axes are introduced as shown previously, the coercivity is strongly reduced. This is due to the fact that the misaligned parts create 90° domain walls at their interfaces. Thus, the magnetization reversal mechanism changes from nucleation (for single crystals) to domain

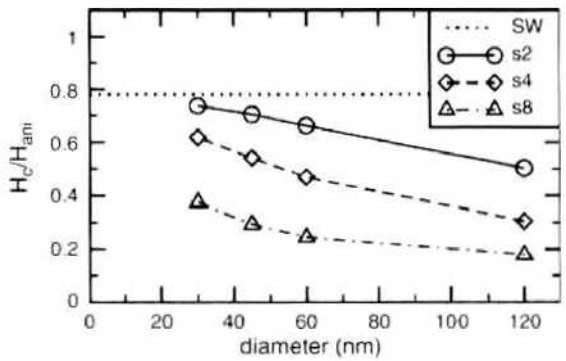


Figure 15. Coercivity (in units of the anisotropy field $H_{ani} = 2K/J_s$) as a function of particle diameter for a particle with a single anisotropy axis and reduced surface anisotropy (γ-Fe_2O_3 material parameters with in-plane anisotropy). The anisotropy axes in all six core parts are parallel to the z-axis; "SW" indicates the switching field of a Stoner-Wohlfarth particle for an external field applied at 3° from the anisotropy axis; "s2"-"s8" stands for the thickness of the surface shell as given in Table 2

Table 2. Thickness t, and volume of the surface shell V_{shell} in comparison with the core volume V_{core} for a particle diameter of 60 nm.

Symbol	t, (nm)	V_{shell} (nm^3)	V_{core} (nm^3)	V_{shell}/V_{core}
s2	1	8000	38,000	0.21
s4	2	16,000	30,000	0.53
s8	4	29,000	17,000	1.71

wall pinning. The oxide shell contributes to this effect and leads to a further reduction as shown in Fig. 16. In single crystals, the oxide shell provides the nucleation site, whereas in the case of multiple easy axes, it supports the domain wall depinning from the interface. This gives rise to the maximum of the coercivity as a function of the particle size.

If we assume strong surface anisotropy for FePt nanoparticles [81], effective anisotropy values as high as $K_{eff} = K_S/t = 46.2$ MJ/m^3 may arise in the surface shell, where t indicates its thickness. This corresponds to a ratio $K_{eff}/K_1 = 6$. As a result, we find a strongly disturbed magnetization distribution for a uniaxial FePt nanoparticle with a diameter of 45 nm and a surface layer thickness of 1.5 nm. The influence of the surface anisotropy on the coercivity of uniaxial FePt nanoparticles is shown in Fig. 17. For small particle sizes (but still large compared with the exchange length), the coercivity approaches the nucleation field because the thickness of the shell is of the order or below the domain wall thickness. As the particle size (and the shell thickness) increases, the coercivity is reduced by 50%. The disorder in the magnetization distribution on the surface of the particle facilitates the nucleation process and reduces the coercivity. For large particles the reduced coercity remains constant because the nucleation process does not depend on the particle size when the thickness of the shell is larger than the domain wall size. For particles with a diameter of the order of a few nanometers, Monte-Carlo simulations have shown this competition between surface and bulk anisotropy, which leads to throttled and "hedgehog" spin structures [81].

Surface oxidation during the processing of FePt nanoparticles can lead to the formation of Fe-oxides with reduced magnetocrystalline anisotropy. The simulations show a reduction of the coercivity up to 75% depending on the particle size and thickness of the oxide shell. A similar effect is observed for "pure" FePt nanoparticles due to intrinsic surface anisotropy. Strong surface anisotropy perpendicular to the surface facilitates the nucleation and magnetization reversal, which leads to a similar reduction of the coercivity.

3.2. Rod-Shaped Co-Nanowires (One-Dimensional Structures)

The knowledge of the switching process of interacting granular Co nanowires is of technical relevance and has been investigated using finite element micromagnetics [82]. Co nanowires are usually grown in highly ordered anodic aluminium templates using electrodeposition. This technique yields completely metal-filled aluminium membranes. The nanowires are arranged in a hexagonal lattice. The numerical results clearly explain the influence of the microstructure on the magnetization reversal process. The nucleation and expansion of

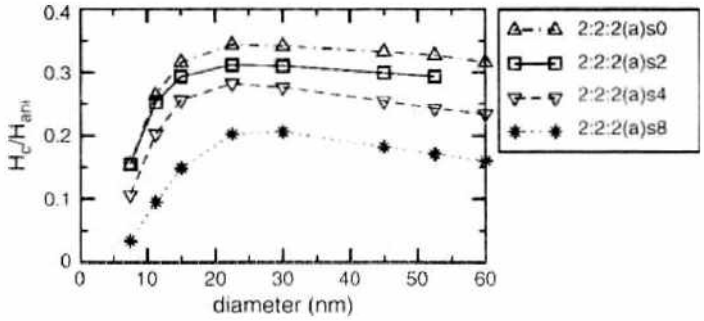


Figure 16. Coercivity as a function of particle diameter for different thicknesses of the surface shell ("s0" gives the results without an oxide shell). A polycrystalline particle with 2 : 2 : 2 configuration of the magnetocrystalline anisotropy axes has been assumed (cf. Fig. 14).

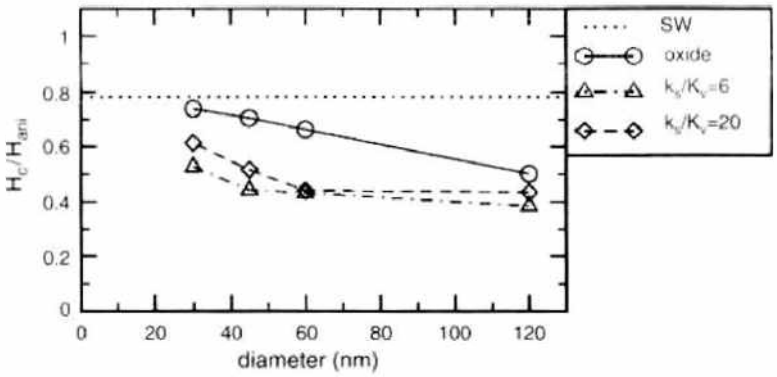


Figure 17. Coercivity as a function of particle diameter for an oxide shell with reduced anisotropy in comparison with strong surface anisotropy. A thin shell (s2) has been assumed.

reversed domains are calculated by solving the Gilbert equation of motion for different damping constants using a moving mesh technique. The studied magnetic Co nanowire has a diameter of about 55 nm and total length up to 1000 nm. TEM investigations show two different types of hcp-structured grains. For one the c-axis is randomly oriented in a plane perpendicular to the long axis of the wire and the other has the c-axis parallel to the long axis. The typical length of these grains is 100–250 nm. Grains of the second type have a length of less than 100 nm and the c-axes are parallel to the long axis. The first kind of grains has a total volume fraction of the Co nanowire of 70–90%. The materials parameters used for the simulations are $J_s = 1.76$ T, $A = 13$ pJ/m, and $K_1 = 4.5 \times 10^5$ J/m^3 [83]. The simulations show that the results are insensitive to a reduction of the exchange coupling between the grains. The nanowire has two stable magnetic states: all magnetic moments are parallel to the long axis, pointing either in one direction or in the other. We apply an external field parallel to the long axis in the opposite direction of the magnetization. If the field is large enough, a reversed domain will be formed, a head-to-head domain wall will be built, and it will propagate through the whole nanowire, until it arrives at the other end. There it is annihilated, leaving the system in the second possible stable state. As a result of the micromagnetic simulations it has been proved that the domain wall velocity and the structure of the domain wall strongly depend on the diameter d of the nanowire. Two types of domain walls can be formed during their motion through the nanowire as shown in Fig. 18 [84]: transverse walls and vortex walls.

As result of the simulations, it has been shown that the domain wall velocity depends on the domain wall structure (Fig. 19). In the first case, the magnetization in the center of the

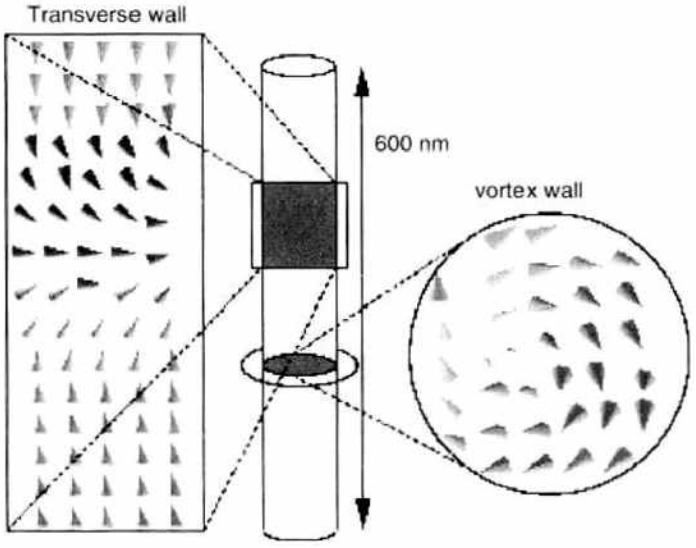


Figure 18. Magnetization distribution in a transverse wall (cut plane parallel to the wire axis) and in a vortex wall (cut plane perpendicular to the wire axis through the vortex core). The wire diameter $d = 20$ nm.

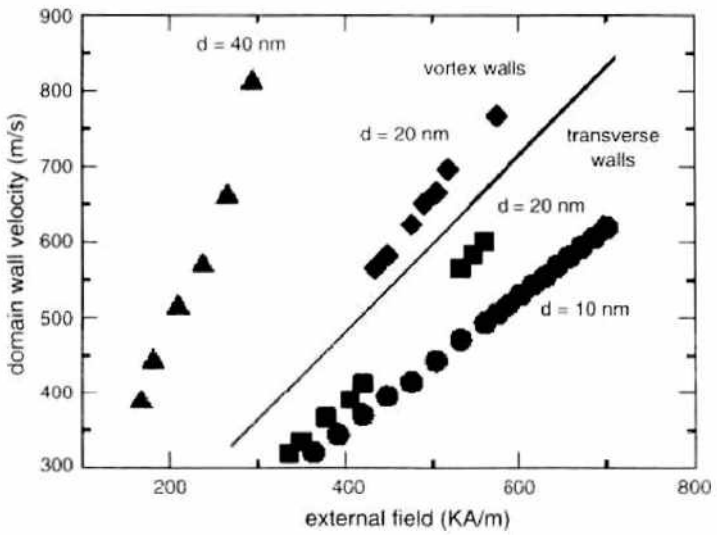


Figure 19. Domain wall velocity for different diameters d of the Co nanowire. The velocity depends on the domain wall structure. The simulations were made for a damping constant of one.

wall points perpendicular to the long axis of the wire. As the wall moves, the transverse component of the magnetization circles around driven by gyromagnetic precession. In the second case, a vortex is formed in a plane normal to the long axis of the wire. The magnetization aligns parallel to the wire surface. Within the domain wall, the magnetization rotates around the long axis that leads to the formation of a Bloch point in the middle of the wire. The formation of the vortex decreases the magnetostatic energy at the expense of the exchange energy. For small diameters, it is energetically preferable to form a transverse wall that minimizes the exchange energy because of the parallel alignment of the magnetization in the domain wall. However, there are no flux closure states in the wire and so the magnetostatic energy is not reduced. However, for small diameters, the exchange energy is the most dominant energy contribution to the total Gibbs free energy, and so its minimization determines the wall configuration. With increasing diameter, the magnetostatic energy becomes more important in the minimization process. Although the vortex wall increases the exchange energy, the minimization of the magnetostatic energy due to the lower demagnetizing field makes it energetically preferable. Since the difference in the total Gibbs free energy is less than 1% at a critical diameter of $d = 20$ nm, it is possible to obtain both domain wall structures. For $d < 20$ nm, only transverse walls are observed, and for $d > 20$ nm only vortex walls are observed.

The dependence on the damping constant is not the same for both structures. For $d = 10$ nm, just transverse walls are formed. With increasing damping constant, the velocity increases from 50 m/s for a damping constant of 0.05 to 520 m/s for a damping constant of 1 at an applied field of 500 kA/m. For higher damping constants, the domain wall velocity increases faster for higher external fields. In the transverse wall, gyromagnetic precession plays a major role during the wall motion. On the contrary, for the vortex walls, the domain wall velocity increases with decreasing damping constant, reaching 2000 m/s for $d = 40$, a damping constant of 0.05 and an applied field of 250 kA/m.

The development of an adaptive mesh refinement algorithm helps to bridge the length scales used in micromagnetic simulations [35]. The finite element mesh is dynamically adapted to the current magnetization distribution. The mesh is dynamically adapted to the current magnetization distribution. Finite elements are introduced into regions where a high resolution is required. Afterward the mesh is coarsened again. Different attempts have been made in micromagnetics. Miltat and Labrune [85] analyzed two-dimensional Néel-type walls using a discretization of the sample into rectangular prisms of variable size having small elements in the center of the wall. Tako et al. [32] introduced *a posteriori* refinement that subdivides triangular elements with large errors into four smaller elements. Hertel and Kronmüller [33] applied an adaptive finite element method to calculate domain configurations and vortex motion in thin-film elements. In the following, the case is described in which

mesh refinement and mesh coarsening is applied in order to obtain a high-density mesh that moves together with the domain wall (Fig. 20).

To solve the partial differential equations, we use the finite element method. The magnetic wire is divided into tetrahedral finite elements. The space discretization leads to a system of coupled magnetic moments sitting at the nodes of the finite element mesh. The effective field at the nodes follows from the derivative of the Gibbs free energy with respect to the magnetic moment. The Gilbert equation of motion has to be solved for each magnetic moment. The equations are coupled by the exchange and magnetostatic interactions between the magnetic moments. In order to resolve a magnetic domain wall, the element size has to be smaller than the characteristic length, given by the minimum of the exchange length parameters (Eqs. [1] and [2]). If the element size is too big a so-called domain wall collapse will occur: the magnetization becomes aligned antiparallel at neighboring nodes and the torque on the magnetization vanishes. A large number of finite elements is required for the study of wall motion in magnetic wires using a uniform fine grid with an element size smaller than the critical exchange length. In order to keep the number of finite elements small and avoid the domain wall collapse [86], an adaptive refinement scheme can be applied.

The outline of adaptive algorithms is as follows. Starting from an initial finite element mesh τ_0, we produce a sequence of refined grids τ_M until the estimated error is below a given tolerance ε. Generally, we have to distinguish three refinement strategies [87]. First there is the possibility of moving the nodes from positions with nearly uniform magnetization to the wall or vortices regions (r-refinement). Then there is the possibility of adding new nodes and elements into the elements having a big error indicator (h-refinement). The third way is to interpolate the direction cosines β_i, by polynomials of higher order instead of linear functions (p-refinement). The aim of an adaptive refinement algorithm is to get the "optimal" mesh, where the number of nodes is as small as possible while keeping the error below a given tolerance. All strategies are based on the idea of an equidistribution of the local error indicator to all mesh elements. Babuska and Rheinboldt [88] state that a mesh is almost optimal when the local errors are approximately equal for all elements. Adaptive mesh algorithms have to identify the regions where a higher spatial resolution is required. Therefore so-called error indicators are computed from the current finite element solution. In micromagnetics a reliable error indicator is based on the constraint condition for the norm of the magnetization vector [34]. Strong deviations of the norm show regions with non-uniform magnetization [89].

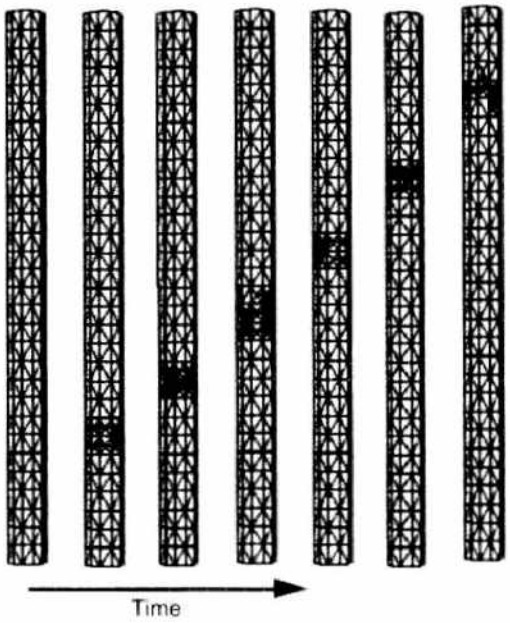


Figure 20. Sequence of meshes during the motion of the domain wall. The pictures give the triangular mesh on the surface of the wire. The mesh is dynamically adapted to the current magnetization distribution. Finite elements are introduced into regions where a high resolution is required. Afterward the mesh is coarsened again.

In the following, we outline the h-refinement scheme used in the micromagnetic simulations of domain wall motion. The finite element grid is adjusted to the current wall position during the solution of the Gilbert equation of motion. The mesh is refined in regions with nonuniform magnetization, whereas elements are taken out where the magnetization is uniform. Thus the fine grid moves together with the wall because the mesh can be coarsened as soon as the wall has passed by. Figure 21 shows the flow chart of the corresponding algorithm: We start with the initial mesh τ_0, initial magnetization $\mathbf{M}_0$, and the starting time t_0. At each time step i, *cvode* integrates the Gilbert equation of motion (Eq. [38]) and gives the new values for the magnetization $\mathbf{M}_i$ at time t_i. Furthermore, the error indicator η_T for all elements T of the given mesh τ_m is calculated. We control the adaptive refinement method with σ_1, σ_2 and a certain threshold value ε, which represents the allowed error.

Through proper refinement and coarsening steps, we obtain a mesh with the error indicator η_T for all elements. First, we treat the question, if coarsening of the current mesh is possible: Is the percentage of elements that fulfills $\eta_T < \sigma_2 \cdot \varepsilon$ higher than a given value (coarse criterion)? Then there are regions that just require a coarser mesh. So we can go back to the initial mesh $\eta_T < \tau_0$ and subsequently refine the mesh in regions with high-error indicators.

The other possibility is that we cannot coarsen our mesh. Then we must decide whether the mesh has to be refined. The condition is if there are elements with $\eta_T > \varepsilon$, then a refinement process [90] is required. We refine all elements whose error indicator fulfills $\eta_T > \sigma_1 \cdot \varepsilon$ with $\sigma_1 \in (0, 1)$ [91]. The algorithm leads to the new mesh τ_{m+1}. If the directions of the magnetization differ very much between neighboring elements and more than one refinement step is required to reduce the error indicator in a significant number of elements, then the refinement steps are repeated. Because a system of coupled equations has to be solved in every iteration of this algorithm, the number of iterations should be as small as possible. Thus, the marking strategy should select not too few mesh elements for refinement in each cycle. Therefore not just elements with $\eta_T < \varepsilon$ are marked for refinement, but in addition all elements fulfilling $\eta_T < \sigma_1 \cdot \varepsilon$.

In both cases, the saved magnetization $\mathbf{M}_{old}$ of the former time step is interpolated on the currently active mesh and the time is set to the former value t_{old}. This means if the error is too large or if the mesh can be coarsened, the time step is rejected and the finite element mesh is adjusted. Otherwise, if the mesh cannot be coarsened or does not have to be refined, the time integration is continued on the given grid. Thus the simulation proceeds in time only if the space discretization error is below a certain threshold.

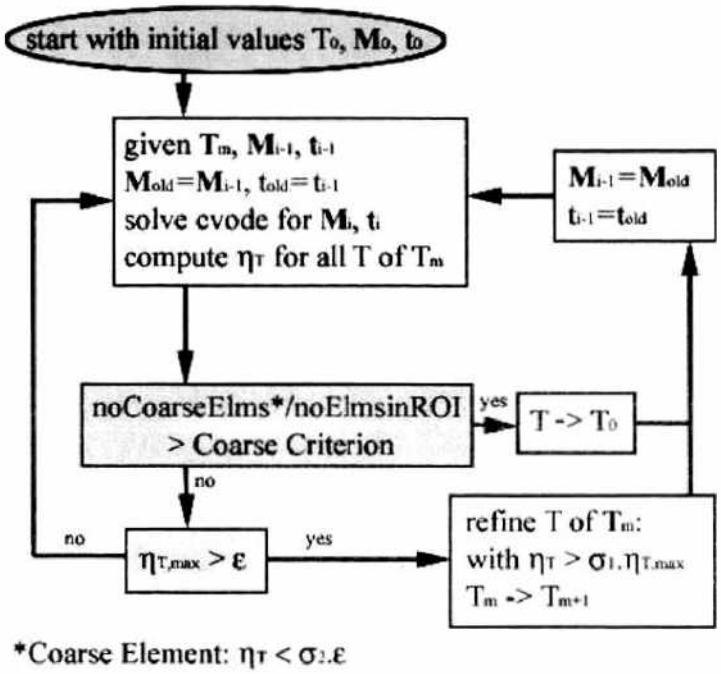


Figure 21. Flow chart of the adaptive mesh micromagnetic algorithm. The simulation proceeds in time only if the space discretization error is below a certain threshold ε.

This algorithm guarantees that the simulation proceeds in time only if the space discretization error is below a certain threshold. Numerical studies showed that the moving mesh reduces the total CPU time by more than a factor of 4 (Fig. 22).

The simulation of an array of nanowires would exceed the computational power of our workstations. Therefore, we just model three interacting nanowires. The influence of the other nanowires due to the strayfield is taken into account by a demagnetization factor. In the simulation, the external field is applied parallel and perpendicular to the long axis. Initially, the Co nanowire is fully saturated parallel to the long axis. Then the external field is instantaneously applied with a strength of 1400 kA/m. For each applied field value, we integrate the Gilbert equation of motion until the equilibrium state is reached and then the external field is reduced by steps of 28 kA/m. After the simulation of the total hysteresis curves, we have to consider that we just simulate three nanowires, whereas the experimental measurement is the result of the interaction within a two-dimensional array of nanowires. We see a good qualitative agreement with the experimental results. The nanowires show just a small hysteretic behavior. The values for the coercive fields are by a factor 2 bigger than the experimental values. For the perpendicular case, the magnetization starts to rotate in the smaller grains, which have a magnetocrystalline easy direction parallel to the long axis. Because the external field is sufficiently decreased also, the magnetization in the bigger grains with a magnetocrystalline easy direction perpendicular to the long axis starts to rotate. But in these grains, the reversal is finished earlier because the external field has to be strong to rotate the magnetization out of the magnetocrystalline anisotropy direction in the smaller grains. To the contrary, in the parallel case, the reversal starts in the bigger grains. We also performed simulations neglecting the granular structure of the Co nanowire. This means that the nanowire is just one single crystal with uniaxial anisotropy either parallel or perpendicular to the long axis. When the field is applied perpendicular to the long axes of the nanowires (Fig. 23), we see that the coercive field H_c is slightly increased with the increasing number of nanowires because the magnetostatic interaction between them. In addition, the shape of the hysteresis curves changes with more nanowires because, in the case of one nanowire, the arrangement of the granular structure in our finite element model influences the shape of the curve. With increasing number of nanowires, the number of grains is increased also and therefore the curve becomes smoother.

The simulations clearly show that we have to take into account the granular structure to achieve a good qualitative agreement with the experiment.

3.3. Permalloy Nanodots

The recent advances in microfabrication techniques [92] have stimulated interest in the properties of submicron-sized patterned magnetic elements [104, 105]. Promising applications include magnetic random access memory, high-density magnetic recording media, and

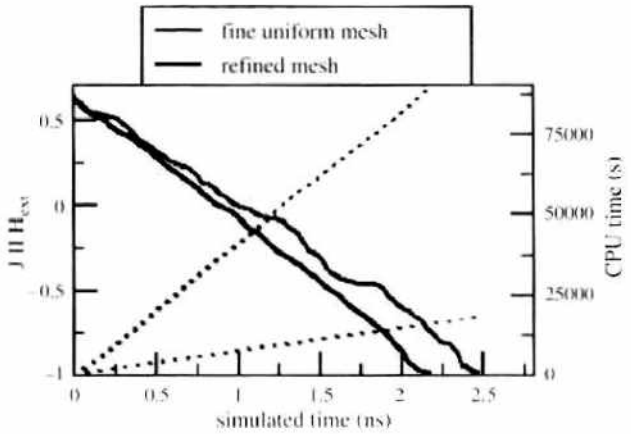


Figure 22. Domain wall motion in a Co nanowire. Full lines: Comparison of the magnetization using a very fine uniform mesh and an adaptive refined mesh. Dotted lines: The fine uniform mesh requires by a factor of 4 more CPU time.

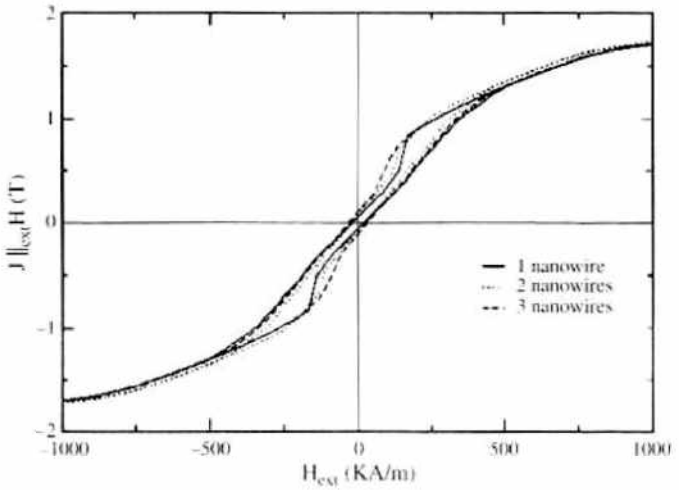


Figure 23. Hysteresis curves for different number of Co nanowires. The field is applied perpendicular to the long axis.

magnetic sensors [93]. However, in order to exploit the special behavior of magnetic nanoelements, it is necessary to study and understand their fundamental properties. The static and dynamic magnetic switching properties of magnetic nanodots with curling in-plane magnetization distribution (vortex) have widely been studied experimentally as well as theoretically. As an example, the magnetization states of cylindrical permalloy nanodots of different sizes and aspect ratios with analytical models and numerical finite element simulations, especially magnetic vortex states, will be compared. Direct experimental evidence for the existence of these magnetic vortex states has been found by the method of magnetic force microscopy. Shinjo and coworkers [94] have used magnetic force microscopy (MFM) to characterize magnetic nanodots of permalloy ($Ni_{80}Fe_{20}$) with a thickness of 50 nm and a radius between 300 nm and 1000 nm, for example. However, the lateral resolution is not high enough to estimate the diameter of the vortex core. In addition, the MFM tip is sensitive only to the out-of-plane component of the strayfield gradient, and the interaction between the magnetization of the nanodot and the MFM tip plays an important role for the contrast. These problems can be overcome using spin-polarized scanning tunneling microscopy, and the direct observation of the magnetization distribution in nanoscale iron islands with magnetic vortex cores have been reported [95]. Lorentz TEM allows *in situ* magnetizing experiments with thin samples, and it has been used to characterize the magnetization distribution in individual circular and elliptical particles [96]. The hysteresis loops of magnetic nanodots have been measured by vibrating sample magnetometer [97] and magneto-optical methods [98, 99]. Single domain and vortex states have been successfully identified. Furthermore, these magnetic vortex states are an interesting object for high-frequency magnetization dynamics [100] experiments, which are important for high-density magnetic recording media, where high-frequency field pulses of the magnetic write head store the information by reversing the magnetization. Typical material parameters of permalloy ($Ni_{80}Fe_{20}$), which are used in micromagnetic simulations, are the following: $J_s = 1.0$ T, $A = 13$ pJ/m, $K_1 = 0$, and $\alpha = 0.01$.

The discretization of a cylindrical nanodot into finite elements is rather crucial in order to describe magnetization distributions close to the vortex core. Figure 24 shows different finite element meshes of a cylindrical nanodot with a radius of $R = 100$ nm and a height of $L = 20$ nm. In order to investigate the influence of the finite element mesh on the results, three meshes with different mesh densities have been created (cf. Table 3). The first mesh is a uniform tetrahedral mesh with an average mesh size of 20 nm (Fig. 24[a]). Because the thickness of the nanodot is 20 nm, there is only one layer of elements. The second dot has an average mesh size of 6 nm, which resulted in four layers (Fig. 24[b]). The third mesh is an "adapted" mesh, with a very high density of nodes in the center, where the core of the vortex is found in zero external field (Fig. 24[c]). This is sensible because the vortex represents a singularity, which requires a very high numerical accuracy and therefore a lot of nodes. The outer regions are meshed with larger finite elements because the magnetization distribution

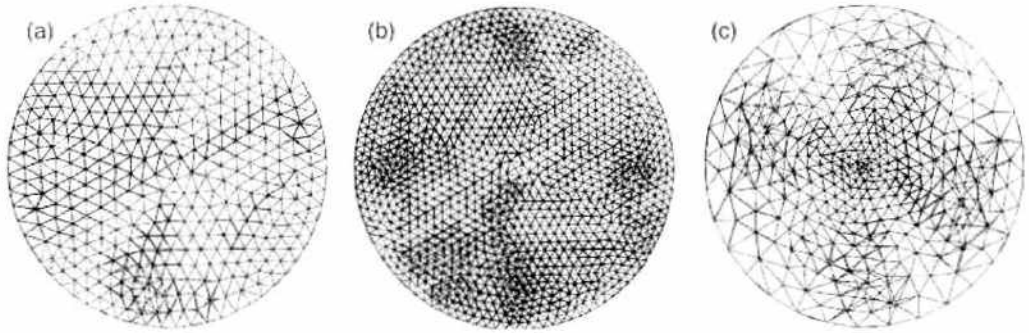


Figure 24. Finite element meshes of different mesh density for a circular nanodot: coarse mesh (a), fine mesh (b), and adapted mesh (c).

is rather uniform. First, the static reversal properties of the permalloy nanodots are numerically simulated and compared with analytical calculations. The finite element simulations have been initialized with the magnetization distribution of the rigid vortex model and an approximate core radius of about 11 nm.

Then the Landau-Lifshitz equation of motion for the magnetization has been integrated with a damping constant $\alpha = 1$ in zero field, and the magnetization relaxed to its equilibrium distribution, which minimizes the total Gibbs free energy. For the coarse mesh, the magnetization distribution given in Fig. 25(a) has been found. Obviously, the resolution of the mesh is too low to properly resolve the vortex. Thus, M_z is zero in the whole dot. The second mesh with a mesh size of 6 nm is fine enough to resolve the core (Fig. 25[b]). The adapted mesh, which has a high resolution with many small elements in its center and a gradually decreasing resolution toward the circumference, resolves the core just as well as the fine mesh (Fig. 25[c]), but its number of nodes and elements is comparable to that of the coarse mesh.

Therefore, it is very well suited for simulations, where the vortex core can be expected to be in the center of the nanodot. If the vortex core moves out of the center (e.g., due to an external field), it might reach a region with a low-resolution mesh. Then, the vortex core cannot be resolved properly anymore and the results become very unreliable. As a result, it is necessary to use either a uniform high-resolution mesh or to apply adaptive mesh refinement, which increases the mesh density at run time as required and optimizes (minimizes) the number of nodes and elements in the finite element mesh.

An analytical model for the magnetization distribution in zero field has been developed for a the rigid vortex state using a variational principle by Usov and coworkers [101, 102]. The rigid vortex model assumes a "rigid vortex," which does not change its shape in an external field. Together with a certain magnetization distribution it gives an approximation for the magnetization distribution of a curling state (vortex state) in a fine cylindrical particle. The core radius is obtained from the minimization of the total energy (exchange and magnetostatic energy). For permalloy with $R = 100$ nm and $L = 20$ nm a core radius of 11 nm is obtained.

The profile of M_z along the x-axis through the center of the dot for the different meshes is given in Fig. 26. The coarse grid clearly fails to resolve the vortex. However, the fine and the adapted grid are in excellent agreement. This emphasizes the importance of suitable meshing because the fine grid consists of more than four times more elements than the

Table 3. Details of the finite element meshes with different mesh density for a circular nanodot with a radius of $R = 100$ nm and a height of $L = 20$ nm corresponding to Fig. 24(a)–24(c).

Mesh data	Coarse mesh	Fine mesh	Adapted mesh
Mesh size	20 nm	6 nm	2–40 nm
Nodes	1437	6455	1397
Elements	5816	30,979	6256
Surface triangles	1758	5232	1390

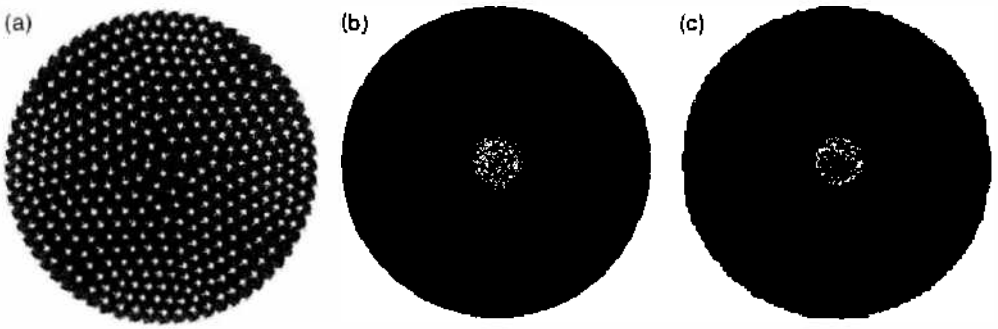


Figure 25. Magnetization distribution of the vortex state on the (a) coarse mesh, (b) fine mesh, and (c) adapted mesh. (a) The vortex core cannot be resolved, (b) the vortex core is properly resolved, and (c) the vortex core is nicely resolved, but the total number of elements and vertices is similar to that of the coarse mesh. Dark gray corresponds to in-plane magnetization and bright color to out of plane component magnetization M_z.

adapted one, which leads to much longer computation times. However, if we define the vortex core radius as that radius, where M_z is zero, we find a value of approximately 25 nm, which is considerably larger than the value of 11 nm predicted by the rigid vortex model. M_z is also quite uniform across the thickness of the dot. The results show that the vortex core obtained from the numerical simulations is larger (18.5 nm) than assumed by the analytical rigid vortex model (11 nm) due to a "broadening" of the M_z distribution (if the core radius is defined by $M_z = 0$). Furthermore, it is interesting to note that the finite element simulation shows that there is a region with $M_z > 0$ outside the core. Thus, we find positive surface charges in the core of the vortex, which are surrounded by negative surface charges. Only outside of approximately half the radius (50 nm) almost all surface charges disappear. It has been verified that there is very little variation of the magnetization distribution across the thickness of the nanodot.

Figure 27 shows the hysteresis curve for a circular nanomagnet with in-plane external field. For very high external fields (applied in the plane of the nanodot), the magnetization is almost uniform and parallel to the external field (Fig. 28[a]). As the field decreases (solid line in Fig. 27), the magnetization distribution becomes more and more nonuniform, which is caused by the magnetostatic strayfield. Upon further decrease of the external field, the symmetry of the magnetization distribution breaks and a "C" state (Fig. 28[b]) develops. At the nucleation field (about 5 kA/m for our example), a vortex nucleates on the circumference and quickly moves toward its equilibrium position (close to the center of the nanodot). As a result, we find a sudden drop in the average magnetization. When the external field is reduced to zero, the vortex moves into the center of the nanodot (Fig. 28[c]). If the external field is increased in the opposite direction, the vortex is forced out of the center of the dot. For about −70 kA/m, the vortex is pushed out of the nanodot (annihilation; Fig. 28[d]), and we find the second jump in the hysteresis curve to (almost) saturation.

This characteristic behavior has also been found experimentally using Hall-micromagnetometry by Hengstmann et al. [103], who measured the strayfield of individual permalloy disks using a submicron Hall magnetometer. The hysteresis loops of arrays of Supermalloy nanomagnets have been measured by Cowburn et al. [98] using the Kerr effect.

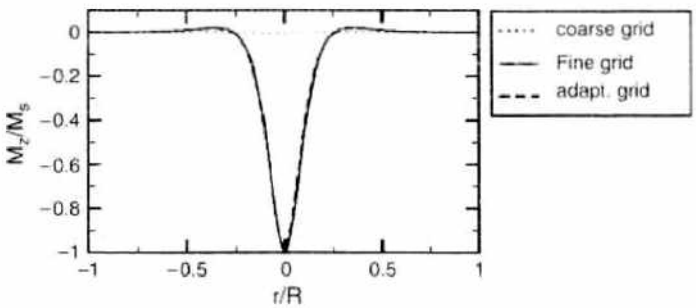


Figure 26. Profile of M_z along the x-axis through the center of the nanodot for different meshes.

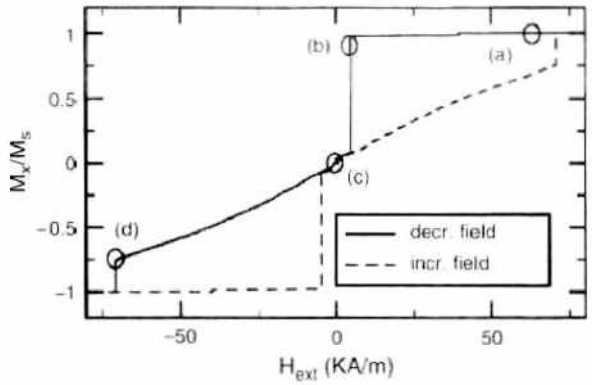


Figure 27. Hysteresis curve of a circular Permalloy nanodot with $R = 100$ nm, $L = 20$ nm for an in-plane external field. The circles mark the position on the hysteresis curve at which the snapshots in Fig. 28 have been taken.

Their characteristic loop shape has then been used to identify the single-domain in-plane and the vortex phase. The rigid vortex model can describe very well the susceptibility, magnetization distribution, and vortex annihilation field for low fields as well as the vortex nucleation field for a wide range of dot sizes [104–108]. The experimentally observed nucleation fields appear to be bigger than those predicted by the rigid vortex model [109]. This is probably because the simplest "C-shape" nucleation is not always an appropriate approach to describe the magnetization reversal in circular dots. For very high fields, we have an almost uniformly magnetized nanodot. For decreasing field the total energy increases (almost) linearly. The dashed line for positive field values indicates the total energy for the vortex state. At the intersection of the solid and the dashed line (at a value of about 35 kA/m for the external field), the vortex state and the uniform magnetization have equal energy. However, they are separated by an energy barrier that arises from the magnetostatic energy, which in turn is caused by the strayfield on the circumference of the nanodot as the vortex is pushed out of the center. Thus, the vortex state is a metastable state for external fields higher than 35 kA/m and the uniform state is metastable for external fields below 35 kA/m. The exchange energy remains approximately constant for negative external fields until the annihilation field is reached. Since all exchange energy is stored in the vortex core, this indicates that the vortex core remains undisturbed for even very large vortex shifts. For a twice as large nanodot with $R = 200$ nm and $L = 40$ nm, we find a nucleation field of 28 kA/m and an annihilation field of 84 kA/m. The corresponding hysteresis loop is given in Fig. 29. In general, the initial susceptibility, the vortex nucleation, and the annihilation fields depend on the dot's saturation magnetization M_s and should scale universally as a function of the dimensionless dot-aspect ratio $L = R$ [105, 108].

The investigation and improved understanding of the dynamic processes in magnetic nanostructures become more and more important because magnetic nanoparticles are promising candidates for high-speed, high-density magnetic storage (e.g., hard disks and MRAMs) and sensor devices [110, 111]. Recent advances in nanometer scale fabrication

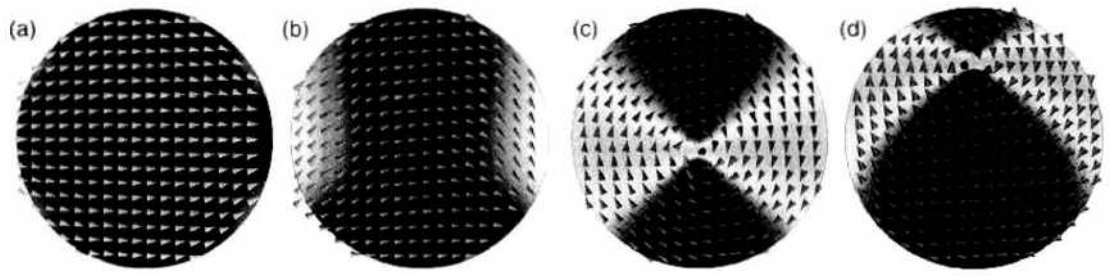


Figure 28. Typical magnetization distributions along the hysteresis loop. The snapshots have been taken at the corresponding position on the hysteresis curve indicated in Fig. 27. (a) Almost homogeneous magnetization, (b) "C" state before the vortex nucleates, (c) centered vortex in zero field, (d) magnetization distribution before annihilation of the vortex. The external field is applied parallel to the magnetization direction of pattern (a).

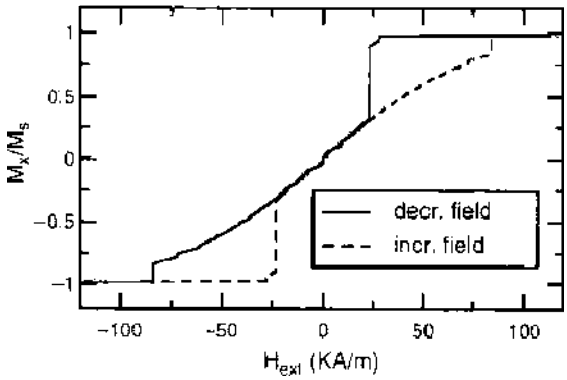


Figure 29. Hysteresis curve of a nanodot with a radius of 200 nm and a thickness of 40 nm.

technology allow detailed experimental investigations on a nanometer length scale [97, 112]. In high-speed storage devices, magnetization reversal occurs by applying short-field pulses [113]. The characteristic switching time depends on the reversal mode and is usually in the nanosecond regime. The magnetization reversal excites many spin-excitation modes (spin waves), whose understanding is important to determine the field-dependent spin instability regions, where spontaneous or thermally assisted magnetization reversal might occur [114]. The magnetization dynamics under short-field pulses have been investigated in saturated NiFe disks [115] and in closure domains in Co disks [116]. However, the magnetization dynamics of the magnetic vortex state in thin permalloy disks is markedly different from those in the uniformly magnetized state and also from the spin waves observed in thin magnetic films. The dynamic behavior of the magnetic nanodots has been studied by instantaneously applying an external field of 6.4 kA/m (80 Oe) in plane perpendicular to the dot axis (z-axis). Even though the experiment was started from the equilibrium magnetization distribution in zero field, M_x and M_y show a quite irregular behavior during the first 0.5 ns. During this time, the vortex core "adapts" to the applied external field and deforms while it already starts its precession toward equilibrium. A low damping constant of $\alpha = 0.05$ has been used. The size of the finite elements is quite important. A coarser mesh leads to a bad approximation of the vortex core and an inaccurate result. The simulations using the uniform meshes give results that are in good agreement with experiments and analytical considerations. The precession frequency of 0.65 GHz is also confirmed by the results of Guslienko and coworkers [117]. In addition it has been found that the magnetostatic energy oscillates in phase with M_x or M_y, respectively. This has to be ascribed to variations in the surface charge density on the circumference. The time evolution of $\langle M_x \rangle$ (the average of M_x over the whole nanodot) for a dot with with an aspect ratio of $L/R = 20$ nm/100 nm $= 0.2$ is shown in Fig. 30. The damped oscillation, which is caused by the spiral motion of the vortex core toward its equilibrium position, is observed. The corresponding Fourier spectrum reveals a sharp peak at a frequency of about 0.75 GHz.

Figure 31 shows the results of the translation mode eigenfrequencies of various nanodots with a radius $R = 100$ nm and a thickness between 10 nm and 40 nm. The results are in

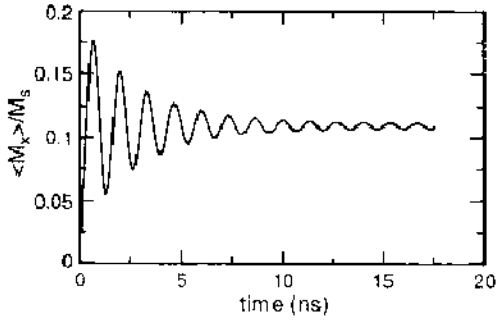


Figure 30. Simulation results of the oscillation of $\langle M_x \rangle$ as a function of time for vortex precession in a nanodot with $L/R = 20$ nm/100 nm $= 0.2$ under applied in-plane field $\mu_0 H_x = 0.01$ T.

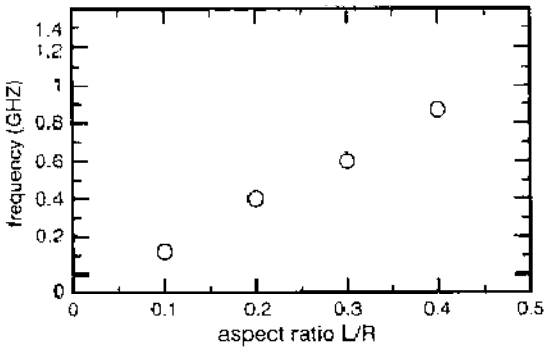


Figure 31. Translation mode eigenfrequencies versus aspect ratio L/R for nanodots with $R = 100$ nm.

good agreement with the results of a finite difference model and the analytical "two-vortices" model presented in Ref. [117]. The decreasing total energy (dissipation due to damping with $\alpha = 0.05$ in the Landau-Lifshitz equation of motion) and the swapping between magnetostatic and Zeeman energy (which shifted by 180°) are shown in Fig. 32. The exchange energy remains constant because the vortex core, which accounts for most of the exchange energy, precesses without changing its shape. This confirms the analytical description of the translational mode suggested in Ref. [118]. Direct experimental observation of this mode in an isolated vortex using time-resolved Kerr microscopy has recently been reported by Park et al. [119]. There is good qualitative agreement with the analytical and numerical models but still a few questions concerning the quantitative discrepancies and damping times remain open.

3.4. Layered-Shaped Granular Layers (Two-Dimensional Structures)

In common hard disk drives, the information is stored in the magnetization state of a ferromagnetic granular film. A granular film is a composition of grains that are regions with the same crystal structure favoring the magnetization to align parallel to one axis, called *easy axes*. Usually neighboring grains have the same crystal structure, but the easy axes are different. In modern hard disk drives, the grain-size diameter is less than 10 nm. In longitudinal recording, the easy axes are randomly oriented in the plane of the film. One bit is represented by the magnetization state of about 200 to 300 neighboring grains. Because of the in-plane anisotropy in longitudinal recording, the magnetization is parallel to the film plane. Small grains are required because the minimum bit length, below which neighboring transitions become indistinguishable, is determined by the transition width, which in turn depends on the grain size of the film. Usually the grains are weakly exchange coupled to each other. For too strongly exchange coupled grains, the magnetization in neighboring grains aligns parallel, and effectively larger grains are formed. Thus, magnetic interaction increases the effective grain size. As a consequence, larger bit lengths would be required. However, completely decoupled grains are thermally unstable because the thermal stability decreases with decreasing grain volume. Thus, an important task in longitudinal recording is to find the optimal exchange coupling strength between grains.

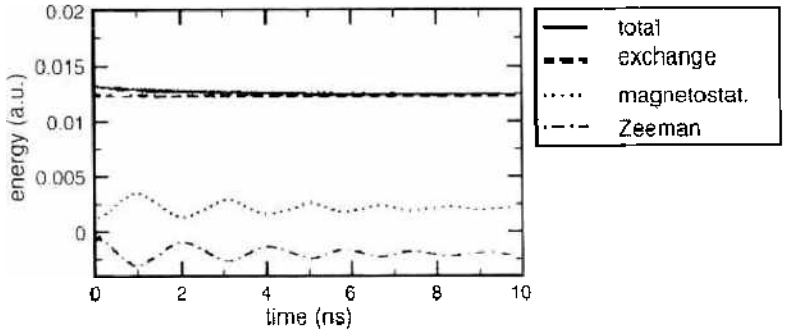


Figure 32. Energy over time for a dot with $L/R = 0.1$ and $\mu_0 H_z = 0.01$ T.

Thin films for conventional longitudinal magnetic recording are typically based on $Co_{84-x}Cr_{16}M_x$ alloys with additions of M = Pt, Ta, B [120]. The phenomenon of Cr grain boundary segregation was also found experimentally by energy-filtered TEM imaging [121]. Recording data densities larger than 10 Gbits/in^2 are possible when the media have thermally stable, magnetically decoupled grains with grain sizes less than 10 nm. To increase data storage density (i.e., to decrease the bit size and to increase the signal to noise ratio), higher symmetry textures and narrow grain size distributions are necessary. In order to separate individual data bits, arrays of patterned thin-film elements are fabricated by lithographic and/or imprint techniques. As an example, the magnetization reversal dynamics of a longitudinal recording medium has been simulated for the granular microstructure described in Fig. 2 consisting of 420 in-plane randomly oriented grains with about 8 nm grain size and 15 nm film thickness. In order to simulate the write process in a 100 Gbits/in^2 medium, a bit size of 50×130 nm^2 was assumed with a head speed of 20/ms corresponding to a trapezoidal external field profile with 2.5 ns bit writing time, including 0.1 ns rise and 0.1 ns decay time. The bit structure is shown in Fig. 33, and the maximum external field is about 1.3 H_k.

The micromagnetic simulations start form the state of randomly magnetized grains and clearly show in Fig. 34(a) that a domain wall structure in the thin-film medium consisting of fully coupled grains after writing bit A (after 2.5 ns) is formed during the bit-writing process. However, in the case of fully decoupled grains (Fig. 34[b]), the simulations show individually switched grains after writing bit A and B (after 5 ns). Partially switched grains produce a large contribution to the media noise. The material parameters used for the simulations are $J_s = 0.37$ T, $K_1 = 0.2$ MJ/m^3, $A = 10$ pJ/m, $\alpha = 0.02$. For the exact simulation of a real longitudinal medium, the reduced magnetic exchange in the intergranular region between the grains due to the Cr-segregation has to be known and taken into account.

In order to study the role of the variation of the intergranular exchange coupling between neighboring grains, we have studied the magnetization reversal processes of two and three grains, respectively, with a diameter of 16 nm and different in-plane orientations and considering variable intergrain exchange constants values. Assuming perfectly coupled grains, we obtain a complex magnetization reversal process with the formation of domain walls as shown in Figs. 35(a) and 37(a). Fully exchange-decoupled grains lead to the individual reversal process within each grain (i.e., a more or less homogeneous precessional rotation similar to the one of a single particle described in Fig. 10(a). The quasistatic demagnetization curves of Fig. 36 clearly show the "two-step" behavior in the case of two completely decoupled grains. According to the misorientation of the decoupled grains, different switching fields are obtained for both grains A and B. The switching is first observed in the misoriented grain B (Figs. 35[b] and 37[b]) and C (Fig. 37[b]). In the case of exchange coupled grains, both grains switch simultaneously (Figs. 35[a] and 37[a]). Figure 38 shows the dependence of the switching time on the degree of intergrain exchange coupling between the three grains. The direct coupled grains show the shortest switching time of about 300 ps and the completely

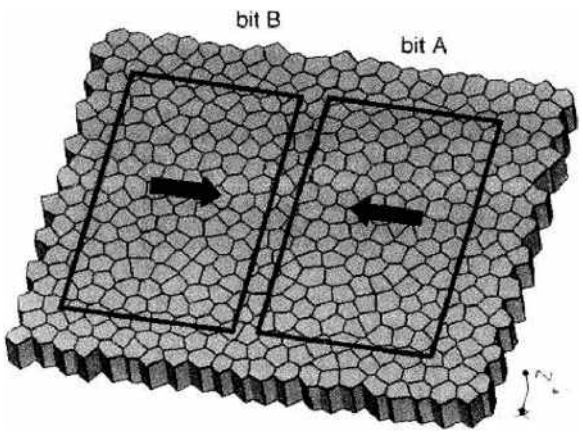


Figure 33. Micromagnetic model of 420 in-plane randomly oriented grains used for the simulation of the switching of two bits A and B in a CoCrPtX medium for longitudinal recording.

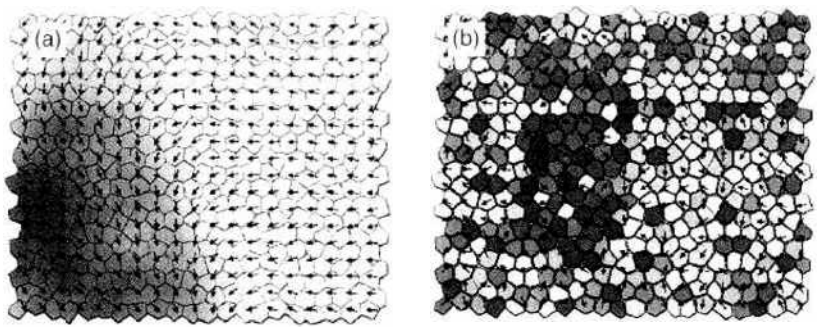


Figure 34. Micromagnetic simulations showing the difference of the switching of the grains after writing single bits in a CrCoPt longitudinal medium. Fully coupled grains after 2.5 ns (a) and fully decoupled grains after 5 ns (b).

decoupled grains show the precessional oscillation of the total polarization leading to a complete switching after about 1 ns. In reality an intergranular exchange in the order of about 10%–30% of the bulk value seems to be sufficient to stabilize individual bits and to keep the switching time and magnetization oscillations small.

The idea of perpendicular recording is to represent bits with magnetization directions perpendicular to the film plane. A perpendicular orientation of the magnetization reduces the demagnetizing field in the high-density limit. Thus, opposite bits act as domains that reduce the stray-field. To achieve a perpendicularly magnetized configuration, textured films with easy axes perpendicular to the film plane are used. Another advantage of the aligned grains in perpendicular recording is a narrow switching field distribution. In conventional longitudinal media the grains are oriented randomly. Because the switching field depends on the angle between the easy axis and the external field, some grains may not switch, leading to a broadening of the transition between the bits.

In comparison to the modeling of the bit writing in a longitudinal medium with 100 Gbit/in^2 areal storage density, the micromagnetic model of Fig. 33 was used to simulate the switching of individual grains in a (Co, Fe)(Pt, Pd) thin-film medium with perpendicular magnetocrystalline anisotropy. The material parameters used for the simulations are $J_s = 0.30$ T, $K_1 = 0.1$ MJ/m^3, $A = 10$ pJ/m, $\alpha = 0.02$. A random misorientation of the grains up to $\pm 5°$ is taken into account. Figure 39 shows a typical TEM image of a CoPd thin film with decoupled grains with an average grain size of about 10 nm. For the writing and stabilizing of individual bits, the intergranular exchange coupling becomes important as in the case of the longitudinal magnetic recording medium. It is obvious from the domain images of Fig. 40 obtained during the writing of bit A that the individual bit structure cannot be stabilized in a granular film consisting of completely coupled grains. Figure 41 shows the results of micromagnetic simulations that reveal a clear bit pattern after writing bit A and B

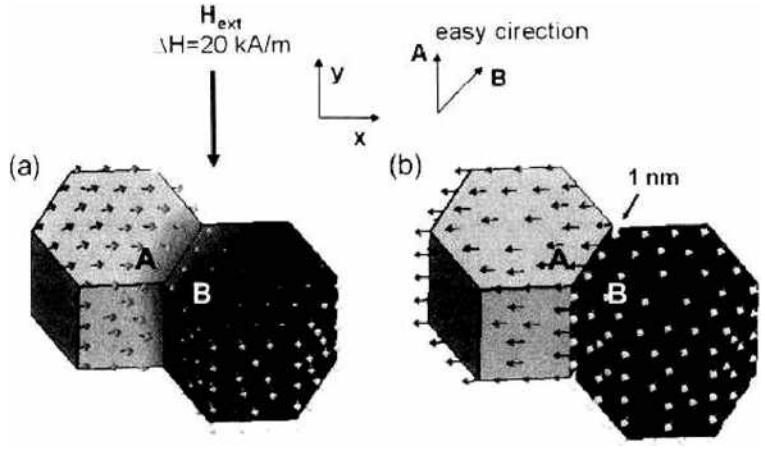


Figure 35. Magnetization reversal of two grains. (a) A domain wall is formed near the grain boundary of direct coupled grains. (b) Individual homogeneous precessional rotation of magnetization of decoupled grains. The misoriented grain B switches first.

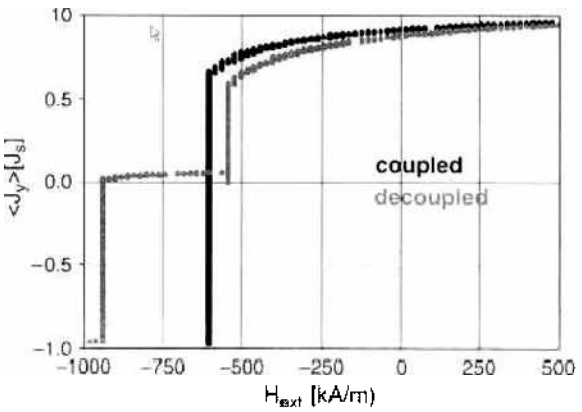


Figure 36. Quasistatic demagnetization curves of the "2 grain" model of Fig. 35 showing an individual switching of decoupled grains and simultaneous switching of fully exchange coupled grains.

(after 5 ns) considering fully decoupled grains. The jagged bit shape obviously results in a contribution to the media noise.

3.5. Nanocrystalline Composite Magnets

Nanocomposite permanent magnets have been widely studied experimentally and theoretically during the past decade [122–132]. These magnets consist of a mixture of magnetically hard and soft phases. Nanocomposite magnets show a high remanence and a reasonable large coercive field if both phases are sufficiently exchange coupled. Nanocomposite magnets with excellent hard magnetic properties were obtained for various different compositions. The soft magnetic phase is either α-Fe or Fe_3B. As hard magnetic phase $Nd_2Fe_{14}B$, $SmCo_5$, $Sm_2Fe_34N_x$, and $Pr_2Fe_{14}B$ were used. Possible application of nanocomposite permanent magnets are bonded magnets used in consumer electronic applications, where the miniaturization requires magnets that are easy to magnetize [132, 133]. In addition, the reduction of the total rare-earth content may lead to a cost reduction. Whereas the enhanced remanence of nanocomposite magnets is an advantage as compared with isotropic, single-phase nanocrystalline NdFeB magnets, the coercive field obtained in two-phase systems is rather low. Thus recent studies investigated different ways to improve the coercive field of nanocomposite magnets. Kanekiyo and Hirosawa [129] showed that the hard magnetic properties of nanocomposite magnets can be significantly improved by modifying the alloy compositions. The substitution of Nd by Dy increases the hard-phase anisotropy, giving rise to an improved coercivity without a significant loss in remanence. Fukunaga and co-workers

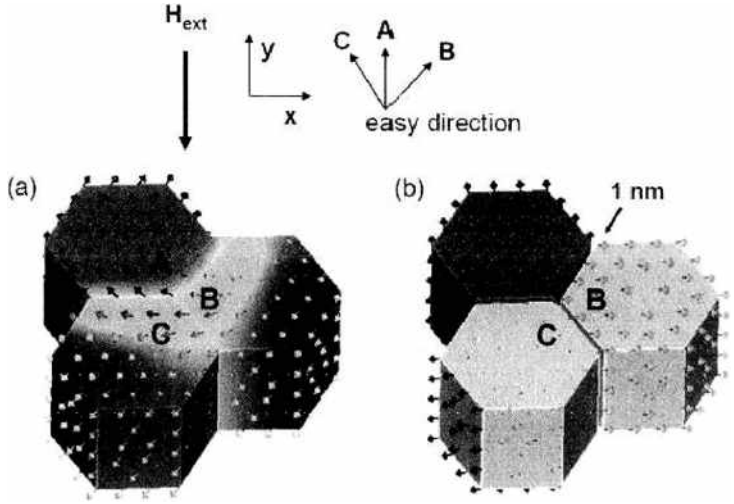


Figure 37. Transient magnetization patterns during switching. (a) Fully exchange coupled grains switch simultaneously. (b) Fully decoupled grains switch independently depending on the misorientation of the grains.

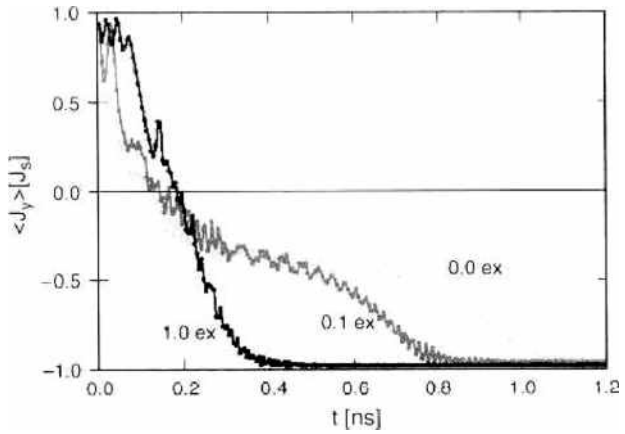


Figure 38. Switching time in dependence of the intergranular coupling according to the "3 grain" model of Fig. 37. The shortest switching time is found for the direct coupled (1.0 ex) grains. Completely decoupled grains (0.0 ex) show a strong oscillation of the polarization during switching.

[126] proposed a two-phase microstructure consisting of NdFeB and α-Fe grains embedded within a residual amourphous phase in order to improve the coercive field of nanocomposite magnets. Such a nanocomposite system was realized by Hamano and co-workers [131] using a composition $Nd_8Fe_{70}Co_8Nb_2B_6$ with a maximum $H_c = 575$ kA/m.

Finite element modeling treats magnetization processes within on a length scale of several nanometers and thus give a quantitative correlation between the microstructure and the magnetic properties of nanocomposite magnets. The numerical solution of the Gilbert equation of motion shows how reversed domains nucleate and expand. Magnetization reversal starts at the grain boundaries dissolving a strongly nonuniform magnetic state. The intrinsic magnetic properties near the grain boundaries strongly influence the remanence and the coercive field of two phase, nanocrystalline magnets. The magnetic properties are assumed to decrease gradually from their bulk values in the intergranular phase. A linear reduction of the exchange constant to 20% of its bulk increases the coercive field by up to 50% without a significant reduction of the remanence in NdFeB/α-Fe magnets with a mean grain size of 10 nm. In order to explain the influence of intergrain exchange interactions, see Fig. 42.

Figure 42 schematically illustrates remanence enhancement and magnetization reversal in two-phase nanocrystalline permanent magnets. In order to explain the influence of intergrain exchange interactions Kneller [122] proposed a model where in the magnetization of the different phases is connected by mechanical springs. At a high-external field, the magnetization of all grains becomes oriented upward. As the field is gradually reduced to zero.

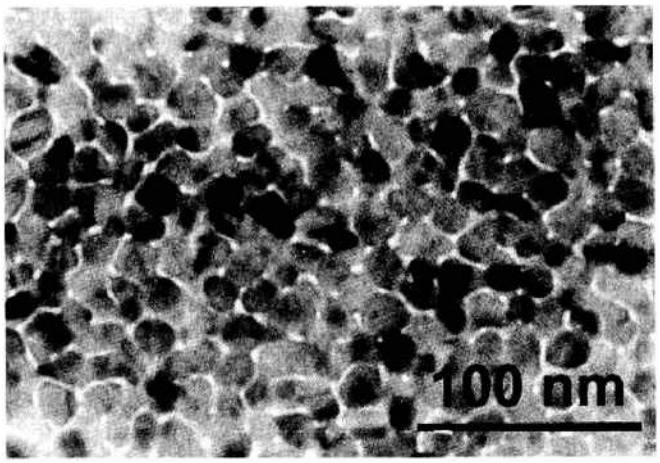


Figure 39. TEM image of the granular microstructure of a CoPd multilayer recording medium used for perpendicular recording.

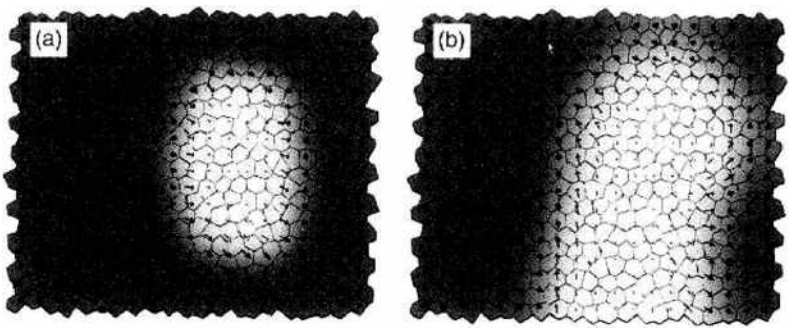


Figure 40. Transient domain images during the writing of bit A in the perpendicular (Co, Fe)(Pd, Pt) thin-film medium with fully coupled grains after (a) 1.5 ns and (b) 2.5 ns.

the magnetization of the hard magnetic grains rotate toward the direction of their local easy axes. The magnetization of the soft magnetic grains remains parallel to the saturation direction. The springs of neighboring hard magnetic grains like to pull the magnetization into different directions. These competitive effects cancel and at zero applied field, the magnetization of the soft phase points parallel to the average magnetization direction of neighboring hard magnetic grains. Under the influence of a reversed external field, the magnetization of the soft phase rotates reversibly. If the field is removed, the magnetization of the soft grains will rotate back to its original direction that is the so-called exchange-spring effect. At higher opposing fields, the magnetization of the soft grains may even point downward. The connecting springs provide a high torque onto the neighboring hard magnetic magnetization, leading to the irreversible switching of the hard magnetic grains. A reduction of the stiffness constant reduces the torque onto the hard magnetization and thus shifts irreversible switching toward higher values of the opposing field. An appropriate choice of the exchange stiffness will keep remanence enhance and improve the coercive field. Thus, it is possible to tailor the magnetic properties of nanocomposite magnets, changing the amount and the intrinsic properties of an intergranular phase. An appropriate reduction of the intergrain exchange may improve the coercive field but still keep the effect of remanence enhancement. Inoue et al. [123] and Hamano et al. [124] successfully prepared α-Fe/$Nd_2Fe_{14}B$ magnets with a remaining amorphous phase. The amorphous intergranular phase is expected to reduce the exchange interactions between the hard and the soft grains, leading to an improved coercive field. A coercive field of $H_c = 405$ kA/m was obtained for nanocomposite magnets produced from a $Nd_8Fe_{76}Co_8Nb_2B_6$ meltspun alloy. The corresponding remanence and energy-density product were $J_r = 1.16$ T and $(BH)_{max} = 147$ kJ/m^3 [124]. Atom probe characterization of $Nd_8Fe_{75.5}Co_8B_6Nb_1Cu_{0.5}$ show that intergranular phase is enriched with

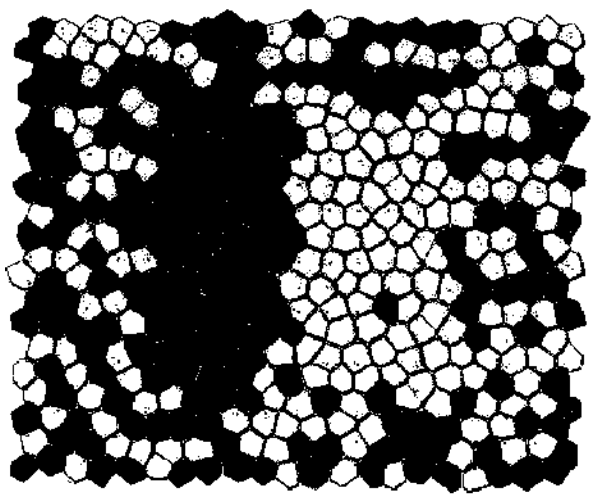


Figure 41. Simulation of the bit writing in a perpendicular (Co, Fe)(Pd, Pt) thin-film with fully decoupled grains after 5 ns.

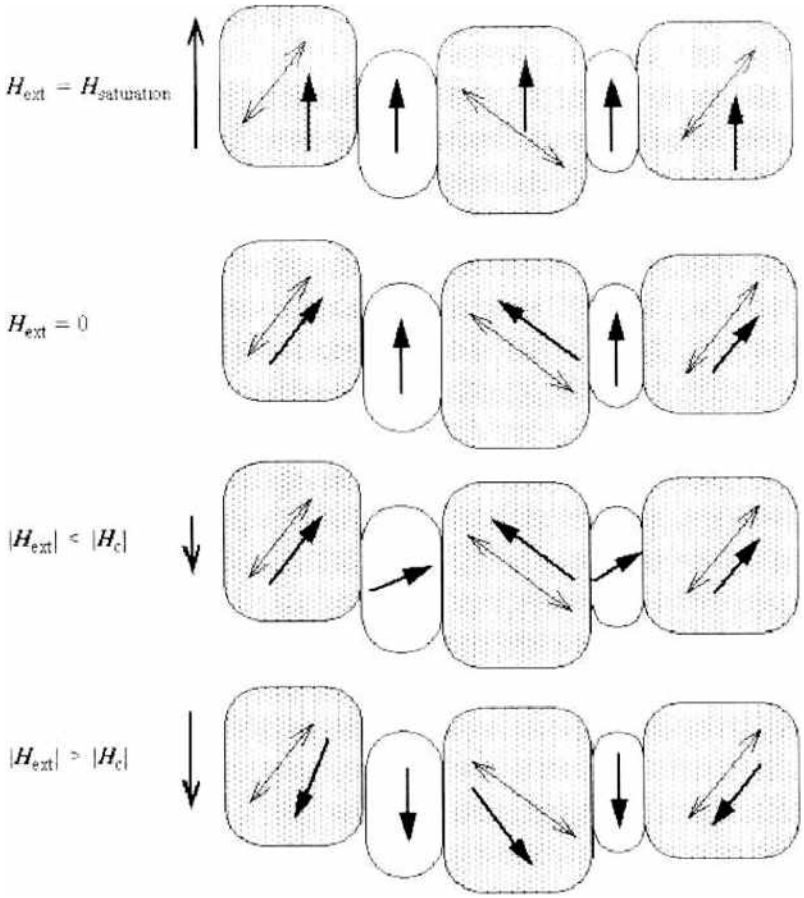


Figure 42. Schematics of remanence enhancement and magnetization reversal in two-phase nanocrystalline permanent magnets. The magnetocrystalline anisotropy directions of the hard phase (dark) are randomly oriented. At zero applied field, the magnetization of the soft phase (bright) points parallel to the average magnetization direction of neighboring hard magnetic grains. At the critical field, exchange interactions between the hard and the soft phase causes the irreversible switching of a neighboring hard magnetic grain.

Nb and B [125]. Mössbauer studies suggest that the proposed intergranular phase is actually ferromagnetic with magnetic properties that gradually decrease toward the grain boundary. Recently, Fukunaga and co-workers [126] numerically investigated the effect of reduced strength of intergrain exchange interactions in α-Fe/$Nd_2Fe_{14}B$ systems. Their results confirm the increase of coercivity owing to reduced intergrain exchange interactions, provided that grain size is sufficiently small. The increase of coercive field without a significant loss of the remanence enhances the energy density product. A reduction of the intergrain exchange constant to 20% of its bulk value increases $(BH)_{max}$ from about 200 kA/m to 310 kA/m, assuming a grain size of 8 nm. Previously, Fischer and Kronmüller [127, 128] studied the influence of intergrain exchange interactions in single-phase and two-phase nanocrystalline NdFeB magnets, using micromagnetic finite element simulations. Partial decoupling of the grains increases the coercive field and decreases the remanence in single-phase, nanocrystalline magnets. The reduction of the exchange constant in an intergranular phase with a thickness of 3 nm to 20% of its bulk value decreases the coercive field by a factor 1/2 for a mean grain size of 20 nm and a volume fraction of 50% α-Fe. This work uses dynamic micromagnetic finite element simulations to investigate the effects of intergranular phases on the magnetic properties of α-Fe/$Nd_2Fe_{14}B$ magnets. The exchange constant, A, is reduced gradually toward the grain boundaries in a simple model system consisting of single soft magnetic grain surrounded by hard magnetic neighbors. Both the width of the intergranular phase and the amount of reduction of A influence the coercive field, in addition to the grain size. The optimum parameters are derived from the numerical results and applied to large-scale simulations using a realistic microstructure. The simulations provide considerable insight into the mechanisms that determine the remanence and the coercive field.

Particularly, the results show how intergrain exchange interactions trigger the nucleation and expansion of reversed domains.

3.5.1. Intergrain Exchange Interactions and Magnetization Reversal of Individual Grains

Figure 43 shows a simple finite element model of an individual grain within a two-phase nanocrystalline magnet. A single α-Fe grain is surrounded by 20 $Nd_2Fe_{14}B$ grains. The anisotropy directions of the hard magnetic grains are randomly oriented. The hard magnetic grains are truncated so that the volume fraction of the soft grains is 36%. The grains are subdivided into tetrahedral finite elements. The total number of elements of the individual grain is 34.000. The element size is approximately 1 nm. The grains are either in direct contact or are separated by an intergranular phase of width W. Within this intermediate phase, the exchange constant varies as shown in Fig. 44. The exchange constant decreases linearly from its bulk value, A_{bulk}, to rA_{bulk} at the interface between two neighboring grains. The same procedure is used for both types of grain boundaries, hard/hard and soft/hard interfaces. The gray scale plot of Fig. 43 gives the exchange constant at the surface of the structure. For numerical convenience, the Gilbert damping constant was set to $\alpha = 1$.

In addition to the width, W, and the reduction, r, the grain diameter of the grain, D, was varied in the calculations. Thus the model system is characterized by the three parameters (W, r, D). The intrinsic material parameters used for the calculations are shown in Table 4 [122, 134]:

Figure 45 compares the numerically calculated demagnetization curves for a system with perfect exchange interactions between the grains A: ($W = 0$, $r = 1$, $D = 10$ nm), and for a system with an intergranular phase B: ($W = 2$ nm, $r = 0.2$, $D = 10$ nm). In addition to the total magnetic polarization given by the solid line, the graphs give the contributions from the hard magnetic phase (dashed line) and the soft magnetic phase (dotted line). Whereas the intergranular phase has only a minor effect on the remanence, which changes from $J_r = 1.27$ T to $J_r = 1.24$ T, the coercive field increases from $H_c = 440$ kA/m to $H_c = 580$ kA/m, as the intergrain exchange interactions are reduced. Furthermore, the overall shape of the demagnetization curve remains the same for both simulations.

The curves denoting the contribution of $Nd_2Fe_{14}B$ and the contribution of α-Fe to the total magnetic polarization are nearly parallel, which indicate that hard and soft phases

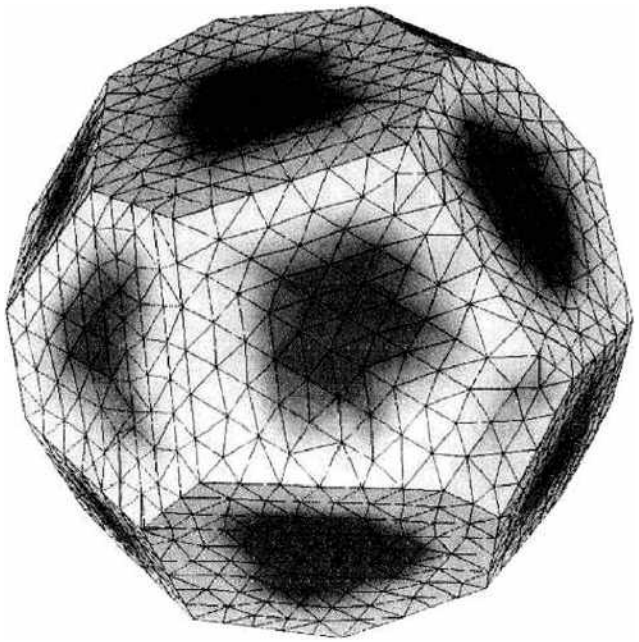


Figure 43. Finite element model at the outer surface of an individual grain within a two-phase nanocrystalline magnet. The variation of the exchange constant near the grain boundaries is shown in the gray scale map. The strength of the exchange interactions decreases linearly toward the grain boundaries ($W = 3.2$ nm, $r = 0.2$).

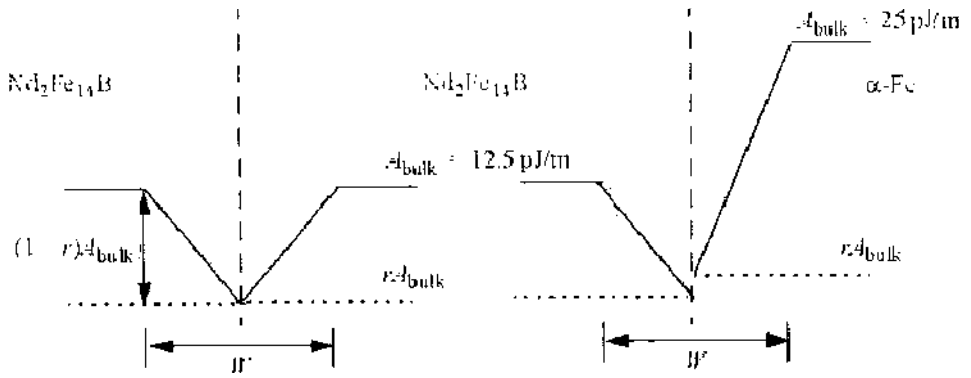


Figure 44. Intrinsic properties of the intergranular phase. Within a region of thickness $W/2$ the exchange constant decreases linearly from its bulk value to rA_{bulk}.

switch together when the external field reaches a critical value. For small opposing fields, the magnetization of the soft grain rotates reversibly out of its equilibrium direction at zero applied field. During this process, the magnetization remains moderately uniform with the α-Fe particle. However, intergrain exchange interactions provide a torque onto the neighboring hard magnetic grains, causing the magnetization to rotate out of their local magnetocrystalline anisotropy axes near the grain boundaries. Because both phases are exchange coupled, the deviations of the magnetization from the local anisotropy axes near the grain boundaries increase with increasing rotation of the magnetization within the α-Fe grain. This gives rise to an increase of the magnetocrystalline anisotropy energy with decreasing external field. This process is more pronounced in system A where the grains are in direct contact. Owing to the perfect coupling, the anisotropy energy is considerably larger in system A than in system B, with reduced intergrain exchange interactions. Owing to the intergranular phase, the magnetization within the α-Fe grain may rotate without causing major deviations of the hard-phase magnetization from the local easy axes. This process which finally initiates irreversible switching occurs at much higher opposing fields as compared with perfect coupling.

Figure 46 compares the magnetization distribution in the undercritical state, just before irreversible switching is initiated, for the systems A, B, and C ($W = 3.2$ nm, $r = 0.2$, $D = 10$ nm). With increasing width of the intergranular phase, the effective coupling between the grains decreases. The plots clearly show that the maximum possible rotation increases with decreasing intergrain exchange interactions. If the external field exceeds a critical value, the rotation of the magnetization within the soft grain becomes too large and the neighboring hard magnetic grain starts to reverse irreversibly. The irreversible switching is initiated at the interface between the different phases. The plots illustrate the nucleation and expansion of a reversed domain at $H_{ext} = 440$ kA/m for system A. The reversed nucleus is plotted at different times during irreversible switching. A similar process, which will occur at higher opposing fields, is observed for systems B and C. The numerical results are summarized in Fig. 47, which gives the coercive field as a function of the exchange constant rA_{bulk} next to the grain boundaries. The coercive field increases linearly with r for a grain diameter of the α-Fe grain of $D = 10$ nm. Increasing the width of the intergranular phase from $W = 2$ nm to $W = 3.2$ nm shifts the line toward higher values of the coercive field. Finally, Fig. 48 gives the calculated demagnetization curves for different diameters of the α-Fe grain. While keeping $W = 3.2$ and $r = 0.2$ constant, D was changed from 10 nm to 20 nm. The results clearly

Table 4. Intrinsic magnetic properties. The spontaneous magnetic polarization, J_s, the bulk value of the exchange constant, A_{bulk}, and the anisotropy constants K_1 and K_2.

Phase	Anisotropy	J_s (T)	A_{bulk} (pJ/m)	K_1 (MJ/m³)	K_2 (MJ/m³)
$Nd_2Fe_{14}B$	Uniaxial	1.61	12.5	4.6	0.66
α-Fe	Cubic	2.15	25	0.046	0.015

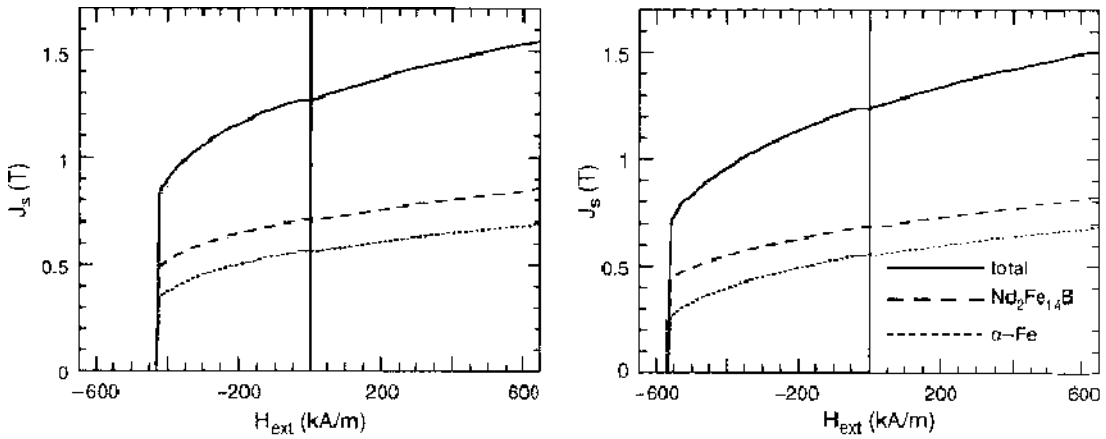


Figure 45. Numerically calculated demagnetization curve for the model system of Fig. 43. The grain diameter of the α-Fe grain was 10 nm. (A) No intergranular phase, perfect exchange interactions between the grains. (B) Intergranular phase with width $W = 2$ nm and reduction $r = 0.2$.

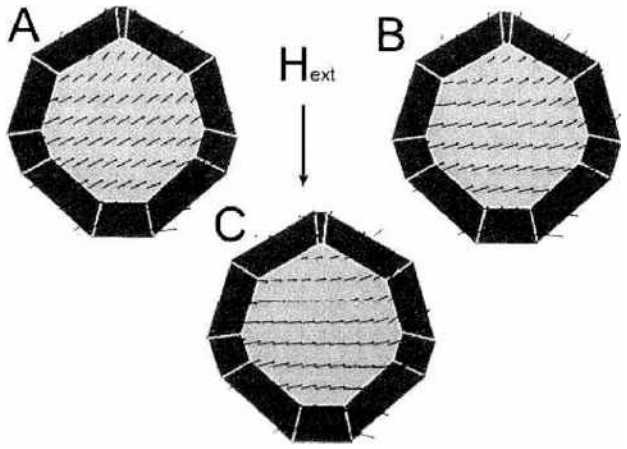


Figure 46. Magnetization distribution in the undercritical state just before irreversible switching is initiated. (A) No intergranular phase, perfect exchange interactions between the grains. (C) Intergranular phase with width $W = 3.2$ nm and reduction $r = 0.2$. The arrows give the projection of the magnetization onto a slice plane parallel to the external field.

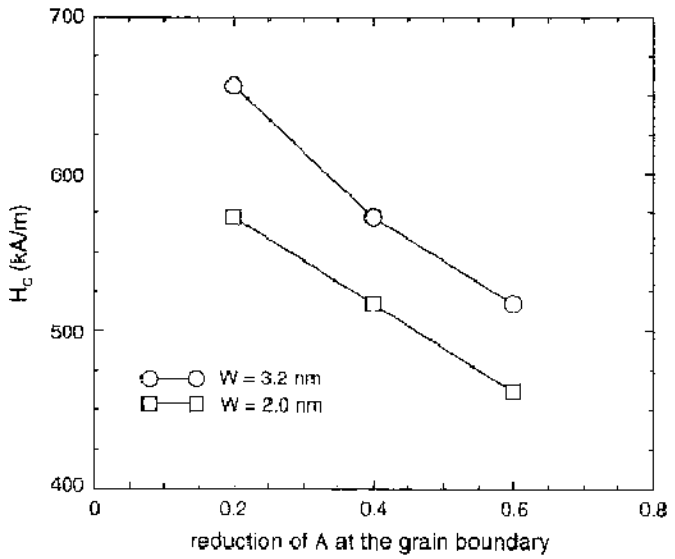


Figure 47. Dependence of the coercive field on the width W of the intergranular phase with reduced exchange interactions. The plots give the coercive field as a function of the reduction r for a grain diameter $D = 10$ nm.

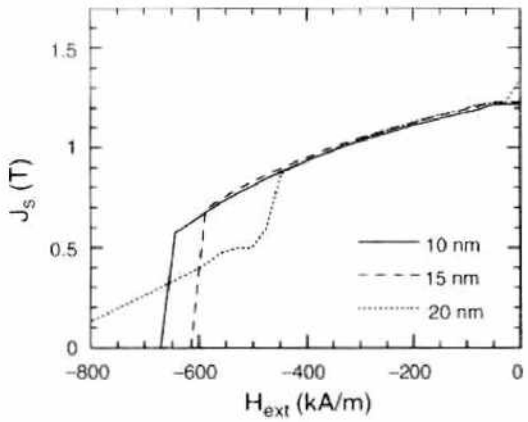


Figure 48. Calculated demagnetization curves as a function of the grain size. Intergranular phase with width W = 3.2 nm and reduction $r = 0.2$.

show a grain diameter $D > 15$ nm deteriorates the squareness of the demagnetization curve. The results confirm that an improvement of the magnetic properties due to a remaining amorphous phase is only possible for sufficiently small grain size.

3.5.2. Complex Grain Structure

Remanence enhancement and magnetization reversal were simulated for a realistic grain structure of a two-phase α-Fe/$Nd_2Fe_{14}B$ with remaining amorphous phase, using the optimum parameters $W = 3.2$ nm, $r = 0.2$, and $D = 10$ nm. The grains are obtained from a Voronoi construction [135, 136], where the grains are assumed to grow with constant velocity from randomly located seed points. In order to avoid strongly irregular shaped grains, the magnet is first divided into cubic cells. Within each cell a seed point is chosen at random. After grain growth simulation, the intrinsic magnetic properties are assigned to the grains such that the volume fraction of α-Fe is 50% and the magneto-crystalline anisotropy axes of $Nd_2Fe_{14}B$ are randomly oriented. Between the grains an intergranular region with reduced exchange constant is assumed in order to mimic the remaining amorphous phase [125]. Figure 49 shows the phase distribution and the intergranular phase at the surface of model magnet, consisting of 125 grains. The mean grain size was 10 nm. The average thickness of the intergranular with was 3.2 nm. The exchange constant of the intergranular phase was reduced to 20% of its bulk value. Now a step function was used to model the decrease of the exchange constant, in contrast to the previous calculations where the exchange constant was reduced linearly toward the interface. This assumption keeps the number of finite

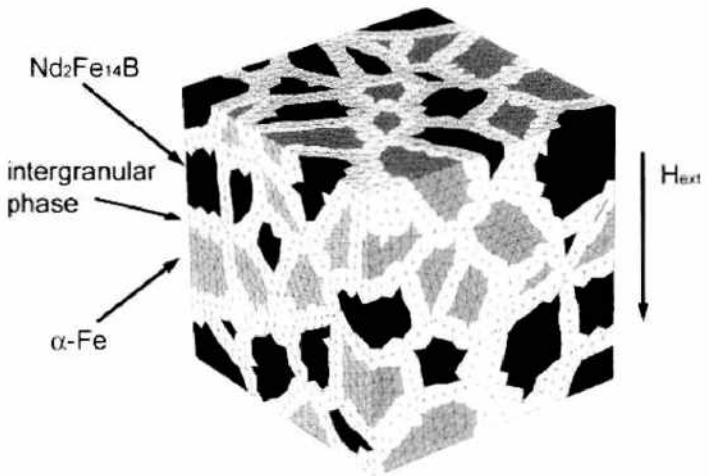


Figure 49. Model of a nanocomposite α-Fe/$Nd_2Fe_{14}B$ magnet with residual amorphous phase, consisting of 125 grains. The volume fraction of α-Fe is 50%. The mean grain size is 10 nm. The thickness of the amorphous intergranular phase is 3.2 nm.

elements small as compared with the number of elements required for a gradual change of the intrinsic magnetic properties. The total number of finite elements used to descretize the model show in Fig. 49 is about 76,000.

Figure 50 shows the magnetization distribution in the remanent state and at applied fields of $H_{ext} = -420$ kA/m. The nearly uniform magnetic state for zero-applied fields gives rise to high remanence. The remanence is $J_r = 1.37$ T, giving a ratio of the remanent to saturation polarization, is $J_r/J_{sat} = 0.73$, which considerably exceeds the theoretical limit for the remanence of noninteracting grains. The remanence ratio depends on the crystal symmetry, the crystallographic orientation of the easy directions, and the volume fraction and the saturation polarization of the phases [137]. The remanence ratio of a two-phase nanocomposite magnet is

$$\frac{J_r}{J_{sat}} = \frac{1}{J_{sat}}\{v_1 m_1 J_{s1} + (1 - v_1) m_2 J_{s2}\} \tag{104}$$

where m_1, m_2 denote the remanence ratio of phase 1 and phase 2 and v_1, v_2 are the volume fraction of the phases. Without any interactions, the remanence ratio of $Nd_2Fe_{14}B$ (uniaxial anisotropy) and α-Fe (cubic anisotropy) is $m_1 = 0.5$ and $m_2 = 0.83$ [138], respectively. For a volume fraction of 50% α-Fe, J_{sat} is $J_{s1} + J_{s2}$. With $J_{s1} = 1.61$ of $Nd_2Fe_{14}B$, $J_{s2} = 2.15$, the remanence ratio of noninteracting particle becomes $J_r/J_{sat} = 0.68$. The calculated remanence ratio $J_r/J_{sat} = 0.73$ exceeds this theoretical limit. This result clearly indicates that intergrain exchange interactions enhance the remanence of nanocomposite magnets, even so intergrain exchange interactions are considerably reduce due to the presence of the amorphous phase. The magnetic polarization J is nearly parallel to the anisotropy direction within the hard magnetic grains. Intergrain exchange interactions determine the direction of J within the α-Fe grains, where J is parallel to the average magnetization direction of all neighboring $Nd_2Fe_{14}B$ grains. This average direction corresponds to the saturation direction for an isotropic distribution of the hard phase easy axes. At an external field of $H_{ext} = -420$ kA/m the magnetization distribution becomes strongly nonuniform, as the soft magnetic grains reverse reversibly. Owing to the reduction of the exchange constant at the interface between the different phases, soft grains may reverse while most of the hard magnetic grains remain unswitched. The total magnetic polarization at this field is $J = 0.48$ T. This value is lower than the corresponding value of the simple-model system. Neighboring soft magnetic grains form continuous regions of soft magnetic phase that increase the effective soft magnetic feature size and deteriorates the loop shape.

Micromagnetic finite element simulations show that an intergranular phase with reduced exchange interactions improves the coercive field of nanocrystalline α-Fe/$Nd_2Fe_{14}B$ magnets, provided that the grain size is sufficiently small. The coercive field increases from $H_c = 440$ kA/m to $H_c = 660$ kA/m for a mean grain size of 10 nm, without a significant reduction of the remanence, if the grains separated by an intergranular phase with a thickness of 3.2 nm. The minimum value of the exchange constant in the boundary phase was 20% of its bulk value.

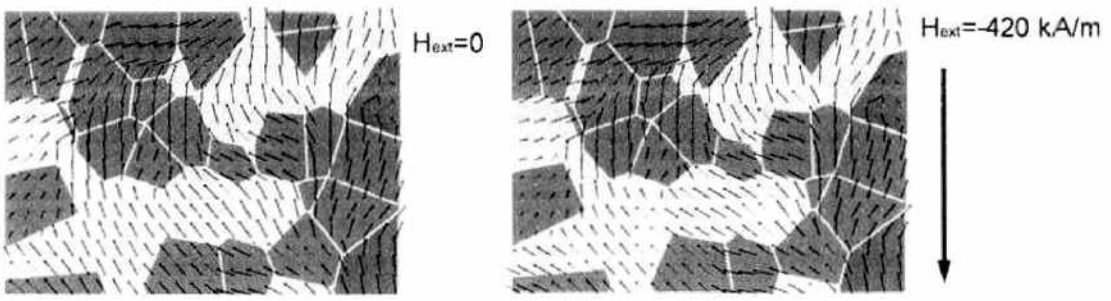


Figure 50. Magnetization distribution at the remanent state and under the influence of a reversed field. The arrows give the projection of the magnetization onto a slice-plane parallel to the external field. α-Fe may reverse without the switching of neighboring $Nd_2Fe_{14}B$ grains.

ACKNOWLEDGMENTS

The authors would like to express their thanks to Drs. Dieter Suess, Hermann Forster, and Rok Dittrich for all their helpful, inspiring, and fruitful discussions. This work has partly been supported by the EC project HIDEMAR (GRD1-2001-40316) and by the Austrian Science Fund projects Y132PHY and P14899.

REFERENCES

1. L. Landau and E. Lifshitz, *Phys. Z. Sowjetunion* 8, 153 (1935).
2. W. F. Brown, *Phys. Rev.* 58, 736 (1940).
3. W. F. Brown, "Micromagnetics," Interscience, New York, 1963.
4. R. Comerford, *IEEE Spec.* 37, 36 (2000).
5. S. R. Das, *IEEE Spec.* 40, 19 (2003).
6. C. H. Möller et al., *Appl. Phys. Lett.* 80, 3988 (2002).
7. S. Tehrani et al., *IEEE Trans. Magn.* 36, 2752 (2000).
8. J.-G. Zhu, Y. Zheng, and G. A. Prinz, *J. Appl. Phys.* 87, 6668 (2000).
9. K. J. Strnat et al., *J. Appl. Phys.* 38, 1001 (1967).
10. M. Sagawa et al., *J. Appl. Phys.* 55, 2083 (1984).
11. J. J. Croat et al., *J. Appl. Phys.* 55, 2078 (1984).
12. G. C. Hadjipanayis, R. C. Hazelton, and K. R. Lawless, *J. Appl. Phys.* 55, 2073 (1984).
13. S. Sasaki, J. Fidler, and M. Sagawa, in "High Performance Sintered Magnets Made by New RIP from SC Alloys," Proceeding XI. International Symposium on Magnetic Anisotropy and Coercivity in Rare Earth Transition Metal Alloys (H. K. a. M. H. a. M. Okada, Ed.), 2000, pp. 109–118.
14. Y. Kaneko, *IEEE Trans. Magn.* 36, 3275 (2000).
15. W. Rodewald et al., *IEEE Trans. Magn.* 38, 2955 (2002).
16. W. Scholz et al., *IEEE Trans. Magn.* 39, 2920 (2003).
17. G. C. Hadjipanayis et al., *IEEE Trans. Magn.* 36, 3382 (2000).
18. J. Fidler and T. Schrefl, *J. Phys. D* 33, R135 (2000).
19. J. Fidler et al., *Comput. Mater. Sci.* 25, 554 (2002).
20. J. Fidler et al., *J. Appl. Phys.* 91, 7974 (2002).
21. T. R. Koehler and D. R. Fredkin, *IEEE Trans. Magn.* 28, 1239 (1992).
22. A. Hubert and R. Schäfer, "Magnetic Domains," 1998.
23. V. D. Tsiantos et al., *J. Magn. Magn. Mater.* 242–245, 999 (2002).
24. T. L. Gilbert, *Phys. Rev.* 100, 1243 (1955).
25. J. Alberty, C. Carstensen, and S. A. Funken, *Numerical Algorithms* 20, 117 (1999).
26. J. Bey, *Computing* 55, 355 (1995).
27. J. G. Zhu, *Mater. Res. Soc. Bull.* 20, 49 (1995).
28. P. G. Ciarlet, "The Finite Element Method for Elliptic Problems," 1978.
29. R. Ribó, M. A. d, R. Pasenau, and E. Escolano, http://gid.cimne.upc.es, International Center For Numerical Methods in Engineering (CIMNE), 2002.
30. W. Rave, K. Ramstöck, and A. Hubert, *J. Magn. Magn. Mater.* 183, 329 (1998).
31. C. Carstensen and F. Funken, *Computing* 62, 243 (1999).
32. K. M. Tako et al., *J. Appl. Phys.* 81, 4082 (1997).
33. R. Hertel and H. Kronmüller, *IEEE Trans. Magn.* 34, 3922 (1998).
34. W. Scholz, T. Schrefl, and J. Fidler, *J. Magn. Magn. Mater.* 196–197, 933 (1999).
35. H. Forster et al., *J. Appl. Phys.* 91, 6914 (2002).
36. H. Forster, Ph.D. Thesis, Vienna University of Technology, 2003.
37. A. Aharoni, *IEEE Trans. Magn.* 27, 3539 (1991).
38. Q. Chen and A. Konrad, *IEEE Trans. Magn.* 33, 663 (1996).
39. B. Streibl, Dipl. Thesis, Vienna University of Technology, 1998.
40. A. Aharoni, "Introduction to the Theory of Ferromagnetism," Monographs on Physics, Oxford University Press, 1996.
41. L. Berger, *Phys. Rev. B* 54, 9353 (1996)
42. S. Ingvarsson et al., *Phys. Rev. B* 66, 214416 (2002).
43. K. Capelle and B. L. Gyorffy, *Europhys. Lett.* 61, 354 (2003).
44. T. Gilbert, Armour Research Foundation, Report No. 11, 1955.
45. C. W. Gardiner, 1996.
46. D. R. Fredkin and T. R. Koehler, *IEEE Trans. Magn.* 26, 415 (1990).
47. T. R. Koehler, *Physica B: Condens. Matter* 233, 302 (1997).
48. D. A. Lindholm, *IEEE Trans. Magn.* 20, 2025 (1984).
49. D. Süss, Dipl. Thesis, Vienna University of Technology, 1999.
50. D. Süss, T. Schrefl, and J. Fidler, *IEEE Trans. Magn.* 36, 3282 (2000).
51. S. Benson et al., TAO home page, http://www-unix.mcs.anl.gov/tao/, 2001.
52. S. Benson and J. Moré, Revision 2.1.3, Tech. Rep. ANL-95/11, Revision 2.1.3: Argonne National Laboratory, 2002.

53. G. D. Byrne and A. C. Hindmarsh, PVODE home page. http://www.llnl.gov/CASC/PVODE/. 2002.
54. G. D. Byrne and A. C. Hindmarsh, *Int. J. High Perf. Comput. Appl.* 13, 354 (1999).
55. S. D. Cohen and A. C. Hindmarsh, Lawrence Livermore National Laboratory Report UCRL-MA-118618, 1994.
56. S. D. Cohen and A. C. Hindmarsh, *Comput. Phys.* 10, 138 (1996).
57. P. N. Brown and A. C. Hindmarsh, *J. Appl. Math. Comp.* 31, 40 (1989).
58. D. Suess et al., *J. Magn. Magn. Mater.* 248, 298 (2002).
59. A. C. Hindmarsh and L. R. Petzold, *Comput. Phys.* 9, 34, 148 (1995).
60. G. Henkelman and H. Jònsson, *J. Chem. Phys.* 113, 9978 (2000).
61. R. Dittrich et al., *J. Magn. Magn. Mater.* 250, 12 (2002).
62. R. Dittrich, Ph.D. Thesis, Vienna University of Technology, 2003.
63. D. Weller et al., *IEEE Trans. Magn.* 36, 10 (2000).
64. E. C. Stoner and E. P. Wohlfarth, *Philos. Trans. R. Soc. London* 240, 599 (1948).
65. E. C. Stoner and E. P. Wohlfarth, *IEEE Trans. Magn.* 27, 3475 (1991).
66. S. Chikazumi, "Physics of Ferromagnetism." 1997.
67. R. Kikuchi, *J. Appl. Phys.* 27, 1352 (1956).
68. Y. Yu and J. W. Harrell, *IEEE Trans. Magn.* 30, 4083 (1994).
69. L. Néel, *Ann. Geophys.* 5, 99 (1949).
70. M. Lederman et al., *J. Appl. Phys.* 75, 6217 (1994).
71. A. D. Kent et al., *J. Appl. Phys.* 76 (1994).
72. J. G. S. Loket, *Phys. Rev. B* 58, 12201 (1998).
73. W. Wernsdorfer et al., *Phys. Rev. Lett.* 77, 1873 (1996).
74. S. Guéron et al., *Phys. Rev. Lett.* 83, 4148 (1999).
75. S. Sun et al., *Science* 287, 1989 (2000).
76. B. Bian et al., *J. Appl. Phys.* 87, 6962 (2000).
77. T. Klemmer et al., *Scripta Metall. Mater.* 33, 1793 (1995).
78. W. Scholz et al., *J. Magn. Magn. Mater.* (2004).
79. S. Anders et al., *Microelectron. Eng.* 61–62, 569 (2002).
80. L. L. Afremov and A. V. Panov, *Phys. Met. Metallogr. (USSR)* 86, 269275 (1998).
81. Y. Labaye et al., *J. Appl. Phys.* 91, 8715 (2002).
82. K. Nielsch et al., *IEEE Trans. Magn.* (2002).
83. W. Yang, D. N. Lambeth, and D. E. Laughlin, *J. Appl. Phys.* 87, 6884 (2000).
84. H. Forster et al., *J. Magn. Magn. Mater.* 249, 181 (2002).
85. J. Miltat and M. Labrune, *IEEE Trans. Magn.* 30, 4350 (1994).
86. M. J. Donahue, *J. Appl. Phys.* 83, 6491 (1998).
87. A. Raizer et al., *J. Appl. Phys.* 67, 5803 (1990).
88. I. Babuska and W. Rheinboldt, *SIAM J. Num. Anal.* 15, 736 (1978).
89. A. Bagneres-Viallix, P. Baras, and J. B. Albertini, *IEEE Trans. Magn.* 27, 3819 (1991).
90. H. P. Langtangen, "Computational Partial Differential Equations—Numerical Methods and Diffpack Programming." 1999.
91. A. Schmidt and K. G. Siebert, "Numerical Aspects of Parabolic Free Boundary Problems—Adaptive Finite Element Methods." 2000.
92. P. Candeloro et al., *Japan. J. Appl. Phys.* 41, 5149 (2002).
93. R. P. Cowburn, *J. Phys. D* 33, R1 (2000).
94. T. Shinjo et al., *Science* 289, 930 (2000).
95. A. Wachowiak et al., *Science* 298, 577 (2002).
96. M. Schneider, H. Hoffmann, and J. Zweck, *J. Magn. Magn. Mater.* 257, 1 (2003).
97. C. A. Ross et al., *Phys. Rev. B* 65, 144417 (2002).
98. R. P. Cowburn et al., *Phys. Rev. Lett.* 83, 1042 (1999).
99. G. Gubbiotti et al., *IEEE Trans. Magn.* 38, 2532 (2002).
100. V. Novosad et al., *Phys. Rev. B* 66, 052407 (2002).
101. N. A. Usov and S. E. Peschany, *J. Magn. Magn. Mater.* 118, L290 (1993).
102. N. A. Usov and S. E. Peschany, *Fiz. Met. Metalloved.* 12, 13 (1994) [*Phys. Met. Metallog.*].
103. T. M. Hengstmann et al., *J. Appl. Phys.* 90, 6542 (2001).
104. K. Y. Guslienko and K. L. Metlov, *Phys. Rev. B* 63, 100403(R) (2001).
105. K. Y. Guslienko et al., *Phys. Rev. B* 65, 024414 (2002).
106. M. Schneider, H. Hoffmann, and J. Zweck, *Appl. Phys. Lett.* 77, 2909 (2000).
107. A. Fernandez and C. J. Cerjan, *J. Appl. Phys.* 87, 1395 (2000).
108. K. Y. Guslienko et al., *Appl. Phys. Lett.* 78, 3848 (2001).
109. V. Novosad et al., *IEEE Trans. Magn.* 37, 2088 (2001).
110. M. Hehn et al., *Science* 272, 1782 (1996).
111. G. A. Prinz, *Science* 282, 1660 (1998).
112. C. A. Ross, *Annu. Rev. Mater. Res.* 31, 203 (2000).
113. A. Lyberatos et al., *J. Appl. Phys.* 91, 2236 (2002).
114. O. Chubykalo et al., *Phys. Rev. B* 65, 184428 (2002).
115. W. K. Hiebert, A. Stankiewicz, and M. R. Freeman, *Phys. Rev. Lett.* 79, 1134 (1997).
116. Y. Acremann et al., *Science* 290, 492 (2000).

117. K. Y. Guslienko et al., *J. Appl. Phys.* 91, 8037 (2002).
118. K. Y. Guslienko and A. N. Slavin, *J. Appl. Phys.* 87, 6337 (2000).
119. J. P. Park et al., *Phys. Rev. B* 67, 020403R (2003).
120. J. C. Lodder, *Thin Solid Films* 291–292, 474 (1996).
121. J. E. Wittig et al., *Mater. Res. Soc. Symp. Proc.* 517, 211 (1998).
122. E. F. Kneller, *IEEE Trans. Magn.* 27, 3588 (1991).
123. A. Inoue et al., *Mater. Trans. JIM* 36, 962 (1995).
124. M. Hamano et al. in "15th International Workshop on Rare-Earth Magnets and Their Applications." Werkstoff-Informationsgesellschaft, Dresden, Germany, 1998.
125. Y. Q. Wu et al., *IEEE Trans. Magn.* 35, 3295 (1999).
126. H. Fukunaga, J. Kuma, and Y. Kanai, *IEEE Trans. Magn.* 35, 3235 (1999).
127. R. Fischer and H. Kronmüller, *J. Magn. Magn. Mater.* 184, 166 (1998).
128. R. Fischer and H. Kronmüller, *J. Appl. Phys.* 83, 3271 (1998).
129. H. Kanekiyo and S. Hirosawa, *J. Appl. Phys.* 83, 6265 (1998).
130. V. Archambault and D. Pere, *Mater. Res. Proc.* 577, 153 (1999).
131. H. Hamano et al., *Mater. Res. Proc.* 577, 153 (1999).
132. V. Panchanathan, *IEEE Trans. Magn.* 31, 3605 (1995).
133. J. J. Croat, *J. Appl. Phys.* 81, 4804 (1997).
134. M. Sagawa et al., *J. Appl. Phys.* 57, 4094 (1985).
135. F. P. Preparata and M. I. Shamos, "Computational Geometry." Springer, New York, 1988.
136. T. Schrefl and J. Fidler, *J. Magn. Magn. Mater.* 111, 105 (1992).
137. K. H. Mueller et al. in "8th International Symposium on Magnetic Anisotropy and Coercivity in RE-TM Alloys." University of Birmingham, Birmingham, UK, 1994.
138. G. Bertotti, "Hysteresis in Magnetism—For Physicists, Materials Scientists and Engineers." 1998.

CHAPTER 12

Quantum Theory of Spintronics in Magnetic Nanostructures

J. Mathon,[1] A. Umerski[2]

[1] *Department of Mathematics, City University, London, United Kingdom*
[2] *Department of Applied Mathematics, Open University, Milton Keynes, United Kingdom*

CONTENTS

ISBN: 1-58883-049-7

Handbook of Theoretical and Computational Nanotechnology
Edited by Michael Rieth and Wolfram Schommers
Volume 7: Pages (677–746)

1. INTRODUCTION

Magnetic nanostructures occupy a special place among all the other nanostructures because the spin angular momentum of electrons (holes) plays in them the role of a new degree of freedom which can be exploited to design devices with novel quantum-mechanical functions. Thus a new area of condensed matter physics called spintronics was created some 15 years ago. When we talk in this review about magnetic nanostructures, we have in mind structures that may be macroscopic in two dimensions but their length in the third dimension is on a nanoscale. They are often referred to as magnetic multilayers and a typical magnetic multilayer is shown schematically in Fig. 1. Magnetic multilayers pose great challenge both to experiment and theory. Experimentally, they have to be grown with atomic control over the thicknesses of individual layers and with interfaces between the layers so perfect that, ideally, the momentum of carriers parallel to the layers is conserved. This is now possible for metallic magnetic multilayers, which are the subject of this review. The restriction to metallic multilayers is for two reasons. First research into properties of metallic magnetic multilayers is by far most advanced but the second reason is that currently they are the only systems that can be used in spintronics. Multilayers based on magnetic semiconductors are being studied extensively, but, at present, their low Curie temperature prevents their commercial application in spintronics.

Theoretically, magnetic multilayers are also challenging since they are inherently inhomogeneous systems. However, for perfect epitaxial layers, the inhomogeneity is only in one dimension and we shall see that such a quasi-one-dimensional inhomogeneity, even when random, can be modeled quite realistically.

The key feature that distinguishes magnetic multilayers from conventional (macroscopic) magnets is the experimentally established fact that the transport of charge is mediated in them by carriers whose spin remains conserved across the whole thickness of the multilayer. In a conventional magnet, the spin of an individual charge carrier flips between its up and down projections over a characteristic distance l_{sf} (spin diffusion length), which is of the order of several tens of nanometers. This is much shorter than the dimensions of conventional electronics components. It follows that spin memory is lost in macroscopic samples and, therefore, the spin angular momentum plays no role in transport of charge. This is why for decades spin was ignored by conventional semiconductor electronics. However, when an ultrathin layer structure with a thickness smaller than l_{sf} is prepared the spin "remembers" its orientation across the whole thickness of the structure, which means that carriers with different spin orientations do not mix and flow independently as if in two separate wires connected in parallel [1]. If the multilayer contains magnetic components, then the two spin channels are inequivalent. This is because the numbers of charge carriers with up and down spin are unequal in a ferromagnet, and, even more importantly, up- and down-spin carriers are scattered at different rates at nonmagnet/magnet interfaces. This has interesting and highly exploitable consequences. Because of different scattering rates for up- and down-spin carriers at nonmagnet/magnet interfaces, the total resistance of a magnetic nanostructure depends on the magnetic configuration of all its magnetic components. This in turn can be altered by an applied magnetic field and, therefore, the resistance of a magnetic nanostructure can be changed by the applied field. The effect is known as the giant magnetoresistance (GMR) since the relative change of the resistance can be very large, of the order of 100%. When two magnetic electrodes are separated by an insulating tunneling barrier, a similar effect, called tunneling magnetoresistance (TMR), can occur. The GMR effect was discovered about 15 years ago [2], and a large TMR effect was first observed in 1995 [3]. With the discovery of the GMR effect, the era of spintronics had begun. Both GMR and TMR effects have been thoroughly explored over the past 10 years and have found many applications.

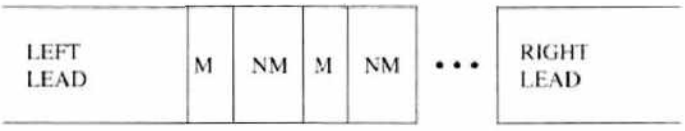


Figure 1. Schematic picture of a magnetic multilayer structure. M-magnetic layers; NM-nonmagnetic layers.

For example, the GMR effect is used to read information stored on the computer hard disk and GMR reading heads (spin valves) are now fitted to all modern computers. The closely related TMR effect can be used to store information in a magnetic random access memory (MRAM) that is currently being developed. It is now, therefore, well established that by altering the magnetic configuration of a magnetic multilayer we can influence the charge current flowing in it. However, it was recognized only much more recently [4] that, conversely, by passing a strong charge current one can alter the magnetic state of a magnetic multilayer. The structure in which this effect occurs consists of a left lead, a thick left magnetic layer (polarizing magnet), a nonmagnetic metallic spacer layer, a thin second magnet (switching magnet), and a semi-infinite right lead. It was proposed by Slonczewski [4] that the current passing through the left magnetic layer becomes spin polarized, and therefore, the flow of charge from the left magnetic layer to the right magnetic layer is accompanied by a flow of spin. The flow of the spin angular momentum is called spin current. The spin current is conserved in the nonmagnetic parts of the structure, but the spin of a carrier entering a ferromagnet can change its orientation, provided such a spin flip is compensated by the corresponding change of the total spin of the magnet so that the total angular momentum of the whole system is conserved. It follows that spin current can be absorbed by the ferromagnet. The rate of change of the total spin, given by the difference between the spin current entering a magnet and that leaving the magnet, is equal to the torque exerted on the magnetic moment of the ferromagnet. If the charge current, and the associated spin current, is strong enough the spin-transfer torque can cause total reversal of the magnetization. This effect is called current-induced switching of magnetization. Quite apart from being fundamentally interesting in its own right, current-induced switching of magnetization has important potential applications since it is envisaged that it could be used to write information in MRAM [5].

Even in the absence of charge current (i.e., when the magnetic multilayer is in equilibrium), there is a flow of spin current between two magnetic layers separated by a nonmagnetic metallic spacer layer provided the magnetic moments of the two magnets are not colinear. This leads to a static torque being exerted by one magnet on the other and the effect is known as oscillatory exchange coupling [6] between the two magnets since the sign of the torque depends in an oscillatory manner on the thickness of the nonmagnetic spacer. The oscillatory exchange coupling was discovered at the same time as the GMR effect, but a close link between this effect and current-induced switching of magnetization was not made until very recently.

It can be seen that the four effects that effectively define the new area of spintronics, that is, GMR, TMR, oscillatory exchange coupling, and current-induced switching of magnetization, all rely on the length of the magnetic structure in at least one direction being shorter than the spin diffusion length l_{sf}. Since this is for most metals of the order of several tens of nanometers spintronics can only operate in nanoscale devices.

There is one feature that distinguishes the GMR from TMR, oscillatory exchange coupling and current induced switching of magnetization. The GMR is the only effect that is observed in two qualitatively different geometries. In the first case, the current flows perpendicular to the layers (CPP) geometry. The second more usual geometry corresponds to the situation when the current flows in plane of the layers (CIP). Since the lateral dimensions of the layer structure in the direction parallel to the layers are macroscopic (often of the order of centimeters), the CIP GMR always takes place in the diffusive limit (the elastic mean free path is much shorter than the lateral dimensions of the layer structure). That means that semiclassical description of the CIP GMR based on the Boltzman equation is appropriate. On the other hand, CPP GMR, TMR, oscillatory exchange coupling and current induced switching of magnetization all take place in the CPP geometry. They are thus inherently quantum effects and, consequently, require quantum treatment of transport of charge and spin.

Yet another factor that distinguishes the CIP GMR from the rest is the role of impurities. As we shall see, the CIP GMR can only occur in the presence of impurities or in the presence of roughness at the interfaces between magnetic and nonmagnetic layers. The CPP GMR, TMR, oscillatory exchange coupling, and current-induced switching of magnetization can all operate in perfect samples (i.e., in the ballistic limit). In fact, we shall argue that, while

imperfections are essential for the existence of CIP GMR, they are detrimental to all the other aforementioned effects.

On the basis of this observation, we shall begin this review with the discussion of the CIP GMR using a semiclassical Boltzman equation approach. Then, prior to the discussion of the remaining effects that all take place in the CPP geometry, we introduce the general nonequilibrium Keldysh formalism that allows us to calculate in a unified way the charge and spin currents in every part of an arbitrary magnetic multilayer structure. It is most convenient to formulate such a quantum treatment of CPP transport in terms of local one electron Green functions. We, therefore, describe briefly how the local Green functions are calculated in the layer geometry and discuss their general properties. Finally, the Keldysh formalism will be used to discuss and correlate the existing theories of the CPP GMR, TMR, oscillatory exchange coupling, and current-induced switching of magnetization.

2. GIANT MAGNETORESISTANCE

The era of spintronics begun with the discovery [7, 8] that the resistance of a multilayer consisting of a sequence of thin magnetic layers separated by equally thin nonmagnetic metallic layers is low when the magnetizations of the neighboring magnetic layers are parallel [Fig. 2(a)] but becomes much higher when they are ordered antiparallel [Fig. 2(b)]. The most commonly used combinations of magnetic and nonmagnetic layers are cobalt-copper and iron-chromium, but multilayers based on permalloy as the magnetic component are also frequently used. The second key ingredient was the discovery by Stuart Parkin [9] that the relative orientation of the magnetic moments of two neighboring magnetic layers depends on the thickness of the intervening nonmagnetic spacer layer. In fact, he found that the orientation of the magnetic moment of the magnetic layers oscillates between parallel (ferromagnetic) and antiparallel (antiferromagnetic) as a function of the nonmagnetic layer thickness. This phenomenon is referred to as an oscillatory exchange coupling and will be discussed in Section 4.

Assuming that the thickness of the nonmagnetic spacer layer is chosen so that the spontaneous orientation of the adjacent magnetic layers is antiparallel, change of the magnetic configuration from antiferromagnetic to ferromagnetic can be effected by an applied magnetic field. The relative change of the resistance can be larger than 200% and that is the reason why the effect is called *giant magnetoresistance* (GMR) The "optimistic" magnetoresistance ratio, most commonly used, is defined by

$$\frac{\Delta R}{R} = \frac{R^{\uparrow\downarrow} - R^{\uparrow\uparrow}}{R^{\uparrow\uparrow}} \tag{1}$$

where $R^{\uparrow\downarrow}$ and $R^{\uparrow\uparrow}$ are the resistances of the magnetic multilayer in its antiparallel (zero field) and parallel (saturating field) magnetic configurations. The dependence of the GMR ratio of a Fe/Cr multilayer on the thickness of the nonmagnetic chromium layer, observed by Parkin in his original experiment [9], is reproduced in Fig. 3. Oscillations of the GMR as a function of chromium thickness occur because the magnetoresistance effect is measurable only for those thicknesses of chromium for which the interlayer exchange coupling aligns the magnetic moments of all the iron layers antiparallel.

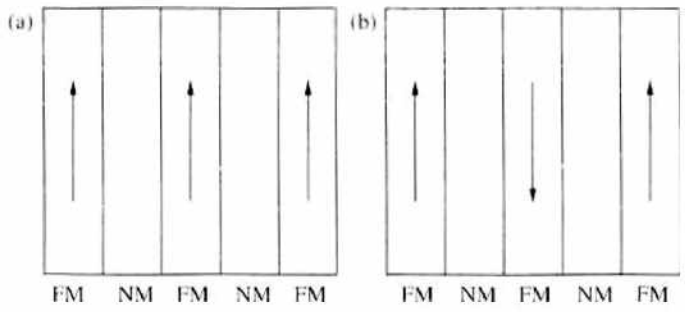


Figure 2. Ferromagnetic (a) and antiferromagnetic (b) configurations of a magnetic multilayer.

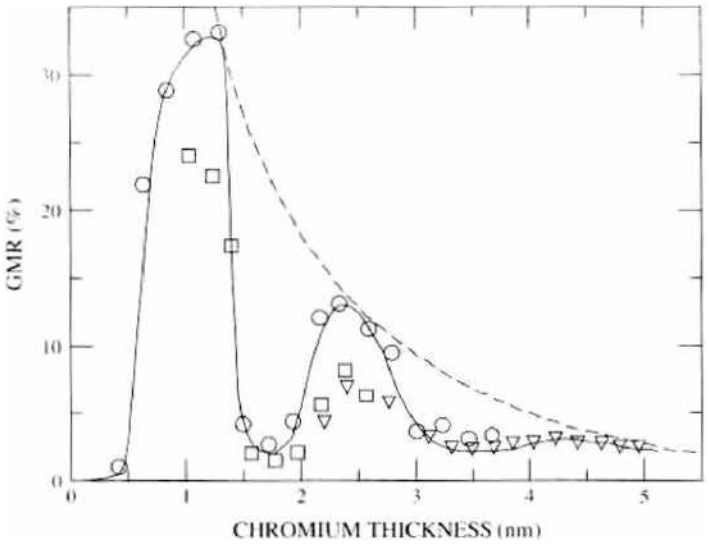


Figure 3. Dependence of the GMR ratio of an Fe/Cr multilayer on Cr thickness.

A typical magnetoresistance curve [10] for an Fe/Cr multilayer of 50 repeats of an iron layer 0.45 nm thick and a chromium layer 1.2 nm thick is shown in Fig. 4 for two temperatures $T = 1.5$ K and 300 K. The gradual decrease of the resistance with increasing magnetic field, seen in Fig. 4, occurs because the magnetic field, which tends to align the moments of the magnetic layers parallel has to overcome the oscillatory exchange coupling which favors the antiparallel arrangement (for this particular thickness of chromium). Complete alignment is achieved only in a saturating field equal in magnitude to the exchange field.

We shall now clarify the physical origin of the GMR using a simple resistor model. There are two principal geometries of the GMR effect. They are shown schematically in Fig. 5. In the first case [Fig. 5(a)], the current flows perpendicular to the layers (CPP geometry). Figure 5(b) illustrates the more usual geometry when the current flows in plane of the layers (CIP). As we shall see, the CPP geometry is easier to treat theoretically but much more difficult to realize experimentally. This is because the transverse dimensions of typical multilayers are of the order of squared centimeters, whereas their thickness is only of the order

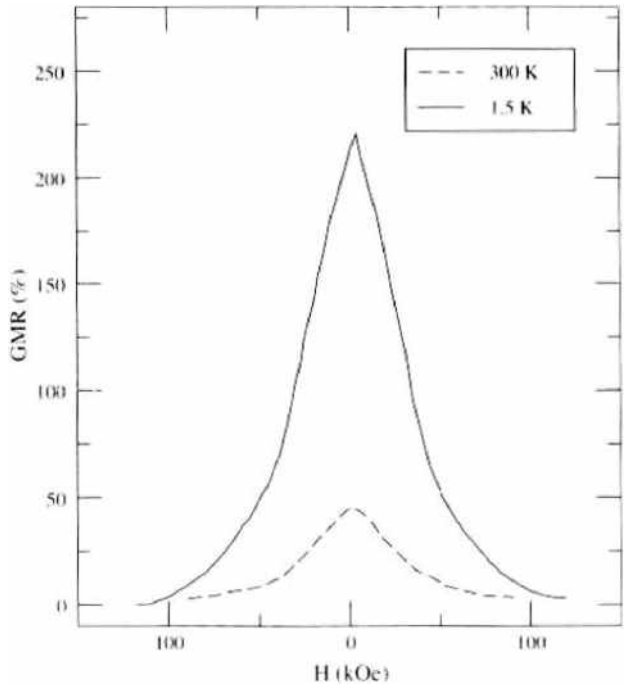


Figure 4. Magnetoresistance curve of an Fe/Cr multilayer

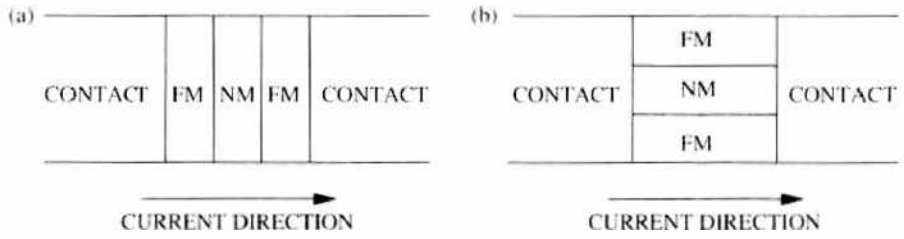


Figure 5. Current perpendicular to plane (a) and current in plane (b) GMR geometries.

of a few nanometers. It follows that the resistance of the multilayer in the CPP geometry is extremely low and, therefore, very sophisticated experimental techniques [11] are required to measure accurately the very small voltage drop across the sample. However, since the underlying physical mechanism is the same in the CPP and CIP geometries, there is no need in the first introductory account to distinguish between them.

Consider a trilayer with two magnetic layers separated by a nonmagnetic metallic spacer layer. As already discussed in Section 1, if the total thickness of the trilayer is smaller than the spin diffusion length, we can assume that electric current in the trilayer flows in two channels, one corresponding to carriers with spin projection ↑ and the other to carriers with spin projection ↓ [1]. Since the ↑ and ↓ spin channels are independent (spin is conserved), they can be regarded as two wires connected in parallel.

The second essential ingredient is that carriers with spin projections parallel and antiparallel to the magnetization of the ferromagnetic layer are scattered at different rates when they enter the ferromagnet. This is called spin-dependent scattering. Let us assume that carriers with spin antiparallel to the magnetization are scattered more strongly. We shall see later that this is the case for the Co/Cu combination, but the opposite is true for the Fe/Cr system. The GMR effect in a trilayer can be now explained qualitatively using a simple resistor model shown in Fig. 6. In the ferromagnetic configuration of the trilayer, carriers with ↑ spin are weakly scattered both in the first and second ferromagnet, whereas the ↓ spin carriers are strongly scattered in both ferromagnetic layers. This is modeled by two small resistors in the ↑ spin channel and by two large resistors in the ↓ spin channel in the equivalent resistor network shown in Fig. 6(a). Since the ↓ and ↑ spin channels are connected in parallel, the total resistance of the trilayer in its ferromagnetic configuration is determined by the low-resistance ↑ spin channel, which shorts the high-resistance ↓ spin channel. It follows that the total resistance of the trilayer in its ferromagnetic configuration is low. However, ↓ spin carriers in the antiferromagnetic configuration are strongly scattered in the first ferromagnetic layer but weakly scattered in the second ferromagnetic layer. The ↑ spin carriers are weakly scattered in the first ferromagnetic layer and strongly scattered in the second. This

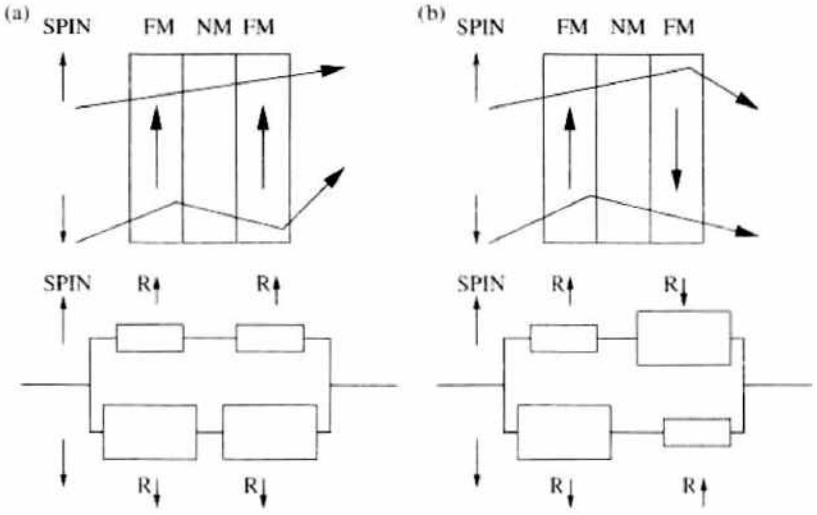


Figure 6. Resistor model of GMR.

is modeled in Fig. 6(b) by one large and one small resistor in each spin channel. There is no shorting now and the total resistance in the antiferromagnetic configuration is, therefore, much higher that in the ferromagnetic configuration.

This simple physical model of the GMR effect is believed to be qualitatively correct but needs to be converted into a quantitative theory that can explain the differences between the CIP and CPP geometries, the observed dependence of the GMR on the layer thicknesses and also the material dependence of the effect. Moreover, we need to understand the microscopic origin of the spin-dependent scattering and clarify under what conditions the ↑ and ↓ spin channels in magnetic multilayers can be treated as independent.

We begin with the discussion of different types of scattering carriers may experience in magnetic multilayers. In the calculation of the resistance, we are mainly concerned with elastic (energy conserving) scattering. In each scattering act, only the direction of propagation of carriers changes. It is essential to distinguish between spin-dependent scattering, which causes the GMR, and spin-flip scattering, which is detrimental to the GMR. The two types of scattering are illustrated in Fig. 7. In the case of spin-dependent scattering, the orientation of the carrier spin is conserved in each scattering event, but the probabilities of scattering for carriers with ↑ and ↓ spin projections are different. However, when a carrier undergoes a spin-flip scattering, its spin orientation changes from ↑ ($s_z = \hbar/2$) to ↓ ($s_z = -\hbar/2$) or vice versa and, at the same time, the spin of the scattering centre changes by $\Delta = \hbar$ so that the total spin is conserved.

There are several sources of spin-flip scattering. When magnetic multilayers are prepared, some of the magnetic atoms may enter the nonmagnetic spacer layer to form magnetic impurities. When a carrier is scattered off a magnetic impurity the spins of the carrier and that of the impurity can interchange provided the impurity spin is free to rotate. This is the case when the impurity spin is not strongly coupled to the spins of the ferromagnetic layers (i.e., when the impurity is not near the ferromagnet/spacer interface).

Carriers can also be scattered from spin waves in the ferromagnetic layers. Spin waves are quasiparticles with spin one, and, therefore, creation (annihilation) of a spin wave in a collision with a carrier leads to a flip of the carrier spin. Since creation (annihilation) of spin waves involves the spin wave energy, this is an inelastic process that is only important at elevated temperatures.

Finally, when impurities with a strong spin-orbit interaction, such as gold, are present in the multilayer, the spin of a carrier incident on such an impurity may be reversed due to the spin-orbit interaction.

Since all these processes mix ↑ and ↓ spin channels, they are detrimental to the GMR. In what follows, we shall assume that spin flip scattering is weak so that no mixing of the ↑ and ↓ spin channels takes place. This assumption may break down for relatively thick multilayers and the implications of spin-flip scattering for GMR are discussed in detail in Ref. [12].

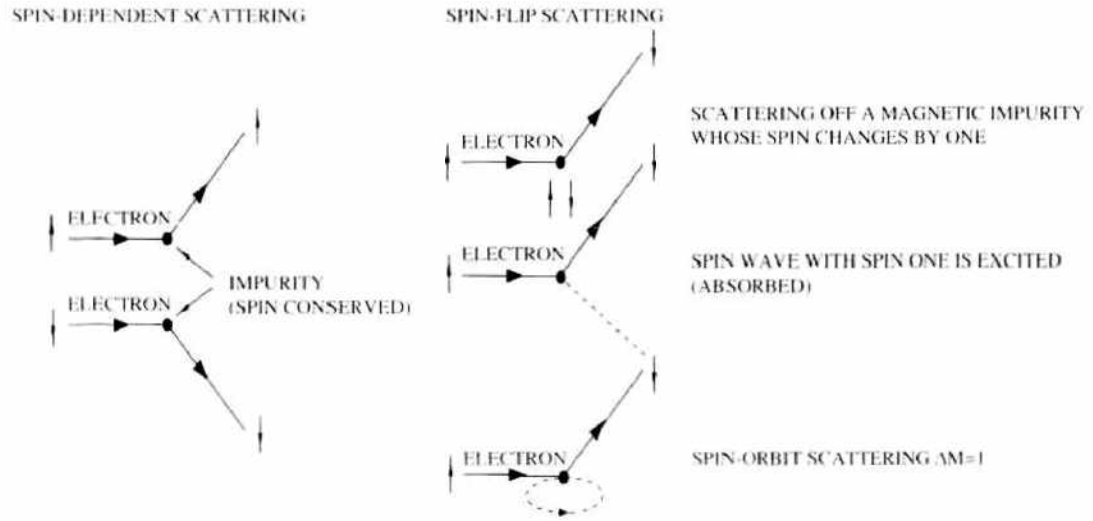


Figure 7. Different types of scattering in magnetic multilayers.

We now turn to the spin-dependent scattering that conserves the carrier spin. The key feature here is that carriers with different spin orientations ($\uparrow$, $\downarrow$) are scattered at different rates when they enter the ferromagnetic layers. Given that carriers obey the Pauli exclusion principle, a carrier can be scattered from an impurity only to quantum states that are not occupied by other carriers. At zero (low) temperatures, all the states with energies E below the Fermi energy E_F are occupied and those with $E > E_F$ are empty. Since scattering from impurities is elastic, carriers at the Fermi level (which carry the current) can be scattered only to states in the immediate vicinity of the Fermi level. It follows that the scattering probability is proportional to the number of states available for scattering at E_F (i.e., to the density of states $D(E_F)$). The densities of states of copper, cobalt, and iron for $\uparrow$ (upper panel) and $\downarrow$ (lower panel) spin orientations are shown in Fig. 8. The Fermi level in copper (and other noble metals) intersects only the conduction band whose density of states $D(E_F)$ is low. It follows that the scattering probability in copper is also low, which explains why copper is a very good conductor. However, the d band in transition metals is only partially occupied; therefore, the Fermi level in these metals intersects not only the conduction but also the d bands. Moreover, since the atomic wave functions of d levels are more localized than those of the outer s levels, they overlap much less, which means that the d band is narrow and the corresponding density of states is high. This opens up a new very effective channel for scattering of carriers into the d band. This new scattering mechanism (Mott scattering [13]) explains why all transition metals are poor conductors compared with noble metals.

In the case of magnetic transition metals, we need to consider an additional crucial factor, namely, that d bands for $\uparrow$ and $\downarrow$ spin carriers are split by the exchange interaction. This amounts to an almost rigid relative shift of the $\uparrow$ and $\downarrow$ spin d bands, which is clearly seen for cobalt and iron in Fig. 8. The $\uparrow$ spin d band in cobalt is full, which means that $D^{\uparrow}(E_F)$ is as low as in copper, but the Fermi level in the $\downarrow$ spin band lies in the d band; therefore, $D^{\downarrow}(E_F)$ is much higher than $D^{\uparrow}(E_F)$. The situation for iron is somewhat different in that the density of states at E_F is higher for $\uparrow$ spin carriers than for $\downarrow$ spin carriers. Also the spin asymmetry in the density of states is not so large for iron as for cobalt. However, in either case, the spin asymmetry of the density of states results in different scattering rates for $\uparrow$ and $\downarrow$ spin carriers, that is, spin-dependent scattering. It should be noted that this mechanism operates even if the scattering potential itself is independent of the spin i.e., nonmagnetic impurities, vacancies or stacking faults in a ferromagnetic metal all lead to spin-dependent scattering. Because the Mott scattering mechanism is effective in bulk ferromagnetic metals, we shall refer to it as bulk spin dependent scattering.

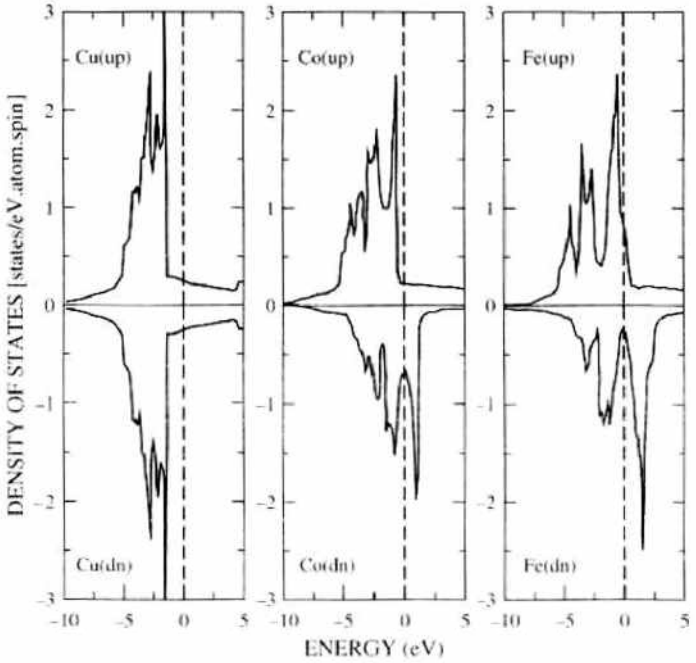


Figure 8. Densities of states of copper, cobalt, and iron. Broken line denotes the position of the Fermi level.

The relative shift of ↑ and ↓ spin bands is simply a consequence of the fact that the potentials seen by ↑ and ↓ carriers in a ferromagnetic metal are different because of the exchange interaction. This provides another mechanism of spin-dependent scattering, which is specific to multilayers. In an infinite ferromagnet, this effect does not, of course, lead to any spin asymmetry of the resistance since as long as the potentials seen by ↑ and ↓ carriers are periodic they do not result in any dissipation of the carrier momentum. However, carriers in a multilayer entering the ferromagnet from the nonmagnetic spacer see a spin dependent potential barrier, which reflects differently carriers with ↑ and ↓ spin orientations. In the CPP geometry, even perfect interfaces thus result in spin-dependent scattering [14]. In the CIP geometry, carriers propagate mainly along interfaces and, therefore, this mechanism is effective only if the interfaces are rough (intermixing of magnetic and nonmagnetic atoms).

As opposed to the bulk Mott mechanism discussed earlier, spin-dependent scattering due to spin dependence of the scattering potentials takes place only at the ferromagnet/nonmagnet interface and is, therefore, called interfacial spin-dependent scattering. To gain better understanding of the interfacial spin-dependent scattering, it is instructive to examine the band structures of the most common combinations of magnetic and nonmagnetic metals used in GMR multilayers. These are Co/Cu and Fe/Cr, and their band structures in the [001] direction are shown in Figs. 9 and 10. It can be seen from Fig. 9 that there is a very good match between the bands of Cu and the ↑ (majority) spin band of Co. One can, therefore, conclude that ↑ spin carriers crossing the Cu/Co interface experience only weak scattering, and that remains true even if Cu and Co atoms are intermixed at the interface. On the other hand, there is a large mismatch between the Cu and Co bands for the ↓ (minority) spin carriers reflecting a large difference between the atomic potentials of the two elements. It follows that ↓ spin carriers are strongly scattered at the Cu/Co interfaces. On the other hand, matching of the Fe and Cr bands is almost perfect for ↓ spin carriers but poor for the ↑ spin carriers. The spin asymmetry of scattering at the Fe/Cr interface has, therefore, a sign opposite to that for Co/Cu interface.

The discussion of spin-dependent scattering based on the mismatch of bands of the magnetic and nonmagnetic components of magnetic multilayers allows us also to understand which combinations of magnetic and nonmagnetic metals should lead to optimum GMR. One clearly seeks as good a match as possible between the bands of the magnetic layers and those of the spacer layer in one spin channel and as large as possible mismatch in the other spin channel. It is clear from Figs. 9 and 10 that Co/Cu and Fe/Cr fulfil very well these requirements.

At a lowest level of the mathematical modeling of GMR, we only need to incorporate the effect of spin-dependent scattering into classical Boltzman equation to determine the resistances of a magnetic multilayer in its ferromagnetic and antiferromagnetic configurations.

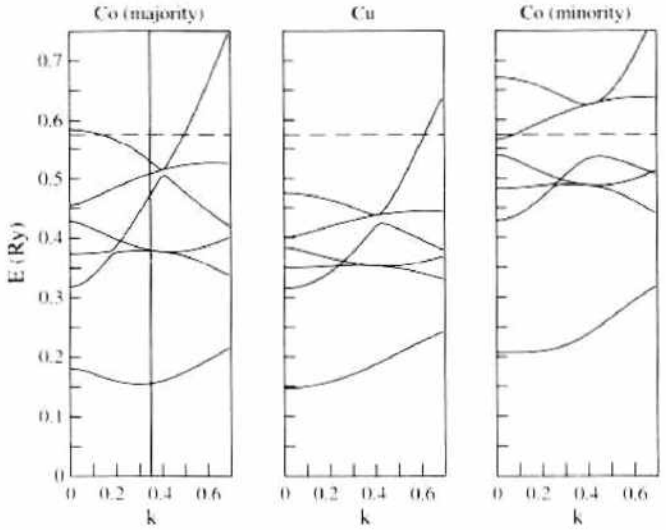


Figure 9. Band structures of cobalt and copper along the [001] direction in the vicinity of one of the Cu Fermi surface necks. Broken line denotes the position of the Fermi level.

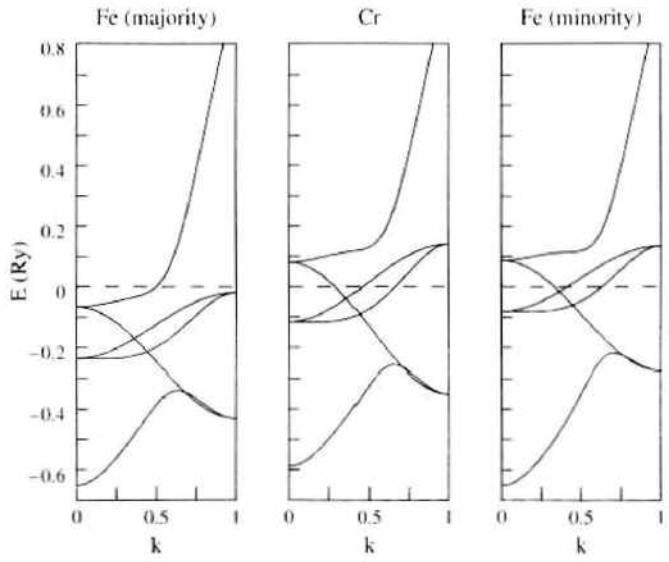


Figure 10. Band structures of iron and chromium in the [001] direction. Broken line denotes the position of the Fermi level.

This was first done by Camley and Barnas [15] (see also Ref. [12]). The results that follow from the classical Boltzman equations in the limit of a long spin-diffusion length can be obtained more simply by making the resistor model of the GMR quantitative. Since the resistor model is widely used, particularly in the interpretation of experimental results in the CPP geometry, we shall describe the arguments used in its derivation [16].

It follows from the Boltzman equation in the relaxation-time approximation [17] that the resistivity ρ of a bulk metal is given by

$$\rho = \frac{m}{ne^2\tau} \tag{2}$$

where e is the electron charge, m is the mass, and n the density of carriers, and τ is the elastic mean free time. The mean free time is inversely proportional to the scattering probability. The scattering probability is, in turn, determined by two factors. The first is the strength of the scattering potential and the second the density of states at E_F available for scattering. As already discussed, the first factor leads to interfacial spin-dependent scattering in magnetic multilayers and the second is the Mott mechanism which results in spin-dependent bulk scattering. We first consider the effect of the bulk spin-dependent scattering. If follows from the above arguments that we can introduce a spin-dependent resistivity for each ferromagnetic metal by

$$\rho_{\mathrm{FM}}^{\uparrow} = 2\frac{\rho_{\mathrm{FM}}}{1+\beta}; \quad \rho_{\mathrm{FM}}^{\downarrow} = 2\frac{\rho_{\mathrm{FM}}}{1-\beta} \tag{3}$$

where ρ_{FM} is the total resistivity of the bulk ferromagnetic metal, $1/\rho_{\mathrm{FM}} = 1/\rho_{\mathrm{FM}}^{\uparrow} + 1/\rho_{\mathrm{FM}}^{\downarrow}$ (assuming that the two spin channels remain independent). The parameter β we have introduced will be referred to as the bulk-scattering asymmetry. We shall treat β as a phenomenological parameter, but we expect from the discussion of the Mott scattering mechanism that $\rho^{\uparrow}/\rho^{\downarrow} \approx D^{\uparrow}(E_F)/D^{\downarrow}(E_F)$. In particular, it follows from Fig. 8 that $\beta < 0$ for cobalt and $\beta > 0$ for iron.

Similarly, we can introduce an interfacial scattering asymmetry assuming that there is a thin interfacial layer whose resistance ρ_{F-N}^{σ} is spin dependent due to the presence of a spin-dependent potential barrier at the ferromagnet/nonmagnet interface. We, therefore, define an interfacial asymmetry parameter γ by

$$\rho_{F-N}^{\uparrow} = 2\frac{\rho_{F-N}}{1+\gamma}; \quad \rho_{F-N}^{\downarrow} = 2\frac{\rho_{F-N}}{1-\gamma} \tag{4}$$

where ρ_{F-N} is the total resistivity of the interfacial layer. We can again deduce from Figs. 9 and 10 that $\gamma < 0$ for cobalt and $\gamma > 0$ for iron. However, the actual magnitude of γ is

difficult to determine microscopically since it depends not only on the difference between the potentials seen by ↑ and ↓ spin carriers at the interface but also on the interfacial roughness and the thickness of the interfacial layer for which ρ^{σ}_{F-N} is introduced. We shall, therefore, treat γ again as a phenomenological parameter.

We are now ready to calculate the GMR. The calculation will be described for bulk spin-dependent scattering and CIP geometry. It is straightforward to include interfacial spin-dependent scattering and derive the resistor model in the CPP geometry.

Consider a periodic superlattice of alternating nonmagnetic and magnetic layers with spin-dependent scattering in the bulk of the ferromagnetic layers. Because the whole superlattice is made up of identical building blocks, superlattice unit cells, it is sufficient to calculate the resistances of a unit cell. In the antiferromagnetic configuration, the magnetic layers with antiparallel magnetizations are inequivalent; therefore, the basic building block we have to consider (magnetic cell) consists of two magnetic layers containing M atomic planes each and two nonmagnetic layers of N atomic planes each. The geometry for which the GMR is going to be calculated and the definition of a magnetic cell are illustrated in Fig. 11. It follows from Eq. (4) that a carrier of a given spin traveling in a superlattice sees regions of different local resistivities. The resistivity is high in those regions where there is a high density of states at E_F available for scattering. There are, therefore, three different local resistivities in the superlattice unit cell: the resistivity of the nonmagnetic spacer layer ρ_{NM}, which is the same for both spin orientations, and the high ρ^{H}_{FM} and low ρ^{L}_{FM} resistivities for the two different spin orientations in the ferromagnet. The low resistivity of the ferromagnet satisfies $\rho_{\mathrm{NM}} \approx \rho^{L}_{\mathrm{FM}}$ both for the Co/Cu and Fe/Cr systems. The distribution of such regions in a superlattice magnetic cell in its ferromagnetic and antiferromagnetic configurations is shown in Fig. 12. It is clear from Fig. 12 that the unit cell of a magnetic superlattice is equivalent to a system of eight resistors, with four resistors in each spin channel. To determine the magnetoresistance, we first need a rule for adding up the four resistors in the same spin channel. Once the total resistances in both spin channels are known, they can be simply added as resistors in parallel to give the total resistance of the magnetic unit cell. This needs to be done for the ferromagnetic (↑↑) and antiferromagnetic (↑↓) configurations. Following this prescription, we find that the resistances $R_{\uparrow\uparrow}$ and $R_{\uparrow\downarrow}$ are given by

$$\frac{1}{R_{\uparrow\uparrow}} = \left(\frac{1}{R_{\uparrow}} + \frac{1}{R_{\downarrow}}\right)_{\uparrow\uparrow}; \quad \frac{1}{R_{\uparrow\downarrow}} = \left(\frac{1}{R_{\uparrow}} + \frac{1}{R_{\downarrow}}\right)_{\uparrow\downarrow} \tag{5}$$

where R_{σ} is the resistance of the unit cell in a spin channel σ.

It is now necessary to determine the rules for adding up the four resistors in the same spin channel. It is clear from Fig. 12 that for the ferromagnetic configuration the problem reduces to the calculation of the resistance of a two-component superlattice with alternating regions of thicknesses a and b having resistances ρ^{a} and ρ^{b}. For the antiferromagnetic configuration, a four-component superlattice needs to be considered. To clarify the underlying physics, it is sufficient to investigate the two-component superlattice. Because the current flows in the direction of the layers forming the superlattice, one might be tempted to conclude that the resistances of the layers should always be added up as for resistors connected in parallel. However, it is easy to demonstrate that, in general, that would be incorrect.

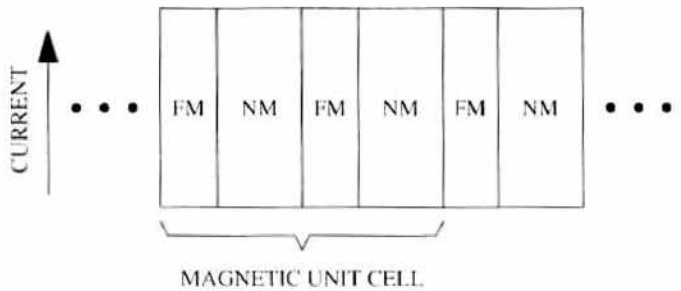


Figure 11. Magnetic superlattice

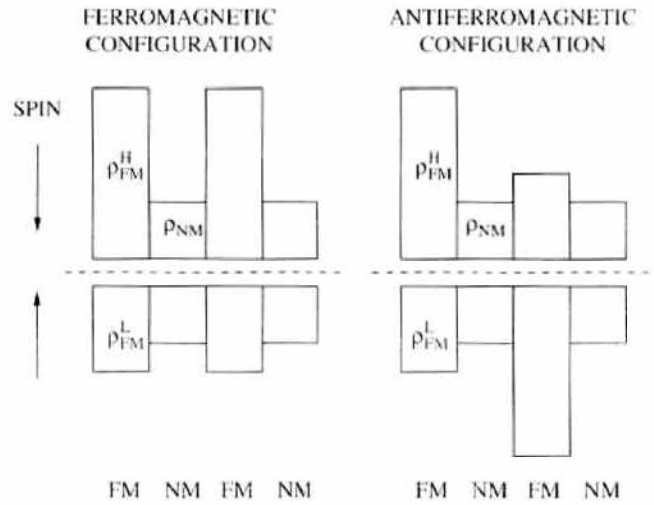


Figure 12. Distribution of local resistivities in a magnetic unit cell.

Let us assume for simplicity that the resistivity ρ^a is higher than ρ^b because there is a higher density of scatterers in the layer a. This is illustrated in Fig. 13. Consider first the simplest case when there is a "partition" between the layers a and b, which prevents carriers crossing the a/b interface. Such a system is clearly equivalent to two *independent* resistors because carriers remain confined to their respective layers. In this case, the above argument applies; that is, the two layers behave as ordinary resistors in parallel. However, there are no impenetrable partitions between the neighboring regions in a superlattice. Carriers can cross easily the interface and undergo scattering in both layers. It follows that the two layers cannot be regarded as independent, and, in general, the simple rules for a conventional network of resistors no longer apply. The reader might conclude that we can get no further without a detailed microscopic calculation. Fortunately this is not so, and there are two physically important limits in which the total resistance of a superlattice can be easily evaluated.

Case A: The mean free path in each layer of a superlattice is much shorter than the thickness of the layer. Because the mean free path is so short, very few carriers starting in one layer reach the neighboring layer. It follows that carriers from different layers do not "mix" and flow in their respective separate resistor channels as if the layers were separated by "partitions." All resistors then behave as resistors in a conventional resistor network and should be added in parallel. Inspection of Fig. 12 shows that there are exactly the same number of resistors of each type in the ferromagnetic and antiferromagnetic configurations, which means that $R_{\uparrow\uparrow} = R_{\uparrow\downarrow}$, and there is no magnetoresistance in this limit.

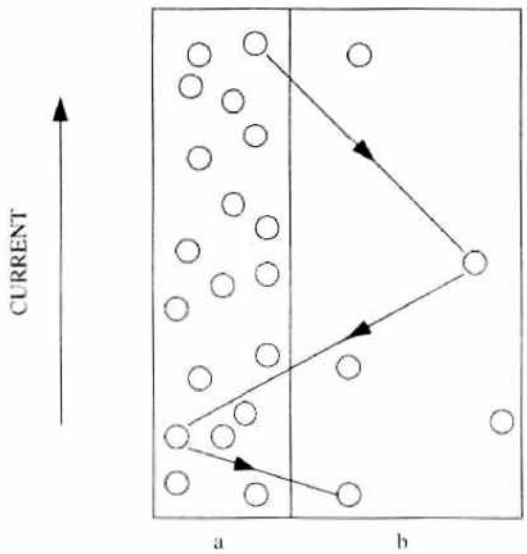


Figure 13. Magnetic superlattice.

Case B: The mean free path in each layer is much longer than the thickness of the layer. For a metallic superlattice with a small number of atomic planes in each layer, we are always close to this limit since typical mean free paths in metals are of the order of tens or even hundreds of interatomic distances. It is, therefore, the limit that is applicable to magnetic superlattices exhibiting the giant magnetoresistance.

Carriers now sample equally layers with low and high resistivity and, therefore, experience an average resistivity. For a two-component superlattice this is given by

$$\bar{\rho} = \frac{a\rho^a + b\rho^b}{a + b} \tag{6}$$

The generalization of Eq. (6) to a four-component superlattice is obvious.

The results for cases A and B obtained here from simple physical considerations can be derived as limits of a more general approach based on the Boltzmann equation [16]. In fact, such a microscopic calculation shows that the limit in which the simple averaging Eq. (6) applies is reached very rapidly and we are close to this limit already for a mean free path comparable with the thickness of the superlattice unit cell.

Applying Eq. (6) to ferromagnetic and antiferromagnetic unit cells shown in Fig. 12, it is easy to show that

$$\left(\frac{1}{R}\right)_{\uparrow\uparrow} = 2(M+N)^2\left(\frac{1}{M\rho^I_{\mathrm{FM}} + N\rho_{\mathrm{NM}}} + \frac{1}{M\rho^{II}_{\mathrm{FM}} + N\rho_{\mathrm{NM}}}\right) \tag{7}$$

$$\left(\frac{1}{R}\right)_{\uparrow\downarrow} = \frac{8(M+N)^2}{M\rho^I_{\mathrm{FM}} + M\rho^{II}_{\mathrm{FM}} + 2N\rho_{\mathrm{NM}}} \tag{8}$$

Finally, using Eqs. (7) and (8), we find that the optimistic magnetoresistance ratio is given by

$$\frac{\Delta R}{R} = \frac{(1-\beta)^2}{4(1 + N/M\mu)(\beta + N/M\mu)} \tag{9}$$

where β is the bulk scattering asymmetry and $\mu = \rho^I_{\mathrm{FM}}/\rho_{\mathrm{NM}}$.

It is now easy to pinpoint the main factors that determine the GMR. Clearly, $\Delta R/R$ is a function of two variables, β and $M\mu/N$. The most important requirement for large GMR is that the spin asymmetry ratio β should be large. For a given β, the GMR increases with increasing $M\mu/N$ but saturates for a large value of this parameter. As a function of the spacer layer thickness N, the GMR decreases monotonically and falls off as $1/N^2$ for large N, which is as observed (see the broken line in Fig. 3). This can be viewed as "shunting" of the cooperative effect of the magnetic layers by an "inactive" spacer layer.

Experiments show (see, e.g., Ref. [18]) that CIP GMR also decreases with increasing thickness of the magnetic layers. This is not reproduced by Eq. (9). The main reason for this failure is our neglect of interfacial spin dependent scattering. To illustrate its effect, we shall adopt the other extreme point of view, that is, assume that interfacial spin-dependent scattering is so strong that bulk scattering can be neglected ($\beta = 0$). We shall further assume that there are I interfacial atomic planes with ρ^{σ}_{F-N} defined by Eq. (4). Finally, we shall make a simplifying assumption $\rho_{\mathrm{NM}} = \rho^I_{\mathrm{FM}}$ (a good approximation for Co/Cu and Fe/Cr multilayers). It is then easy to show that

$$\frac{\Delta R}{R} = \frac{(1-\gamma)^2}{4[1 + (N+M)/2I\nu][\gamma + (N+M)/2I\nu]} \tag{10}$$

where γ is the interfacial scattering asymmetry defined in Eq. (4) and $\nu = \rho^I_{F-N}/\rho_{\mathrm{NM}}$. The GMR now decreases both with increasing thicknesses N, M of the spacer and the ferromagnet. The physical interpretation is that the magnetic layer first grows as a rough interface with strong spin-dependent scattering and then turns into an "inactive" shunting layer. It might seem that the model with dominant interfacial scattering explains better the dependence of the GMR on the ferromagnet thickness. However, numerical solution of the Boltzmann equation with dominant bulk spin dependent scattering leads also to a GMR which decreases

with increasing thickness of the ferromagnetic layers, in good agreement with experiment [18]. One must, therefore, conclude that by analyzing experimental data in the CIP geometry, it is not possible to determine reliably the relative importance of bulk and interface spin dependent scattering. The situation is much clearer in the CPP geometry which will be now briefly discussed.

In applying the simple resistor model to the CPP GMR, we need to make an assumption that the mean free path for spin-flip scattering is longer than the total length of the multilayer in the direction of the current. This is necessary for $\uparrow$ and $\downarrow$ spin channels to remain independent so that we can add up their total resistances $R_\uparrow$ and $R_\downarrow$ in parallel. We shall again introduce bulk scattering asymmetry using Eq. (3), but it is more convenient in the CPP geometry to characterise interfacial scattering by the total resistances of an interface for $\uparrow$ and $\downarrow$ spin channels; that is, we define high R^H_{F-N} and low R^L_{F-N} interfacial resistances. Naturally, they are related to the total interfacial resistance R_{F-N} and to the interfacial scattering asymmetry γ via Eq. (4). Finally, we need to decide how to add up all the resistors in the same spin channel. This is simple in the CPP geometry since electrons move in the direction perpendicular to the layers and, therefore, sample individual layers one by one. All the layers thus behave as conventional resistors connected in series. We shall again consider a superlattice having M atomic planes in each ferromagnetic layer and N planes in each nonmagnetic layer. To calculate the total resistances $R_{\uparrow\uparrow}$ and $R_{\uparrow\downarrow}$ of the superlattice in its ferromagnetic and antiferromagnetic configurations, we need to introduce also the total number N_{MC} of magnetic unit cells. It is then straightforward to show that

$$(AR)_{\uparrow\downarrow} = \frac{N_{MC}}{2}[M(\rho^L_{\mathrm{FM}} + \rho^H_{\mathrm{FM}}) + 2\rho_{\mathrm{NM}}N + 2A(R^H_{F-N} + R^L_{F-N})] \tag{11}$$

where A is the cross-section area of the superlattice and the quantities ρ^L_{FM}, ρ^H_{FM} have already been introduced in the CIP geometry.

Equation (11) can be used to test the validity of the series resistor model. It implies that the total resistance in the antiferromagnetic configuration increases linearly with the thickness of the superlattice (N_{MC}). It is found [19] that Eq. (11) is well obeyed for Co/Cu and Co/Ag multilayers.

One can easily obtain also the total resistance $R_{\uparrow\uparrow}$ of a magnetic superlattice in its ferromagnetic configuration and, hence, the GMR ratio $\Delta R/R$. However, it turns out that it is more useful to examine a closely related quantity

$$[(R_{\uparrow\downarrow} - R_{\uparrow\uparrow})R_{\uparrow\downarrow}]^{1/2} = \frac{N_{MC}}{2}(\beta\rho^*_{\mathrm{FM}}M + 2\gamma AR^*_{F-N}) \tag{12}$$

where $\rho^*_{\mathrm{FM}} = \rho_{\mathrm{FM}}/(1-\beta^2)$, $R^*_{F-N} = R_{F-N}/(1-\gamma^2)$ and β, γ are the bulk and interfacial scattering asymmetries. If we plot the left-hand side of Eq. (12) as a function of the thickness M of the ferromagnetic layer keeping N_{MC} fixed, we obtain a straight line with a slope $N_{MC}\beta\rho^*_{\mathrm{FM}}$ and an intercept $N_{MC}\gamma AR^*_{F-N}$. The slope is thus determined entirely by bulk spin-dependent scattering and the intercept by interfacial spin dependent scattering. It follows that the two types of scattering can be separated in the CPP geometry. An analysis of the CPP GMR experiments for Co/Ag and Co/Cu superlattices [14, 20] based on Eq. (12) shows that $\beta \approx 0.5$ and $\gamma \approx 0.6$–0.8 for both these systems. Bulk and interface scattering are, therefore, comparable.

The equivalent resistor theory of the GMR provides a correct semiquantitative explanation of the effect and is particularly useful for analyzing experiments in the CPP geometry [14]. However, its main shortcomings are that the spin scattering asymmetries β and γ are introduced as phenomenological parameters and the differences between the band structures of the ferromagnetic and non-magnetic layers are ignored. As we shall see, this second factor is particularly important for the CPP GMR and full quantum treatment is necessary to include the contribution to the resistance of multiple reflections of carriers from all the interfaces between magnetic and nonmagnetic layers.

We now briefly discuss how the Boltzmann equation approach, on which the resistor network model is based, can be improved. Early attempts at a more microscopic description

of the CIP GMR were based on the Kubo formula [21] but the electronic band structure of the magnetic and nonmagnetic layers was approximated by a simple parabolic band common to the whole multilayer and spin dependent scattering was introduced phenomenologically. When these two assumptions are made, the results obtained from the Boltzmann and Kubo formulations are essentially equivalent. A more recent refinement is to incorporate in the Boltzmann equation a fully realistic band structure [22]. The main advantage of this approach is that the spin dependent scattering is introduced from first principles and the dependence of the GMR on different magnet/non-magnet combinations can be thus discussed. Another interesting recent development is a calculation [23] of the GMR assuming the Mott scattering mechanism (bulk spin-dependent scattering) but using the Boltzmann equation in the layer geometry combined with a realistic tight-binding band structure. This approach provides a microscopic underpinning of the phenomenological resistor model described here.

However, despite all the attempts at a more realistic modeling of the CIP GMR, a fully predictive theory is still not available. The principal reason for this failure is that the CIP GMR is determined largely by interfacial roughness, which is very difficult to model realistically. Moreover since the interfacial roughness depends on experimental growth methods, annealing, and other factors, its precise nature is usually not known and that makes meaningful comparison between theory and experiment virtually impossible. We shall see that a realistic modeling of the CPP GMR is much easier but, as has already been discussed, measurements in the CPP geometry are much more difficult than in the CIP geometry.

3. NONEQUILIBRIUM KELDYSH FORMALISM

In this section, we show how to calculate the local charge and spin currents flowing in the direction perpendicular to the layers of an arbitrary magnetic layer structure. We shall assume that the magnetic layer structure is sandwiched between two reservoirs with a bias V_b applied between them to produce a spin-polarized current. The structure we consider is shown schematically in Fig. 14. Typically, the two reservoirs will be semi-infinite nonmagnetic leads and the magnetic structure consists of a left magnet separated from the right magnet by a nonmagnetic spacer layer. The spacer layer may be either a nonmagnetic metal or an insulator. Although this is the most common situation, an arbitrary finite number of other layers (magnetic or nonmagnetic) can be easily incorporated into the structure. The broken line in Fig. 14 represents a cleavage plane separating the system into two independent parts so that charge carriers cannot move between the two surface planes labeled $n-1$ and n. It will be seen that our ability to cleave the whole system in this way is essential for the implementation of the Keldysh formalism on which our method is based. This can be easily done within a tight-binding parameterization of the band structure by simply switching off the matrix of hopping integrals $t_{n\nu,n-1\mu}$ between atomic orbitals ν, μ localized in planes $n-1$ and n. We shall, therefore, adopt the tight-binding description in this section bearing in mind that our approach applies, for example, also to LMTO tight-binding band structure implemented for a layer system [24]. Each layer in the structure is therefore described by a tight-binding model, in general multiorbital with s, p, and d orbitals whose one-electron

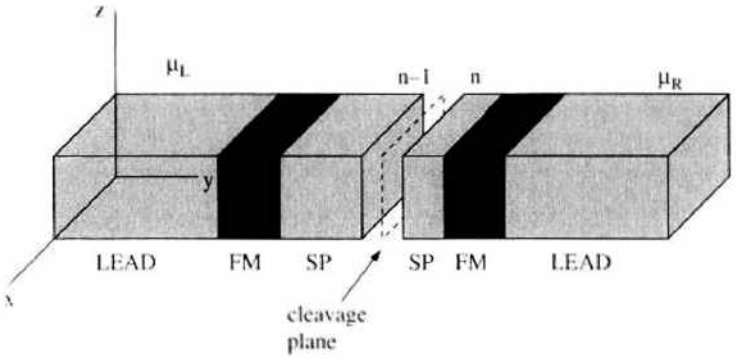


Figure 14. Schematic picture of a magnetic layer structure. All currents flow in the direction of the y axis which is perpendicular to the layers.

parameters are fitted to first-principle bulk band structure. The Hamiltonian is, therefore, of the form

$$H = H_0 + H_{\rm int} + H_{\rm anis} \tag{13}$$

where the one-electron hopping term H_0 is given by

$$H_0 = \sum_{\boldsymbol{k}_\parallel \sigma} \sum_{m\mu,\,n\nu} t_{m\mu,\,n\nu}(\boldsymbol{k}_\parallel) c^\dagger_{\boldsymbol{k}_\parallel m\mu\sigma} c_{\boldsymbol{k}_\parallel n\nu\sigma} \tag{14}$$

where $c^\dagger_{\boldsymbol{k}_\parallel m\mu\sigma}$ creates an electron in a Bloch state, with in-plane wave vector $\boldsymbol{k}_\parallel$ and spin σ, formed from a given atomic orbital μ in plane m. $H_{\rm int}$ is an on-site interaction between electrons in d orbitals, which leads to an exchange splitting of the bands in the ferromagnets and is neglected in the spacer and lead. Finally, $H_{\rm anis}$ contains effective fields corresponding to uniaxial $\boldsymbol{H}_u$ and easy-plane $\boldsymbol{H}_p$ anisotropies that will be needed in the calculation of the current-induced switching of magnetization discussed in Section 7. We do not require the actual form of $H_{\rm anis}$ in the general formulation described in this section but $H_{\rm anis}$ will have to be specified explicitly in Section 7.

Using the equation of motion for the spin angular momentum operator, it is straightforward to show within the tight-binding description (see, e.g., Ref. [25]) that the operator for spin angular momentum current between planes $n-1$ and n is given by

$$\boldsymbol{j}_{n-1} = -\frac{i}{2} \sum_{\boldsymbol{k}_\parallel \mu\nu} t(\boldsymbol{k}_\parallel)_{n\nu,\,n-1\mu} (c^\dagger_{\boldsymbol{k}_\parallel n\nu\uparrow},\, c^\dagger_{\boldsymbol{k}_\parallel n\nu\downarrow}) \boldsymbol{\sigma} (c_{\boldsymbol{k}_\parallel n-1\mu\uparrow},\, c_{\boldsymbol{k}_\parallel n-1\mu\downarrow})^T + \text{h.c.} \tag{15}$$

Here, $\boldsymbol{\sigma} = (\sigma_x, \sigma_y, \sigma_z)$, where the components are Pauli matrices. Equation (15) yields the charge current operator if $\frac{1}{2}\boldsymbol{\sigma}$ is replaced by a unit matrix multiplied by the electronic charge $e/\hbar$, where e is the electronic charge (negative). All currents flow in the y direction, perpendicular to the layers, and the components of the vector $\boldsymbol{j}$ correspond to transport of x, y, and z components of spin.

To use the Keldysh formalism [26–28] to calculate the charge or spin currents flowing between the planes $n-1$ and n, we consider an initial state at time $\tau = -\infty$ in which the hopping integral $t_{n\nu,\,n-1\mu}$ between planes $n-1$ and n is switched off. Then both sides of the system are in equilibrium but with different chemical potentials μ_L on the left and μ_R on the right, where $\mu_L - \mu_R = eV_b$. The interplane hopping is then turned on adiabatically and the system evolves to a steady state. The cleavage plane, across which the hopping is initially switched off, may be taken in either the spacer or in one of the magnets or in one of the leads. In principle, the Keldysh method is valid for arbitrary bias V_b but here we restrict ourselves to small bias corresponding to linear response. This is always reasonable for a metallic system since for larger bias electrons would be injected into the right part of the system far above the Fermi level and many-body processes neglected here would be important. Furthermore, in metallic systems the bias will never be large. For a system with an insulating spacer layer (tunneling junction), a larger bias across the insulating layer can be maintained but the discussion of the bias dependence of perpendicular transport is beyond the scope of the present review. We shall, therefore, always consider only the linear response limit.

Following Keldysh [26, 27], we define a two-time matrix

$$G^+_{\rm RL}(\tau, \tau') = i\langle c^\dagger_L(\tau') c_R(\tau)\rangle \tag{16}$$

where $R \equiv (n, \nu, \sigma')$ and $L \equiv (n-1, \mu, \sigma)$, and we suppress the $\boldsymbol{k}_\parallel$ label. The thermal average in Eq. (16) is calculated for the steady state of the coupled system. The matrix $G^+_{\rm RL}$ has dimensions $2m \times 2m$, where m is the number of orbitals on each atomic site and is written so that the $m \times m$ upper diagonal block contains matrix elements between $\uparrow$ spin orbitals and the $m \times m$ lower diagonal block relates to $\downarrow$ spin. The $2m \times 2m$ hopping matrices $t_{\rm LR}$ and $t_{\rm RL}$ are written similarly and in this case only the diagonal blocks are nonzero. If we denote $t_{\rm LR}$ by t, then $t_{\rm RL} = t^*$. We also generalize the definition of $\boldsymbol{\sigma}$ so that its components

are now direct products of the 2×2 Pauli matrices σ_x, σ_y, σ_z and the $m \times m$ unit matrix. The thermal average of the spin current operator, given by Eq. (15), may now be expressed as

$$\langle \boldsymbol{j}_{n-1} \rangle = \frac{1}{2} \sum_{\boldsymbol{k}_\parallel} \mathrm{Tr} \left\{ [G^+_{\mathrm{RL}}(\tau, \tau) t - G^+_{\mathrm{LR}}(\tau, \tau) t^\dagger] \boldsymbol{\sigma} \right\} \tag{17}$$

Introducing the Fourier transform $G^+(\omega)$ of $G^+(\tau, \tau')$, which is a function of $\tau - \tau'$, we have

$$\langle \boldsymbol{j}_{n-1} \rangle = \frac{1}{2} \sum_{\boldsymbol{k}_\parallel} \int \frac{d\omega}{2\pi} \mathrm{Tr} \left\{ [G^+_{\mathrm{RL}}(\omega) t - G^+_{\mathrm{LR}}(\omega) t^\dagger] \boldsymbol{\sigma} \right\} \tag{18}$$

Again, the charge current is given by Eq. (18) with $\frac{1}{2}\boldsymbol{\sigma}$ replaced by the unit matrix multiplied by $e/\hbar$.

Similarly, the total spin angular momentum on atomic planes on either side of the cleavage plane, in the nonequilibrium state, is given by

$$\langle \boldsymbol{S}_{n-1} \rangle = -\frac{1}{2} i\hbar \sum_{\boldsymbol{k}_\parallel} \int \frac{d\omega}{2\pi} \mathrm{Tr} \left\{ G^+_{LL}(\omega) \boldsymbol{\sigma} \right\} \tag{19}$$

$$\langle \boldsymbol{S}_n \rangle = -\frac{1}{2} i\hbar \sum_{\boldsymbol{k}_\parallel} \int \frac{d\omega}{2\pi} \mathrm{Tr} \left\{ G^+_{RR}(\omega) \boldsymbol{\sigma} \right\} \tag{20}$$

Following Keldysh [26, 27], we now write

$$G^+_{AB}(\omega) = \frac{1}{2}(F_{AB} + G^a_{AB} - G^r_{AB}) \tag{21}$$

where the suffices A and B are either R or L. $F_{AB}(\omega)$ is the Fourier transform of

$$F_{AB}(\tau, \tau') = -i \langle [c_A(\tau), c^\dagger_B(\tau')]_- \rangle \tag{22}$$

and G^a, G^r are the usual advanced and retarded Green functions [29]. Note that in Refs. [26] and [27] the definitions of G^a and G^r are interchanged and that in the Green function matrix defined by these authors G^+ and G^- should be interchanged.

Charge and spin current and spin density are related by Eqs. (17)–(20) to the quantities G^a, G^r, and F_{AB}. The latter are calculated for the coupled system by starting with decoupled left and right systems, each in equilibrium, and turning on the hopping between planes L and R as a perturbation. Hence, we express G^a, G^r, and F_{AB} in terms of retarded surface Green functions $g_L \equiv g_{LL}$, $g_R \equiv g_{RR}$ for the decoupled equilibrium system. The surface Green functions for the decoupled system are determined by the method of adlayers that will be described later in this section. It is then straightforward to show that the spin current between planes $n-1$ and n can be written as the sum $\langle \boldsymbol{j}_{n-1} \rangle = \langle \boldsymbol{j}_{n-1} \rangle_1 + \langle \boldsymbol{j}_{n-1} \rangle_2$, where the two contributions to the spin current $\langle \boldsymbol{j}_n \rangle_1$ and $\langle \boldsymbol{j}_n \rangle_2$ are given by

$$\langle \boldsymbol{j}_{n-1} \rangle_1 = \frac{1}{4\pi} \sum_{\boldsymbol{k}_\parallel} \int d\omega \, \mathrm{Re}\,\mathrm{Tr}\{(B - A)\boldsymbol{\sigma}\}[f(\omega - \mu_L) + f(\omega - \mu_R)] \tag{23}$$

$$\langle \boldsymbol{j}_{n-1} \rangle_2 = \frac{1}{2\pi} \sum_{\boldsymbol{k}_\parallel} \int d\omega \, \mathrm{Re}\,\mathrm{Tr} \left\{ \left[g_L t A B g_R^\dagger t^\dagger - AB + \frac{1}{2}(A + B) \right] \boldsymbol{\sigma} \right\} [f(\omega - \mu_L) - f(\omega - \mu_R)] \tag{24}$$

Here, $A = [1 - g_R t^\dagger g_L t]^{-1}$, $B = [1 - g_R^\dagger t^\dagger g_L^\dagger t]^{-1}$, and $f(\omega - \mu)$ is the Fermi function with chemical potential μ and $\mu_L - \mu_R = eV_b$. In the linear-response case of small bias, which we are considering, the Fermi functions in Eq. (24) are expanded to first order in V_b. Hence the energy integral is avoided, being equivalent to multiplying the integrand by eV_b and evaluating it at the common zero-bias chemical potential μ_0.

Following the method outlined for obtaining Eqs. (23)–(24), similar expressions in terms of retarded surface Green functions may be obtained for the nonequilibrium spin angular

momentum on atomic plane n. Writing again $\langle \boldsymbol{S}_n \rangle = \langle \boldsymbol{S}_n \rangle_1 + \langle \boldsymbol{S}_n \rangle_2$, we obtain

$$\langle \boldsymbol{S}_n \rangle_1 = -\frac{\hbar}{4\pi} \sum_{\boldsymbol{k}_\parallel} \int d\omega \, \mathrm{Im} \, \mathrm{Tr} \{ A g_R \boldsymbol{\sigma} \} [f(\omega - \mu_L) + f(\omega - \mu_R)] \tag{25}$$

$$\langle \boldsymbol{S}_n \rangle_2 = -\frac{\hbar}{2\pi} \sum_{\boldsymbol{k}_\parallel} \int d\omega \, \mathrm{Im} \, \mathrm{Tr} \left\{ \left(A - \frac{1}{2} \right) B g_R^\dagger \boldsymbol{\sigma} \right\} [f(\omega - \mu_L) - f(\omega - \mu_R)] \tag{26}$$

To obtain $\langle \boldsymbol{S}_{n-1} \rangle$ defined by Eq. (19), we must interchange L and R, and t and $t^\dagger$, everywhere in Eqs. (25)–(26). Equations (23)–(26) contain all the information about transport of spin and charge in the direction perpendicular to the layers of an arbitrary magnetic layer structure. The application of this formalism to the oscillatory exchange coupling and to the current-induced switching of magnetization involves the calculation of the spin current in various parts of the layer structure and will be discussed in Sections 4 and 7.

When the Pauli matrix $\frac{1}{2}\boldsymbol{\sigma}$ in the expression for the spin current is replaced by a unit matrix multiplied by $e/\hbar$, we obtain an expression for the charge current within the Keldysh formalism. We can now make a contact with the real space Kubo formula for perpendicular transport which was derived earlier [25] for a multiorbital tight-binding band structure from the two-particle Green function. First of all, we note that in the case of charge current, the contribution coming from Eq. (23), which is proportional to the sum of the Fermi functions $f(\omega - \mu_L) + f(\omega - \mu_R)$, vanishes and the only contribution to the charge current is that given by Eq. (24). Since we are working in the linear-response limit, the charge current is proportional to eV_b. It follows that Eq. (24) yields the total conductance of the system. The real-space Kubo formula [25] applies to a trilayer in which the magnetic moments of the two magnets separated by a nonmagnetic spacer are either parallel or antiparallel. In that case, the total conductance can be written as the sum of the conductances Γ^σ in the up- and down-spin spin channels σ. In this geometry, Eq. (24) simplifies and takes the form

$$\Gamma^\sigma = \frac{4e^2}{h} \sum_{\boldsymbol{k}_\parallel} \mathrm{Tr} \left([T_\sigma \, \mathrm{Im} \, g_L^\sigma] \cdot [T_\sigma^\dagger \, \mathrm{Im} \, g_R^\sigma] \right) \tag{27}$$

where the matrices T_σ and $T_\sigma^\dagger$ are defined by

$$T_\sigma = t[I - g_R^\sigma t^\dagger g_L^\sigma t]^{-1} \tag{28}$$

Here, I is a unit matrix in the orbital space and the Green functions g_L^σ, g_R^σ are $m \times m$ matrices corresponding to the upper ($\sigma = \uparrow$) and lower ($\sigma = \downarrow$) diagonal blocks of the surface Green functions g_L, g_L defined earlier. Equation (27) has exactly the same form as the real-space Kubo formula derived in Ref. [25] and will be used to discuss the CPP GMR in Section 5 and the tunneling magnetoresistance in Section 6.

It is also useful to note that the real-space Kubo formula (27) is equivalent to the formula for the conductance derived by Landauer using methods of the scattering theory [30]. We shall therefore frequently refer to Eq. (27) as the Kubo-Landauer formula.

It can be seen that Eqs. (23)–(27), which determine the spin and charge currents and the transport spin densities without any approximations, all depend on just two quantities, that is, the surface one-electron Green functions for a system cleaved between two neighboring atomic planes. Since these surface Green functions play such an important role, we shall now briefly discuss how they can be calculated.

3.1. Matrix Möbius Transformation and Its Application in the Calculation of Surface Green Functions

The bilinear, or Möbius, transformation $\zeta(z) = (az + b)(cz + d)^{-1}$ is a well-known conformal transformation that transforms circles in the complex z plane into circles in the ζ plane. Umerski [31] showed that a generalization of this transformation to the case where a, b, c, d, z, are $M \times M$ (complex) matrices is very useful for the calculation of surface

Green functions in the tight-binding theory of layered structures. We follow his discussion in this section. We define

$$A_{\bullet}z = (az + b)(cz + d)^{-1} \tag{29}$$

where A is the $2M \times 2M$ partitioned matrix

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \tag{30}$$

It is straightforward to show that

$$A_{\bullet}(B_{\bullet}z) = (AB)_{\bullet}z \tag{31}$$

Suppose that a number of identical atomic planes are deposited on a substrate whose local surface Green's function (SGF) g_0 is a $(M \times M)$ matrix. Typically $M = 9$ or 18 for tight-binding bands with s, p, d orbitals, depending on whether g represents carriers with a single spin projection, or both spin projections. Suppose that after N planes have been deposited the local Green function for the last layer is g_N, as shown in Fig. 15. We now show that g_N is related recursively to g_{N-1} by a Möbius transformation. The hopping matrix $t = t(\boldsymbol{k}_{\parallel})$ and the on-site energy matrix $u = u(\boldsymbol{k}_{\parallel})$ for the deposited planes are both $M \times M$. Using the Dyson equation, it can be shown that [32]

$$g_N = (\omega - u - t^{\dagger} g_{N-1} t)^{-1} \tag{32}$$

where u is the diagonal matrix element of the total Hamiltonian defined by Eq. (13). We can write

$$g_N = t^{-1}[(\omega - u)t^{-1} - t^{\dagger} g_{N-1}]^{-1} = X_{\bullet} g_{N-1}, \quad \text{where} \quad X = \begin{pmatrix} 0 & t^{-1} \\ -t^{\dagger} & (\omega - u)t^{-1} \end{pmatrix} \tag{33}$$

We note that in most cases of interest $t(\boldsymbol{k}_{\parallel})$ is nonsingular and that X is a $2M \times 2M$ matrix. By iterating Eq. (33), and using Eq. (31), we obtain the required surface Green function

$$g_N = X_{\bullet}^{N} g_0 \tag{34}$$

We proceed to find the eigenvalues λ of X to diagonalize X and calculate X^N. In fact, some of the eigenvalues are related to the bulk band structure of the overlayer. The eigenvalues λ satisfy $\det(X - \lambda I) = 0$, where I is the unit matrix. It follows that

$$\begin{pmatrix} I & 0 \\ -\lambda^{-1} t^{\dagger} & I \end{pmatrix} (X - \lambda I) = \begin{pmatrix} -\lambda I & t^{-1} \\ 0 & -\lambda^{-1} t^{\dagger} t^{-1} + (\omega - u)t^{-1} - \lambda I \end{pmatrix} \tag{35}$$

Hence

$$\det(X - \lambda I) = (-1)^N \det(\lambda t^{-1}) \det(\omega - u - \lambda t - \lambda^{-1} t^{\dagger}) \tag{36}$$

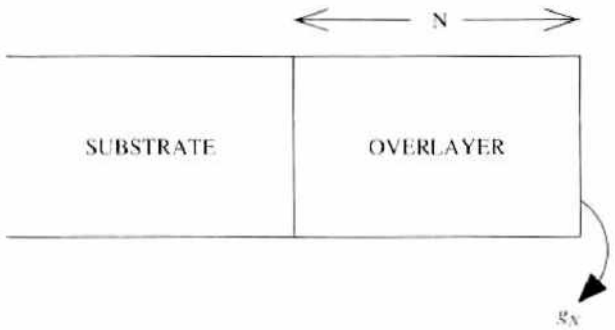


Figure 15. A schematic figure depicting an overlayer system with surface Green's function g_N.

and the eigenvalues satisfy

$$\det(\omega - u - \lambda t - \lambda^{-1} t^{\dagger}) = 0 \tag{37}$$

where the determinant is $M \times M$ and ω stands for ωI, I being the $M \times M$ unit matrix. Furthermore, as $\lambda \to 0$, we see from Eq. (36) that $|\det(X)| = 1$, which means that there are no zero eigenvalues. We also note from Eq. (37), that if λ is an eigenvalue, then $(\lambda^*)^{-1}$ is also an eigenvalue. In general, there are $2(M - P)$ eigenvalues with $|\lambda| \neq 1$ and $2P$ eigenvalues with $|\lambda| = 1$, for some integer P ($0 \leq P \leq N$). If $|\lambda| = 1$, we write $\lambda = e^{ik}$, $-\pi < k \leq \pi$. Equation (37) is then the equation determining the bulk overlayer band structure $\omega = E(k_{\parallel}, k_{\perp})$ with $k_{\perp} = k$. Hence, for a given $k_{\parallel}$, $k_{\perp} = \pm k$ are points on the constant energy surface of the bulk spacer with energy ω. If there is only one such pair of points, then $P = 1$, but for a constant energy surface with more than one sheet, we may have $P > 1$. Thus eigenvalues λ on the unit circle are determined by the propagating states of the overlayer bulk band structure. Those with $|\lambda| \neq 1$ relate to surface states that are exponentially attenuated within the bulk.

The eigenvalues and eigenvectors of X must be determined numerically so that X may be diagonalized by a similarity transformation

$$O^{-1} X O = \Lambda = \begin{pmatrix} \Lambda_1 & 0 \\ 0 & \Lambda_2 \end{pmatrix} \tag{38}$$

where the M eigenvalues in the diagonal $M \times M$ matrices Λ_1, Λ_2 are ordered in ascending order of magnitude down the diagonal (i.e., $|\lambda_1| \leq |\lambda_2| \leq \cdots \leq |\lambda_{2N}|$). Eigenvalues off the unit circle occur in pairs $\lambda = re^{i\theta}$, $(\lambda^*)^{-1} = r^{-1}e^{i\theta}$, one of the pair (with modulus less than 1) appearing in Λ_1 and the other (with modulus greater than 1) in Λ_2. The eigenvalues on the unit circle also occur in pairs $e^{\pm ik}$ ($k > 0$), and we may put e^{ik} in Λ_1 and e^{-ik} in Λ_2; the precise order is not important at this stage but a natural way of doing this is discussed next.

It follows from Eqs. (34) and (38) that

$$g_N = (O \Lambda^N O^{-1})_{\bullet} g_0 \tag{39}$$

We define $f_N = O_{\bullet}^{-1} g_N$ so that

$$f_N = (\Lambda^N O^{-1})_{\bullet} g_0 = \Lambda_{\bullet}^N f_0 \tag{40}$$

Hence, using the definition (29), we find that $f_N = \Lambda_1 f_0 \Lambda_2^{-1}$ and

$$g_N = O_{\bullet} f_N = O_{\bullet} (\Lambda_1^N f_0 \Lambda_2^{-N}) \tag{41}$$

For large N only eigenvalues $e^{\pm ik_{\perp i}} (i = 1, 2, \ldots P)$ on the unit circle contribute; the other diagonal elements of Λ_1^N and Λ_2^N tend to zero as $N \to \infty$. Thus g_N depends on N through quantities such as

$$e^{iN(k_{\perp i} + k_{\perp j})} = e^{iN_i k_{\perp i}} e^{iN_j k_{\perp j}} \quad \text{and} \quad e^{2iNk_{\perp i}} = e^{2iN_i k_{\perp i}} \tag{42}$$

with $N_1 = \cdots = N_P = N$. Hence g_N, and any function of it, is periodic in the variables $N_1, \ldots, N_P$ (which can be considered as continuous) with periods $2\pi/k_{\perp 1}, \ldots, 2\pi/k_{\perp P}$, respectively. Functions such as g_N are termed "quasi-periodic" functions of N.

For the case $P = 1$, there is only one wave vector $k_{\perp 1}$ and g_N is thus periodic in N with period $\pi/k_{\perp 1}$ rather than $2\pi/k_{\perp 1}$. (NB: This can also happen for $P > 1$.)

To calculate g_N from Eq. (41), we need $f_0 = O_{\bullet}^{-1} g_0$ and therefore have to calculate the surface Green function g_0 for a semi-infinite substrate. To do this, it is necessary to be more precise about how we order the eigenvalues on the unit circle. The trick is to include a small positive imaginary part δ in the energy ($\omega \to \omega + i\delta$) and to let $N \to \infty$ before $\delta \to 0$. This shifts the pairs of eigenvalues on the unit circle in such a way that one moves just inside the

circle and the other just outside. The former are in Λ_1 and the latter in Λ_2, so that Λ_1^N and Λ_2^N tend to zero as $N \to \infty$. Hence, if we write

$$O = \begin{pmatrix} o_1 & o_2 \\ o_3 & o_4 \end{pmatrix} \tag{43}$$

it follows from Eq. (41) that

$$g_0 \equiv g_\infty = O_\bullet 0 = o_2 o_4^{-1} \tag{44}$$

Thus g_0 is determined by half the eigenvectors of X, which form the columns of the right-hand half of O.

We now summarize the steps needed to determine the overlayer Green function g_N, which is all that is required to solve all the CPP transport equations derived within the Keldysh formalism. To calculate g_N for the overlayer system shown in Fig. 15, we proceed as follows. First we create the matrix X defined in Eq. (33) using the substrate on-site potentials and hoppings. We diagonalize the matrix X and obtain from Eq. (41) the surface Green function g_0 for the semi-infinite substrate. We then create the new matrix X for the overlayer, using the overlayer on-site potentials and hoppings, diagonalize it, and use Eq. (41) to obtain g_N. The overlayer surface Green function obtained in this way will be a continuous function of N, and quasi-periodic in N, with periods given by $2\pi/k_{\perp 1}, \ldots, 2\pi/k_{\perp P}$ for $P > 1$, or $\pi/k_\perp$ for $P = 1$.

4. OSCILLATORY EXCHANGE COUPLING

In 1986, Grünberg et al. [7] observed that Fe layers separated by a Cr spacer layer were ordered antiferromagnetically. Subsequently, Parkin et al. [9] have made a careful study of the dependence of this effect on the thickness of the chromium spacer layer. Parkin et al. found by measuring the strength of the applied magnetic field, required to saturate the total moment, that, in fact, regions of an antiferromagnetic ordering alternate periodically with intervals of Cr thicknesses in which the moment saturates easily. They deduced from these measurements that there is a magnetic coupling between the Fe layers mediated by the non-magnetic spacer whose sign changes periodically with the spacer layer thickness. The effect is known as the oscillatory exchange coupling. Samples with Cr thicknesses corresponding to an antiferromagnetic coupling exhibited also the GMR effect discussed in Section 2. Oscillations of the GMR effect shown in Fig. 3 of Section 2 thus demonstrate at the same time the existence of an oscillatory exchange coupling. The saturation field experiment of Parkin et al. [9] gives directly the strength of the antiferromagnetic coupling (strength of the saturation field) but clearly cannot give any information about the ferromagnetic coupling. By using the Brillouin scattering of light, Demokritov et al. [33] were able to show directly that the oscillatory exchange coupling actually changes its sign. Their results for Fe/Cr system are reproduced in Fig. 16. The oscillation period observed by Demokritov et al. [33] was very long ≈18 Å. It was shown very convincingly by Unguris et al. [34] that in addition to

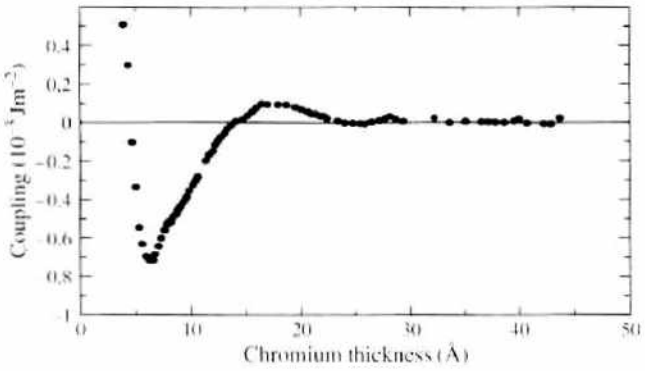


Figure 16. Observed dependence of the oscillatory exchange coupling in Fe/Cr/Fe(001) trilayer on Cr thickness [33].

the long period there is also a short oscillation period of two atomic planes for Fe/Cr(001) system. Oscillatory exchange coupling was subsequently observed in many other magnetic multilayers and of those Co/Cu is perhaps the most typical and best investigated system [35–43].

The physical origin of the oscillatory exchange coupling was explained by Edwards and Mathon [44] who pointed out that majority- and minority-spin carriers in a magnetic multilayer see potential steps of unequal heights at magnet/nonmagnet interfaces. Their theory was based on the observation already discussed in Section 2 that there is an almost perfect match between the minority-spin band of Fe and the bands of Cr but a large mismatch for the majority-spin bands. A very similar picture holds for Co/Cu system but the roles of majority- and minority-spin bands are interchanged (see Section 2). Hence in the ferromagnetic (FM) configuration of the magnetic layers electrons of one spin orientation see a deep potential well in the spacer layer region, whereas carriers of the opposite spin orientation move freely across all the interfaces. In the antiferromagnetic (AF) configuration, carriers of either spin see a potential step. It follows that in the FM configuration carriers of one spin orientation are trapped in a quantum well. As the thickness of the spacer layer varies, quantum well states crossing the Fermi surface cause oscillations of the thermodynamic potential of the system, which results in oscillations of the interlayer exchange coupling. Early theories of this type [45–47] were based on a single-orbital tight-binding model, but they were later generalized to a fully realistic band structure in the case of Co/Cu/Co(001) trilayer [48, 49].

The quantum well (QW) theory of the oscillatory exchange coupling assumes that the perturbation to the system caused by potential steps at the magnet/nonmagnet interfaces is so strong that magnetic carriers may become completely confined in a quantum well. The QW theories must, therefore, be based on a nonperturbative calculation of the total energy (thermodynamic potential). There is also an alternative point of view, which assumes that potential steps are only a weak perturbation to the system. In that case, the usual RKKY-type theory [51], which describes an oscillatory exchange interaction between two magnetic impurities embedded in a nonmagnetic matrix can be easily adapted to a layer geometry. This was done by Bruno and Chappert [52]. The RKKY theory of Bruno and Chappert is very intuitive and gives the correct oscillation periods in terms of the spanning vectors of the spacer layer Fermi surface. However, because the interaction of carriers with the ferromagnetic interface is an adjustable parameter in such a theory it cannot predict the strength and other properties of the coupling.

A refinement of the RKKY second-order perturbation theory is to treat scattering of carriers from interfaces more realistically by including in the perturbation series expansion of the total energy (thermodynamic potential) higher-order corrections [53, 54]. If corrections to all orders of perturbation theory are included, such a perturbative treatment becomes equivalent to the original quantum well theory [45–49]. Since it will be seen that the thermodynamic potential of an arbitrary magnetic multilayer can be calculated quite rigorously without using perturbation theory, there is no advantage in the perturbative formulations [53, 54] except for their pedagogical merit.

The validity of the concept of quantum wells was confirmed by direct observation of QWs in photoemission and inverse photoemission experiments by Ortega et al. [55]. Spin polarization of the QW states was observed by Garrison et al. [56].

All the theories we have discussed are based on the evaluation of the total energy difference between the FM and AF configurations of the magnetic layers in a magnetic multilayer. Another very interesting approach, which appears to be quite different, was pioneered by Slonczewski [57]. He proposed to calculate the mutual interaction between two ferromagnets separated by an insulating barrier from the spin current, which flows in the barrier when the magnetizations of the two magnetic layers are not colinear. His method was adapted to a metallic trilayer by Hathaway and Cullen [58] and Erickson et al. [59] using a parabolic band model. The spin-current method was also reformulated in terms of one-electron Green functions using a single-orbital tight-binding model [60]. We recall that, using the Keldysh nonequilibrium formalism, we derived in Section 3 completely general formulas (23) and (24) for the spin current in an arbitrary magnetic layer structure. In the absence of an applied bias, the spin current in the nonmagnetic spacer, which is required in Slonczewski's method,

is given by Eq. (23) of Section 3. We can, therefore, generalize the spin current method to a fully realistic band structure.

We begin by showing how to express the total energy difference between the FM and AF configurations, that is, the interlayer exchange coupling, in terms of the local density of states and, ultimately, in terms of local one-electron Green functions. Next we derive from the Keldysh formula for the spin current in the nonmagnetic spacer (23) an alternative formula for the oscillatory exchange coupling. We then demonstrate analytically that the spin-current formula for the oscillatory exchange coupling is completely equivalent to the total energy difference formula used in the QW theories (see, e.g., Ref. [49]). The application of this formalism is illustrated for Co/Cu/Co(001) trilayer. Finally, we show that, by exploiting a periodic (quasiperiodic) dependence of all the Green functions on the spacer layer thickness (see Section 3), we can use asymptotic expansions to derive an analytic formula for the coupling. Such a formula, which is valid in the limit of a thick spacer, yields not only all the oscillation periods but also their amplitudes and decay rates in terms of the band structure parameters.

4.1. Equivalence of the Spin Current and Total Energy Calculations of the Oscillatory Exchange Coupling

For the sake of clarity, we confine our discussion in this section to a simple magnetic trilayer, which consists of two semi-infinite magnetic layers separated by a nonmagnetic metallic spacer of N atomic planes. We also assume that the entire trilayer grows epitaxially. However, our formulas and arguments are valid for general multilayers and are readily extendible to nonepitaxial systems modeled by lateral supercells of arbitrary size and complexity. The only difficulty being the additional degree of computational effort required to perform such supercell calculations accurately.

There are two starting points for a quantitative calculation of the exchange coupling. The first proceeds directly from the definition of the exchange coupling between two semi-infinite magnetic layers across a nonmagnetic spacer layer of N atomic planes. The exchange coupling $J(N)$ per surface atom is defined as

$$J(N) = (\Omega_{\mathrm{FM}}(N) - \Omega_{\mathrm{AF}}(N)) \tag{45}$$

where Ω_{FM} and Ω_{AF} are the thermodynamic potentials of the trilayer per unit cross-sectional area for the parallel (FM) and antiparallel (AF) configurations of the magnetic moments, respectively.

To evaluate Eq. (45), we make two approximations. The first is the so-called force theorem that states that a good approximation to the total energy difference between different structures (in this case magnetic) is obtained by comparing sums of one-electron energies using atomic potentials, which are independent of the magnetic configurations. The second approximation is to neglect the dependence of the local densities of states within the magnetic layers on the magnetic configuration. This approximation can be checked and is found to be a good one. We can therefore write

$$J(N) = [(\Omega^{\uparrow}(N) + \Omega^{\downarrow}(N))_{\mathrm{FM}} - (\Omega^{\uparrow}(N) + \Omega^{\downarrow}(N))_{\mathrm{AF}}] \tag{46}$$

where Ω^{σ} is the thermodynamic potential for electrons of spin σ ($\sigma = \uparrow, \downarrow$).

The thermodynamic potential per surface atom for a given magnetic configuration at temperature T is given by

$$\Omega^{\sigma} = -\frac{k_B T}{N_{\parallel}} \sum_{\boldsymbol{k}_{\parallel}} \int_{-\infty}^{+\infty} \ln\left\{1 + \exp\left[\frac{(\mu - \omega)}{k_B T}\right]\right\} \mathcal{A}^{\sigma}(\omega, \boldsymbol{k}_{\parallel}, N)\, d\omega \tag{47}$$

where $N_{\parallel}$ is the number of $\boldsymbol{k}_{\parallel}$ points in the two-dimensional Brillouin zone, μ is the chemical potential, and $\mathcal{A}^{\sigma}$ is the spectral density for particles of spin σ in the trilayer having that configuration. Because of the in-plane translational invariance, we label all the trilayer states

by the plane index i and by the wave vector $k_\parallel$ parallel to the layers. The spectral density $\mathscr{A}^\sigma$ is given by

$$\mathscr{A}^\sigma(\omega, k_\parallel, N) = -\frac{1}{\pi} \operatorname{Im} \operatorname{Tr} \sum_i G_{ii}^\sigma(\omega, k_\parallel, N) \tag{48}$$

where G_{ii}^σ is the diagonal matrix element of the tight-binding one-electron Green function, the trace is over all atomic orbitals, and the sum over i is over all atomic planes in the trilayer.

The problem now is to evaluate the spectral density, bearing in mind that we have an infinite number of atomic plane in our system. This is performed by choosing a pair of neighboring atomic planes in the spacer layer and setting the hopping t between them equal to zero, thereby separating the trilayer into two independent semi-infinite systems. For simplicity, we place such a cleavage plane between the Nth atomic plane in the spacer and the first atomic plane of the right hand ferromagnet. However, the results are independent of the location of the cleavage plane. The cleaved system shown in Fig. 17 then consists of the left ferromagnet with an overlayer of N atomic planes of the spacer and the right ferromagnet. We label the surface Green's functions of these two surface systems by g_{LN} and g_R, and recall that these are matrices, whose indices range over s, p, and d orbitals.

Using Dyson's equation, it can be shown [49] that a closed-form expression for the sum in Eq. (48) can be obtained

$$\mathscr{A}^\sigma(N) = \mathscr{A}_L^\sigma + \mathscr{A}_R^\sigma - \frac{1}{\pi} \operatorname{Im} \operatorname{Tr} \left(\frac{d}{d\omega} \ln(I - g_R^\sigma t^\dagger g_{LN}^\sigma t) \right)$$

where

$$\mathscr{A}_L^\sigma(j) = -\frac{1}{\pi} \operatorname{Im} \operatorname{Tr} \sum_{p=-\infty}^{N} g_{p,p}^\sigma, \quad \mathscr{A}_R^\sigma = -\frac{1}{\pi} \operatorname{Im} \operatorname{Tr} \sum_{p=N+1}^{\infty} g_{p,p}^\sigma \tag{49}$$

Here $g_{p,p}^\sigma$ is the diagonal element of the Green's function in an atomic plane p of the cleaved system. The first two terms $\mathscr{A}_L^\sigma$ and $\mathscr{A}_R^\sigma$ are simply the spectral densities for the left and right surface systems. Since they refer to the cleaved system, they are clearly independent of the magnetic configuration, and hence, do not contribute to the coupling. The last term, on the other hand, gives the difference between the spectral densities of the connected and cleaved systems. It is, therefore, this term that determines the exchange coupling. We note that it depends only on the surface Green's functions of the cleaved system and on the hopping matrix t connecting the left and right surfaces when the two cleaved systems are reconnected (Fig. 17). Substituting this expression in to Eq. (47) and integrating by parts over the energy gives

$$J(N) = \frac{1}{N_\parallel} \sum_{k_\parallel} \int_{-\infty}^{+\infty} f(\omega - \mu) F(\omega, k_\parallel, N) d\omega, \quad \text{where} \quad F = \frac{1}{\pi} \operatorname{Im} \ln \det(R_{\mathrm{FM}}^\uparrow R_{\mathrm{FM}}^\downarrow R_{\mathrm{AF}}^{\uparrow\,-1} R_{\mathrm{AF}}^{\downarrow\,-1}) \tag{50}$$

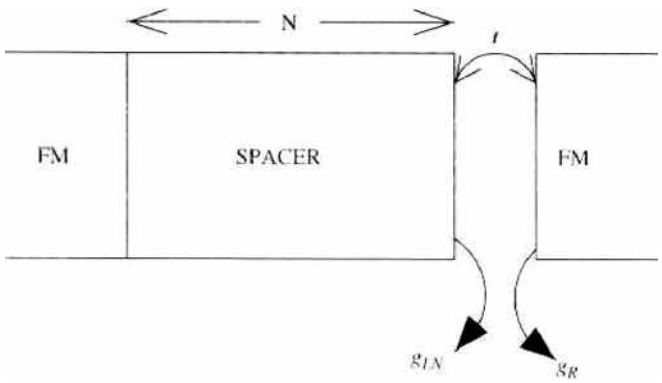


Figure 17. A schematic figure depicting the cleaved trilayer system, the left and right hand surface Green's functions g_{LN} and g_R, and the hopping t of the connected system.

$f(\omega - \mu)$ is the Fermi function with chemical potential μ, and the term $R^{\sigma} = (1 - g_R^{\sigma} t^{\dagger} g_L^{\sigma} t)$ is evaluated in the ferromagnetic (FM) or antiferromagnetic (AF) configurations. Finally, g_L^{σ} and g_R^{σ} are the surface Green's functions on either side of the cleavage plane.

The second method for calculating the exchange coupling is based on the expressions for the spin-current derived from the Keldysh formalism in Section 3. Consider a configuration of the trilayer in which the magnetic moment of the right ferromagnet is rotated in the plane of the layers to make an angle θ with the moment of the left ferromagnet. We recall that all the layers are parallel to the (x, z)-plane of the system of coordinates in Fig. 14 of Section 3. The corresponding thermodynamic potential of the trilayer is $\Omega(\theta)$, and the exchange coupling defined by Eq. (45) can be thus written

$$J(N) = \Omega(0) - \Omega(\pi) = \int_0^{\pi} d\theta \left(-\frac{d\Omega}{d\theta} \right) \tag{51}$$

where the integrand $-d\Omega/d\theta$ is the torque exerted on one magnet by the other. Owing to exchange coupling, the left ferromagnet exerts a torque on the spins of the right ferromagnet, which causes them to precess. Their total rate of change of angular momentum is equal to the torque

$$-\frac{d\Omega}{d\theta} = \left\langle \sum_i \dot{S}_{iy} \right\rangle \tag{52}$$

where the right-hand side is the rate of change of the y component of the spin angular momentum summed over all the spins in the right ferromagnet. By continuity, this is equal to the rate of flow of the y component of the spin angular momentum across the spacer (spin current). Hence

$$-\frac{d\Omega}{d\theta} = \langle j_y \rangle \tag{53}$$

where $\langle j_y \rangle$ is the y component of the spin current. A general expression for the spin-current in the presence of an applied bias was obtained in Section 3 using the Keldysh formalism (see Eq. (23) and Eq. (24)). We are here interested in the case when there is no bias applied to the trilayer. In that case, the chemical potentials on the left and the right of the cleavage plane are identical ($\mu_L = \mu_R = \mu$), and it follows that the y component of the spin-current flowing across the cleavage plane is determined by Eq. (23). Using Eqs. (53) and (23), we find that the exchange coupling $J(N)$ is given by

$$J(N) = \frac{1}{2\pi} \sum_{\mathbf{k}_{\parallel}} \int d\omega \int_0^{\pi} d\theta \, \mathrm{Re}\, \mathrm{Tr} \{(B - A)\sigma_y\} f(\omega - \mu) \tag{54}$$

We recall that the quantities A, B are defined by $A = [1 - g_R t^{\dagger} g_L t]^{-1}$, $B = [1 - g_R^{\dagger} t^{\dagger} g_L^{\dagger} t]^{-1}$.

For a single orbital tight-binding band, the integral with respect to θ in Eq. (54) can be done analytically and it is then easy to show that the total energy formula (50) and the spin current formula (54) for the coupling are identical. For a general multiorbital band structure, it is easier to generalize first the total energy formula (50) to an arbitrary angle between the magnetic moments θ and then differentiate with respect to θ to prove the equivalence of Eqs. (50) and (54).

The total energy formula for the exchange coupling (50) can be readily evaluated numerically either for a tight-binding parameterization of the *ab initio* band structure [48, 49] or by using LMTO tight-binding method [61–63]. The results of these two calculations are virtually identical and we show in Fig. 18 the coupling $J(N)$ for $Co/Cu_N/Co(001)$ obtained by the tight-binding method of Ref. [49]. The tight-binding parameters used in Ref. [49] were obtained from fits to the *ab initio* band structure of Cu and fcc Co [64, 65]. Special care has to be taken to ensure convergence of the sum over $\mathbf{k}_{\parallel}$ and of the energy integral. The latter was carried out in the complex energy plane and replaced by a sum over Matsubara frequencies corresponding to room temperature. The summation over $\mathbf{k}_{\parallel}$ was performed using a dense mesh of 2775 points in the irreducible segment of the two-dimensional Brillouin zone.

The continuous line denotes the result of the numerical calculation at non-integer numbers of atomic Cu planes, obtained by using the analytic extension of the surface Green functions discussed in Section 3. The dashed line was obtained using the stationary phase approximation which will be described next. We shall defer detailed discussion of the numerical results for the exchange coupling shown in Fig. 18 until the end of the section.

4.2. Stationary Phase Approximation Formula for the Exchange Coupling

It might seem from the numerical results presented above that the exchange coupling for Co/Cu/Co(001) oscillates with a single period of 2.6 atomic planes. However, the RKKY theory [52] predicts two oscillation periods, one is the short period of 2.6 atomic planes but a long period of 5.7 atomic planes is also predicted. To understand why the long period oscillation seems to be missing in the realistic total energy calculation, we need some insight into the physics that lies behind such a calculation. This can only be provided by an analytical formula obtained by asymptotic expansions of the total energy expression (50). Such expansions using the stationary phase approximation (SPA) were first derived for the relatively simple case of a (fully solvable) single-orbital tight-binding model [45, 46]. They were later generalized to a realistic band structure but remained restricted to the spacer with a single-sheet Fermi surface. This does, however, includes the important case of Co/Cu/Co(001) system [49]. Only when the Möbius transformation method was developed [31], application of the SPA to the case of a multisheet Fermi surface could be made (see, e.g., Ref. [31]). We shall now explain how such asymptotic expansions are obtained in this most general case.

In Section 3, we established that the Green function $g_N(\omega, \boldsymbol{k}_\parallel)$ at the surface of an overalayer of N atomic planes deposited on the left magnet depends on the overlayer thickness N, energy ω, and in-plane momentum $\boldsymbol{k}_\parallel$. It was also shown in Section 3 that, for large N, the Green function g_N is a quasi-periodic function of the continuous thickness N with periods given by $\{2\pi/k_{\perp 1}, 2\pi/k_{\perp 2}, \ldots, 2\pi/k_{\perp P}\}$. Here, $\{k_{\perp 1}, k_{\perp 2}, \ldots, k_{\perp P}\}$ are the values of the perpendicular wave vector $k_\perp$ satisfying the bulk overlayer dispersion relations $E(\boldsymbol{k}_\parallel, k_\perp) = \omega$ for a given in-plane wave vector $\boldsymbol{k}_\parallel$ and energy ω (different values of $k_{\perp i}$ correspond to different sheets of the spacer Fermi surface). It follows that the integrand $F(\omega, \boldsymbol{k}_\parallel, N)$ in Eq. (50) for the exchange coupling is also a quasi-periodic function of N. It can, therefore, be expanded in a P-dimensional Fourier series:

$$F(\omega, \boldsymbol{k}_\parallel, N) = \sum_{\vec{s}} c_{\vec{s}} e^{iN\phi_{\vec{s}}}, \quad \text{where} \quad \phi_{\vec{s}} \equiv (s_1 k_{\perp 1} + \cdots + s_P k_{\perp P}) \tag{55}$$

and $s_1, s_2, \ldots, s_P$ are integers.

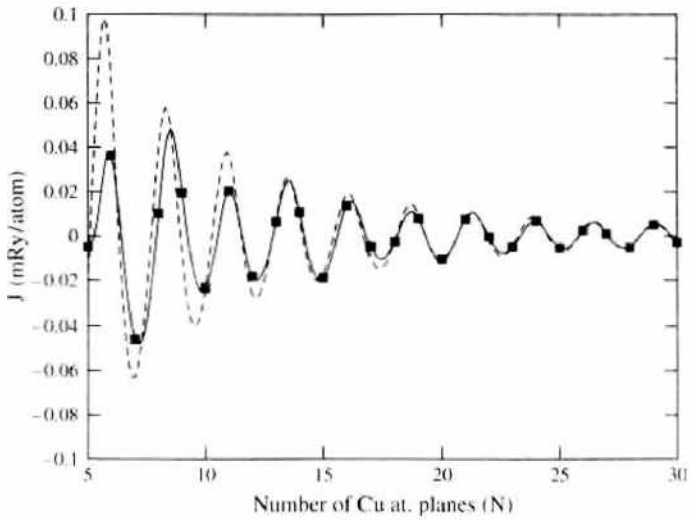


Figure 18. Exchange coupling $J(N)$ in $Co/Cu_N/Co(001)$ as a function of Cu thickness N determined numerically from Eq. (50).

When the Fourier expansion of the function $F(\omega, \boldsymbol{k}_\parallel, N)$ is substituted back in Eq. (50) it can be seen that for large N the imaginary exponential in the integrand oscillates rapidly as a function of $\boldsymbol{k}_\parallel$ and ω. It follows that nonzero contributions to the $\boldsymbol{k}_\parallel$ integral come only from the neighborhood of points in $\boldsymbol{k}_\parallel$-space at which $\phi_{\bar{s}}$ is stationary. Similarly, all the contributions to the energy integral in Eq. (50) cancel except for those coming from the vicinity of a sharp cut off that occurs at an energy equal to the chemical potential μ. It follows that we can expand the argument of the exponential function in powers of $\boldsymbol{k}_\parallel$ about the stationary point(s) and in powers of the energy ω about the cut-off point $\omega = \mu$. Since the Fourier coefficient $c_{\bar{s}}$ is complex,

$$c_{\bar{s}}(\omega, \boldsymbol{k}_\parallel) = |c_{\bar{s}}| e^{i\psi_{\bar{s}}(\omega)} \tag{56}$$

its phase must be included in the argument of the exponential function in Eq. (55), but we may assume that $|c_{\bar{s}}|$ is a slowly varying function of $\boldsymbol{k}_\parallel$ and ω and replace it by its value at the stationary point (cut-off point). This is the essence of the stationary phase approximation. We therefore expand the function $\phi_{\bar{s}}$ in Eq. (55) and the phase of the Fourier coefficients $\psi_{\bar{s}}$, to first order in energy about the chemical potential μ:

$$\phi_{\bar{s}}(\omega) \approx \phi_{\bar{s}}(\mu) + (\omega - \mu)\phi'_{\bar{s}}(\mu), \quad \psi_{\bar{s}}(\omega) \approx \psi_{\bar{s}}(\mu) + (\omega - \mu)\psi'_{\bar{s}}(\mu) \tag{57}$$

We also expand the function $\phi_{\bar{s}}$ to second order in $\boldsymbol{k}_\parallel$ about its stationary points $\{\boldsymbol{k}^0_{\parallel\bar{s}}\}$:

$$\phi_{\bar{s}}(\boldsymbol{k}_\parallel) \approx \phi_{\bar{s}}(\boldsymbol{k}^0_{\parallel\bar{s}}) + \frac{1}{2}(\mathbf{k}_\parallel - \mathbf{k}^0_{\parallel\bar{s}}) \cdot (\partial^2 \phi_{\bar{s}}) \cdot (\mathbf{k}_\parallel - \mathbf{k}^0_{\parallel\bar{s}}) \tag{58}$$

where $(\partial^2 \phi_{\bar{s}})$ is the 2×2 Hessian matrix $(\partial^2 \phi_{\bar{s}})_{\alpha,\beta} \equiv \partial^2 \phi_{\bar{s}} / \partial k_{\parallel\alpha} \partial k_{\parallel\beta}$ evaluated at the stationary point $\mathbf{k}^0_{\parallel\bar{s}}$. Substituting these approximations in Eq. (50), we find that the $\boldsymbol{k}_\parallel$ integrals are now Gaussian and can be easily evaluated analytically. Similarly, the energy integral can be evaluated analytically. We therefore obtain the following formula for the coupling, which is asymptotically exact in the limit of large spacer thickness ($N \to \infty$)

$$J(N) = -\frac{(2\pi)^2 k_B T}{2N A_{BZ}} \sum_{\bar{s}} \sum_{\mathbf{k}^0_{\parallel\bar{s}}} \mathrm{Re}\left(\frac{\tau c_{\bar{s}} e^{iN\phi_{\bar{s}}}}{|\det(\partial^2 \phi_{\bar{s}})|^{1/2} \sinh(\pi k_B T [N\phi'_{\bar{s}} + \psi'_{\bar{s}}])} \right)\Bigg|_{\mu, \mathbf{k}^0_{\parallel\bar{s}}} \tag{59}$$

Here A_{BZ} is the area of the Brillouin zone and $\tau = i$ when both second derivatives in the Hessian matrix are positive, $\tau = -i$ when they are negative, and $\tau = 1$ when the derivatives have opposite signs.

All quantities in Eq. (59) are evaluated at the chemical potential of the spacer (i.e., $\omega = \mu$). We reiterate that the sum over $\{\mathbf{k}^0_{\parallel\bar{s}}\}$ in Eq. (59) is simply a sum over the set of points in the two-dimensional Brillouin zone, at which $\phi_{\bar{s}}(\mathbf{k}_\parallel)$ is stationary.

In the general case of a spacer layer Fermi surface (FS) with many sheets, the evaluation of Eq. (59) is not easy. However for Co/Cu/Co(001), the problem is greatly simplified since the Cu FS has only one sheet in the direction perpendicular to the layers. In this case, the stationary points $\boldsymbol{k}^0_{\parallel s}$ of $\phi_s(\boldsymbol{k}_\parallel)$ coincide with the extrema of the FS $k_\perp(\boldsymbol{k}_\parallel)$. The relevant cross section of the Cu Fermi surface in the direction perpendicular to the layers is shown in Fig. 19. It can be seen that there is a single extremum of $k_\perp(\boldsymbol{k}_\parallel)$ at the belly of the Cu FS (the Γ point $\boldsymbol{k}_\parallel = 0$) and four extrema at the Cu FS necks $\boldsymbol{k}_\parallel a = (\pm 2.53, \pm 2.53)$, where $a = 3.6$ Å is the lattice constant of Cu. The corresponding oscillation periods are: $p^b = 5.7$ atomic planes and $p^n = 2.6$ atomic planes. The SPA formula for Co/Cu/Co(001) thus takes the form

$$J(N) = -\frac{k_B T}{2N} \sum_{\mathbf{k}^0_\parallel} \sum_{s} \mathrm{Re}\left(\frac{\tau c_s e^{2iNsk_\perp}}{2s|\det(\partial^2 k_\perp)|^{1/2} \sinh(\pi k_B T [2Nsk'_\perp + \psi'_s])} \right)\Bigg|_{\mu, \mathbf{k}^0_\parallel} \tag{60}$$

where the sum over $\mathbf{k}^0_\parallel$ is taken over the belly and neck extrema and $k'_\perp = dk_\perp / d\omega$ is the inverse FS velocity.

We can now use the SPA formula (60) to discuss the coupling $J(N)$ shown in Fig. 18, which was obtained numerically from Eq. (50). First of all, it can be seen from Eq. (60) that the coupling strength at low temperatures decreases approximately as $\propto 1/N^2$, which is in good agreement with the dependence $J(N)$ shown in Fig. 18. Deviations from the $1/N^2$ dependence occur because the derivative of the phase ψ_s' of the Fourier coefficient is very large at the neck extrema [49]. The large value of ψ_s' leads also to a much stronger temperature dependence of the coupling [50] than that predicted by early theories [44–47].

The short oscillation period of 2.6 atomic planes visible in Fig. 18 comes clearly from the neck extrema of the Cu FS. However, in agreement with the simple RKKY theory [52], the SPA formula (60) also predicts a long belly period of 5.7 atomic planes. To understand why this long period is not seen in Fig. 18, we need to examine the amplitudes of both the oscillation periods. There are three factors that determine the amplitude: (1) the curvature of the Cu FS surface given by the Hessian matrix $(\partial^2\phi_s)_{\alpha,\beta} \equiv \partial^2\phi_s/\partial k_{\parallel\alpha}\partial k_{\parallel\beta}$; (2) the inverse FS velocity $k_\perp' = dk_\perp/d\omega$; and (3) the magnitude of the Fourier coefficient c_s at the belly (neck) extremum. The first two factors are just the property of the Cu FS, and these are the only factors that determine the coupling strength in the simple RKKY theory [52]. It turns out [52] that the Cu FS curvature and the inverse FS velocity at the belly and necks are comparable. The RKKY theory thus predicts comparable amplitudes of the belly and necks oscillations. The numerical results are clearly incompatible with this prediction of the RKKY theory. The correct explanation is that the magnitude of the Fourier coefficient c_s is the decisive factor, and we shall now demonstrate that $|c_s|$ is several orders of magnitude larger for the neck than for the belly extremum. Using the results for the matrix Möbius transformation derived in Section 3, we can plot the function $F(\mu, \mathbf{k}_\parallel^0, N)$ as a periodic function of the continuous variable N and thus determine easily its Fourier coefficients. The plots of $F(\mu, \mathbf{k}_\parallel^0, N)$ as a function of N over its periodic interval are shown in Fig. 20. It can be seen that the amplitude of $F(\mu, \mathbf{k}_\parallel^0, N)$ is almost three orders of magnitude larger at the neck than at the belly. The other qualitative difference is a discontinuity in $F(\mu, \mathbf{k}_\parallel^0, N)$ at the neck extremum. The magnitude of $F(\mu, \mathbf{k}_\parallel^0, N)$ (of the Fourier coefficient) is determined by the "magnetic contrast" at the Co/Cu interface, i.e., by the degree of match/mismatch between the majority- and minority-spin bands of Co and Cu. This is illustrated in Fig. 21. It can be seen in Fig. 21(a) that the offset of the Cu and Co bands at the belly is small both for the majority- and minority-spin carriers. This is the reason why the amplitude of the belly oscillation period is small.

The most striking feature seen in Fig. 21(b) is that there is a hybridization gap in the minority-spin band structure of Co, which means that minority-spin carriers in the FM configuration are completely confined in the Cu spacer quantum well. On the other hand, majority-spin carriers are only weakly confined. The fact that there is a bound state at the neck is reflected in the discontinuity of $F(\mu, \mathbf{k}_\parallel^0, N)$ seen in Fig. 20. As already discussed, it

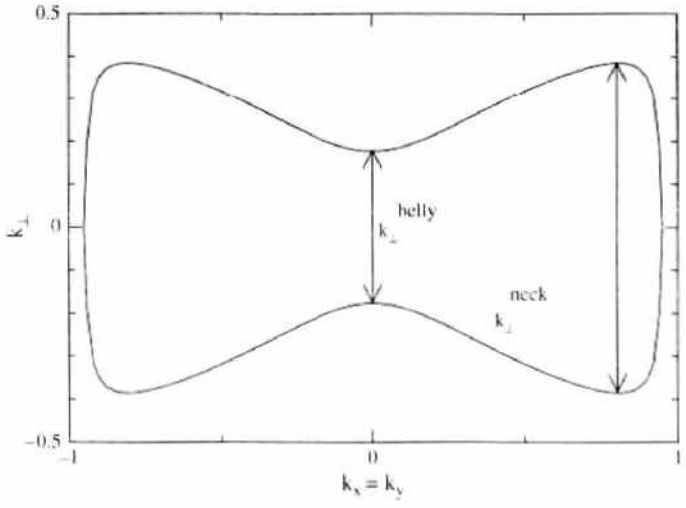


Figure 19. Cross section of the Cu Fermi surface in the direction perpendicular to the layers.

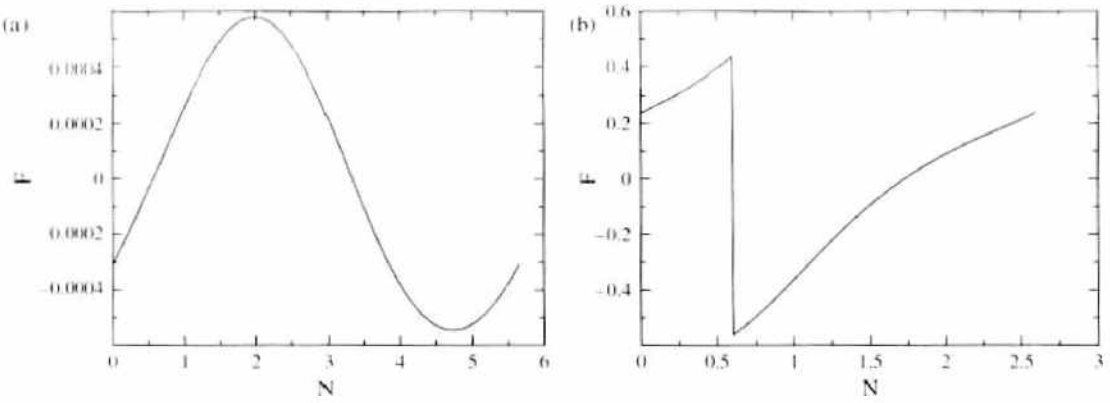


Figure 20. (a) Integrand $F(k_\parallel, \mu, N)$ in Eq. (50) at the belly extremum $k_\parallel = 0$ plotted over the first period $0 \leq N \leq \pi/k$; (b) the same plot for the neck extremum.

is straightforward to Fourier analyze the function $F(\mu, \mathbf{k}_\parallel^0, N)$ and thus determine from the SPA formula (60) separately the contributions to the coupling coming from the belly and the four neck extrema. The results are shown in Fig. 22. As expected, the belly contribution to the coupling is more than two orders of magnitude smaller than that due to the four necks. Combining the neck and belly contributions, we obtain the total coupling in the stationary phase approximation, which is denoted by broken line in Fig. 18. It can be seen that the agreement between the full numerical calculation and SPA is excellent for Cu thickness greater than ≈ 10 atomic planes.

It remains to compare the realistic calculations of the coupling for Co/Cu/Co(001) with the experimental results [35–43]. As already mentioned, the tight-binding calculations [48, 49] are in excellent agreement with the LMTO tight-binding calculations [61–63]. Both sets of calculations also agree with the Korringa-Kohn-Rostoker (KKR) calculations [66, 67] for the same system. However, while the realistic calculations predict that the short (neck) period is dominant, virtually all experiments [35–43] show only the long belly period and hardly any sign of the short period. An exception is the experiment of Weber et al. [68] in which the short period was clearly seen. However, the experimental method used by Weber et al. does not provide a quantitative estimate of the oscillation amplitudes.

The fact that the neck period is not clearly seen in the experiment is not surprising since any small interfacial roughness is likely to average out any short-period oscillation with a period of the order of the interatomic distance. This was initially also the case for Fe/Cr and only perfect samples grown on iron whiskers showed the short-period oscillation [34]. However, what is very surprising in the case of Co/Cu system is that the maximum observed strength of the long period belly oscillation is very large, of the order of one-third of the strength of the calculated neck oscillation. Superficially, one might be satisfied that the

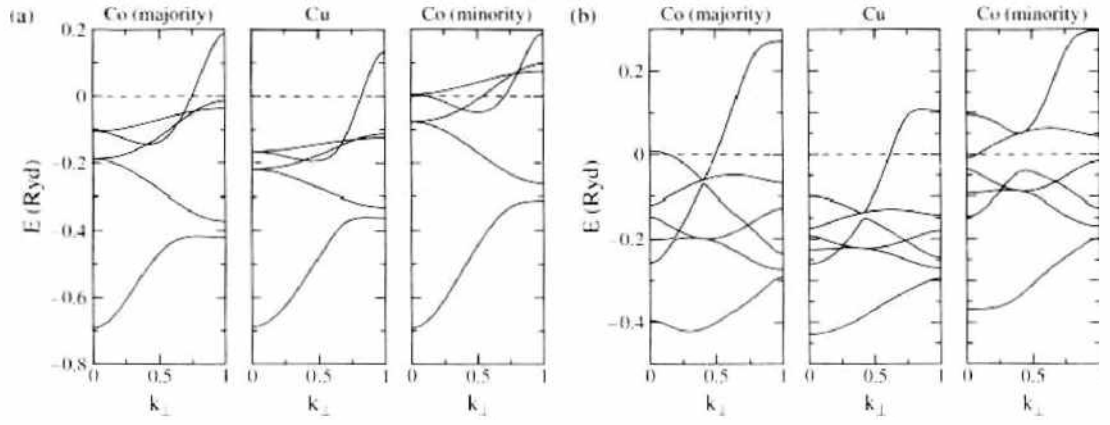


Figure 21. Band structures of Cu and ferromagnetic fcc Co in the direction perpendicular to the layers at (a) the belly and (b) the neck. The Fermi energy μ is denoted by the dashed line.

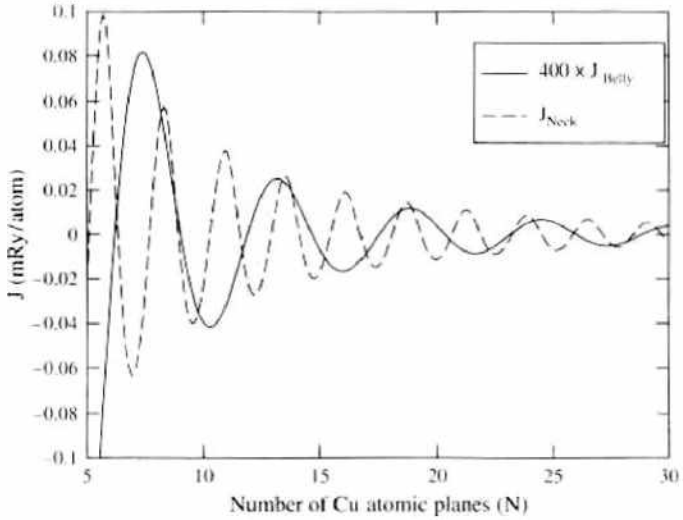


Figure 22. Contributions to the coupling from the belly (continuous line) and the four neck extrema (broken line). Note that the belly contribution is multiplied by a factor of 400.

overall magnitudes of the calculated and observed couplings are comparable but a closer examination reveals a serious problem. We recall that the calculated amplitude of the belly oscillation is more than two orders of magnitude weaker than that of the neck oscillation. While it is quite reasonable to argue that the neck period is suppressed by roughness, it is hard to see how any roughness could enhance the belly period calculated for perfect samples by a factor of the order of a hundred. This would be required to reproduce the observed amplitude of the long-period oscillation. To our knowledge, this problem remains unresolved. We can offer two speculative explanations. One clue might be the dependence of the coupling on the Co thickness. Most of the calculations referred to above were made for semi-infinite Co layers. One exception is the first KKR calculation of Lang et al. [66], which was made for two Co monolayers embedded in a Cu matrix. This shows that the belly oscillation is "revived" for Co monolayers and its amplitude becomes almost comparable with the amplitude of the neck oscillation. The same results were obtained in the tight-binding calculation of Mathon et al. [49]. However, experiments employ relatively thick Co layers and the observed dependence on Co thickness is very weak. It is therefore unlikely that this is the correct explanation of the large amplitude of the observed long-period oscillation.

An alternative explanation is based on the assumption that the short-period neck oscillation is washed out by roughness. However, it is apparent from the SPA formula (60) that not only the fundamental but also higher harmonics contribute to the coupling. These are the terms with $s > 1$ in the sum in Eq. (60). When the coupling is considered as a function of the continuous thickness N, as is assumed in the Möbius transformation method, the period of the first harmonic of the neck oscillation is, of course, one half of the fundamental, that is, 1.3 atomic planes. However, when such a first harmonic oscillation is sampled only on discrete points N corresponding to the actual positions of Cu atomic planes, an oscillation with a period of almost exactly the same length as the belly period is obtained. This is illustrated in Fig. 23.

Assuming that this "long-period" oscillation is more stable with respect to interfacial roughness than the short fundamental neck period, the results shown in Fig. 23 could provide an explanation of the observed long-period ("pseudo-belly") oscillation. The magnitude of the first peak in Fig. 23 is comparable with the maximum observed coupling strength. However, we stress that this hypothesis is highly speculative and requires further investigation.

We have described methods for realistic calculations of the oscillatory exchange coupling with emphasis on applications to the archetypal Co/Cu system. Calculations for other systems have been done but frequently the calculated coupling is much larger than observed. One reason for this discrepancy may be that virtually all calculations assume (sometimes implicitly) that electron interactions in the spacer are negligible. This is obviously not the case for chromium which is known to have a spin density wave in the bulk. In the presence

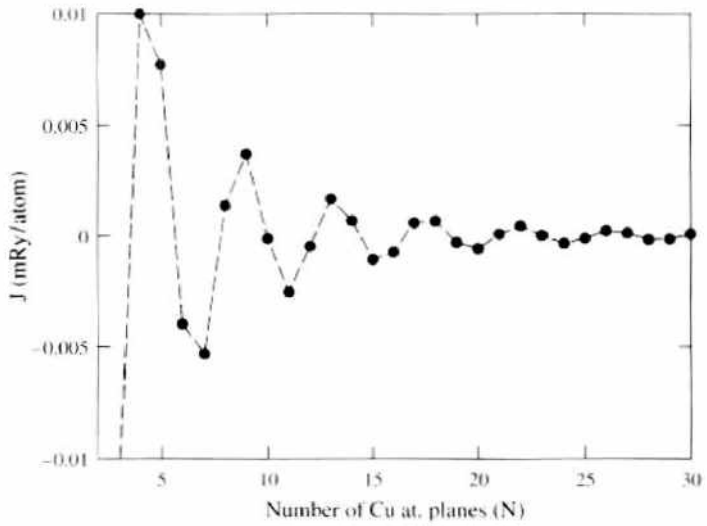


Figure 23. Calculated first harmonic of the neck oscillation sampled on atomic positions of the Cu planes.

of a spin density wave, the calculations of the type described here are clearly inappropriate. This may also be the case for other transition metal spacers even if they do not order spontaneously in the bulk. However, a generalization of the exchange coupling calculations to fully self-consistent calculations for the whole trilayer is numerically very difficult but would be very desirable.

5. BALLISTIC CURRENT-PERPENDICULAR-TO-PLANE MAGNETORESISTANCE

In Section 2, we derived a resistor network model of the CPP GMR. This model is valid in the limit of an elastic mean free path shorter than the thicknesses of all the constituent layers forming a magnetic multilayer, that is, in the classical diffusive limit. However, this conventional point of view, which assumes that GMR arises entirely from interface and bulk impurity scattering, was challenged by Schep et al. [69] who showed that a very large CPP GMR can be obtained in the ballistic limit without any impurity scattering. The classical Boltzmann equation or k-space Kubo formula, which is essentially equivalent to it, cannot be used to discuss the ballistic CPP GMR. This is because the k-space Kubo formula applies to a sample that is infinite in all three dimensions and is, therefore, necessarily in the diffusive limit. The correct treatment of the ballistic limit (the mean mean free path much longer than the whole sample) must be based on the Kubo-Landauer formula (27) of Section 3, which assumes from the outset that the sample is finite in the direction of transport of charge.

The original argument of Schep et al. was based on the Kubo-Landauer formula applied to an infinite superlattice of alternating magnetic and nonmagnetic metallic layers. In this case, the transmission coefficient for each $\boldsymbol{k}_{\parallel}$, required in Eq. (27) of Section 3, is either 1 or 0, depending on whether the corresponding state at the Fermi energy lies in the allowed band of energies or not. One can thus use just the bulk-band structure of an infinite superlattice to apply this "counting" argument to evaluate the conductance. Schep et al. [69] found that the ballistic CPP GMR ratio for Co/Cu superlattices of various compositions is in excess of 100%, which is comparable to but somewhat lower than the highest observed values of the CPP GMR [70]. One reason why the theoretical ballistic CPP GMR is somewhat lower than the observed values may be that the experimental CPP GMR contains a significant contribution due to scattering from imperfections in the bulk of the ferromagnetic layers which was discussed in Section 2. This effect is obviously missing in the ballistic treatment of CPP GMR.

The explanation of a high-ballistic TMR obtained by Schep et al. is simple. The bands of majority-spin carriers of Co at the Fermi surface are very similar to those of Cu since the d bands of majority-spin carriers lie below the Fermi surface (see Fig. 9 of Section 2). It follows that the conductance of majority-spin carriers in the ferromagnetic configuration

is high since the corresponding bands in Co and Cu match very well and there is thus no scattering at the interfaces. On the other hand, the Fermi level lies in the d band of minority-spin carriers in Co and there are no such states in Cu available for transport. In fact, we already showed in Section 4 that a large mismatch between the minority-spin bands in Co and those of either spin in Cu (Fig. 9 of Section 2) leads to the formation of quantum well states in Cu. Since these states are localized in Cu they do not contribute to transport. The conductance in the minority-spin channel is thus low. Similarly, both majority- and minority-spin conductances in the antiferromagnetic configuration of the layers are low. The shunting by the majority-spin channel in the ferromagnetic configuration, discussed in Section 2, then explains a large TMR.

The calculations of Schep et al. for an infinite superlattice were generalized in Ref. [25] to Co/Cu trilayers of finite thickness in the direction perpendicular to the layers. We show in Fig. 24 the dependence of the CPP GMR on the Cu thickness N (number of atomic planes) for a $Co_5/Cu_N/Co_5(001)$ trilayer sandwiched between two Cu semi-infinite leads. These results were obtained from Eq. (27) of Section 3 using the tight-binding parameterization of the Co/Cu band structure described in Section 4. The magnitude of the CPP GMR obtained for trilayers is very similar to that obtained by Schep et al. for an infinite superlattice but a new feature are oscillations of the CPP GMR as a function of Cu thickness. In contrast to the oscillatory exchange coupling discussed in Section 4, the CPP GMR oscillations have a large constant bias equal to the asymptotic value of the GMR reached when the thickness of the Cu spacer layer becomes infinite.

In Fig. 25, we show the dependences of the partial conductances (transmission coefficients) $T^{\sigma}(\boldsymbol{k}_{\parallel})$ in the spin channel σ on $\boldsymbol{k}_{\parallel}$ in the ferromagnetic and antiferromagnetic configurations of a $Co_5/Cu_8/Co_5(001)$ trilayer.

Figure 25 illustrates the band match/mismatch argument given above to explain a large CPP GMR for Co/Cu multilayers. The number of $\boldsymbol{k}_{\parallel}$ channels contributing to the conductance of majority-spin carriers in the ferromagnetic configuration (transmission coefficient ≈ 1) is clearly much larger than that in the minority-spin channel and also much larger that the number of open channels in the antiferromagnetic configuration.

The results for finite Co/Cu trilayers were further extended by Kudrnovsky et al. [71] to the situation when there is intermixing between Co and Cu atoms at the Co/Cu interfaces. They used tight-binding linear muffin-tin orbital method to evaluate the Landauer formula. Interdiffusion at the interfaces was treated using a lateral supercell with supercell dimensions of up to 7×7 atoms. In Fig. 26, we reproduce their results for $Co_5Cu_NCo_5$ trilayers sandwiched between Cu semi-infinite leads. It was assumed in the calculation of Kudrnovsky et al. that there is 15% interdiffusion of Co and Cu atoms at the interfaces.

Two interesting results can be deduced from Fig. 26. First of all, it is clear that the CPP GMR ratio is reduced significantly by the intermixing effect. This was already predicted by

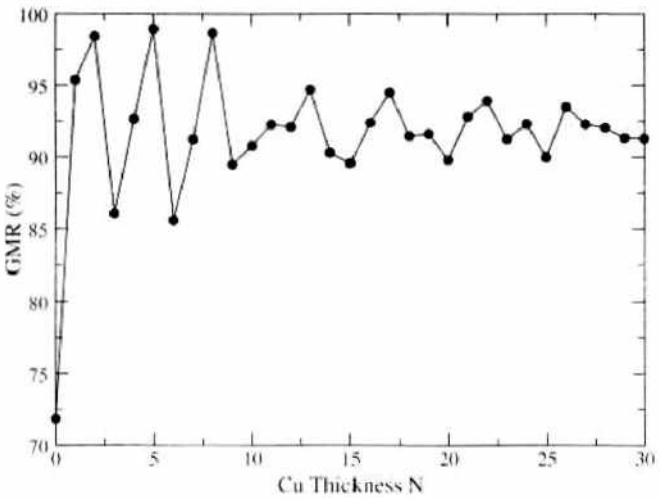


Figure 24. Dependence of the magnetoresistance ratio of a $Co_5/Cu_N/Co_5(001)$ trilayer on the Cu spacer layer thickness N.

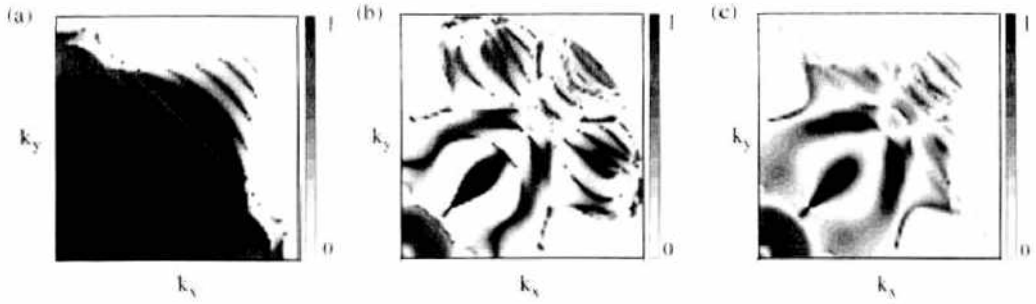


Figure 25. The $k_\parallel$ dependences of the majority-spin $\Gamma^{\uparrow}(k_\parallel)$ (a) and minority-spin $\Gamma^{\downarrow}(k_\parallel)$ (b) conductances in the ferromagnetic configuration, and the $k_\parallel$ dependence of the conductance $\Gamma^{\uparrow\downarrow}(k_\parallel)$ in the antiferromagnetic configuration (c).

Itoh et al. [72] using a single-orbital tight-binding model and the real-space Kubo formula. They showed that the CPP GMR decreases with increasing disorder, whereas the CIP GMR (also calculated by Itoh et al.) increases with increasing disorder.

The second effect of intermixing is that the oscillations of CPP GMR seen in Fig. 24 and in Fig. 26 in the absence of intermixing, are suppressed by disorder. That is most likely the reason why such oscillations have not yet been observed. However, we shall see in Section 6 that they are observable in another CPP experiment (i.e., tunneling magnetoresistance). Since the origin of magnetoresistance oscillations is common to CPP GMR and tunneling MR, we shall now discuss the oscillations of the CPP GMR in some detail. We follow the argument of Ref. [73], where the physical origin of such oscillations was explained using a sc single-orbital tight-binding model of a magnetic trilayer with (001) layering and with a nearest-neighbor hopping t. A trilayer with two semi-infinite magnetic layers separated by a nonmagnetic layer of N atomic planes was considered. It follows that carriers are scattered from a single potential well/barrier having different height for the majority- and minority-spin carriers. To model qualitatively the situation in Co/Cu, it is assumed that there is perfect matching of bands between the ferromagnet and spacer in the majority-spin channel. There is therefore a potential well for minority-spin carriers in the spacer layer but no wells or barriers for the majority-spin carriers. Applying the Kubo-Landauer formula to this model, it is easy to evaluate the CPP GMR. The computed [73] CPP GMR is shown in Fig. 27 for two qualitatively different situations. In the first case [Fig. 27(a)], the Fermi energy $E_F = 0.8$ and the exchange splitting between the majority- and minority-spin bands $\Delta = 0.6$ (all energies are measured in units of hopping t) were chosen so that E_F lies close to the top of the

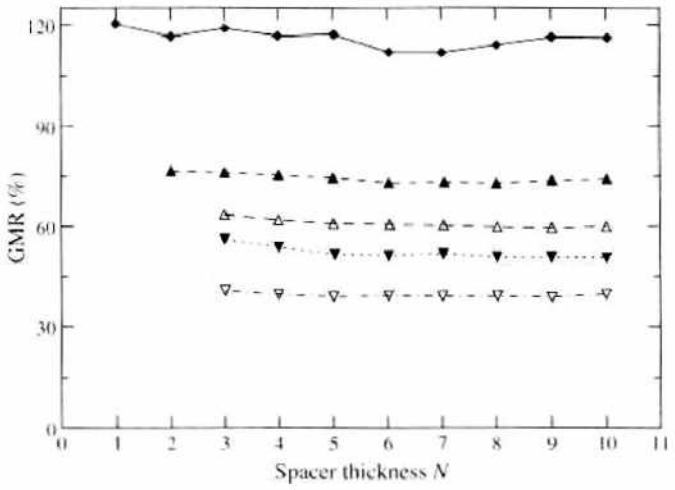


Figure 26. Comparison of trilayers with 15%-interdiffused interfaces with ideal $Co_5/Cu_N/Co_5$ trilayers sandwiched between semi-infinite Cu leads as a function of the spacer thickness N: magnetoresistance ratio for ideal trilayer (diamonds); one of the inner interfaces interdiffused (up triangles); both inner interfaces interdiffused (empty up triangles); two inner and one outer interfaces interdiffused (down triangles); all four interfaces interdiffused (empty down triangles).

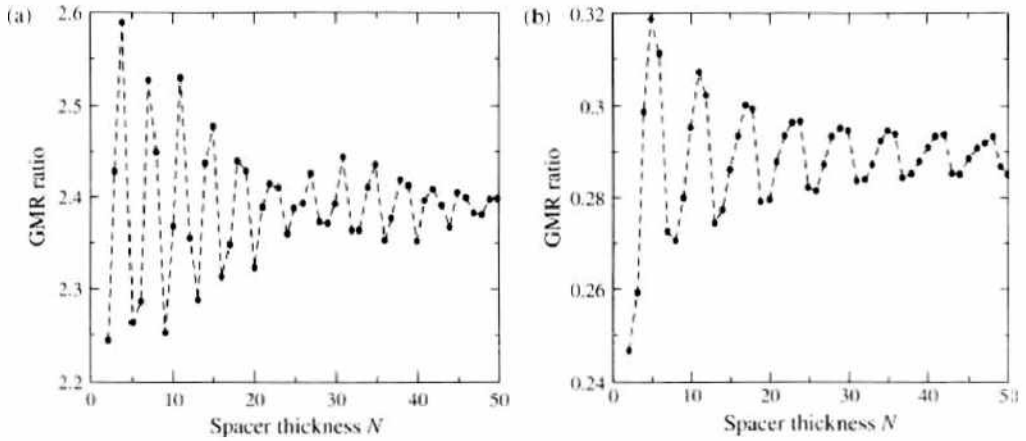


Figure 27. Dependence of the CPP GMR ratio on the nonmagnetic spacer layer thickness for (a) $E_F = 0.8$, $\Delta = 0.6$; (b) $E_F = 0.8$, $\Delta = 0.3$.

spacer potential well in the minority-spin channel. In the second case [Fig. 27(b)], E_F lies some distance above the top of the well ($E_F = 0.8$, $\Delta = 0.3$). In both cases, well-defined oscillations of GMR with spacer thickness N occur. The most remarkable feature seen in Fig. 27(a) are beats which clearly demonstrate the presence of two periods in the case when E_F lies close to the top of the well. We recall that the exchange coupling evaluated for the same model oscillates with a single period determined by the spanning vector of the spacer layer Fermi surface (there is only one such spanning vector for the sc tight-binding model considered here). The presence of two periods indicates that there is a qualitatively different mechanism involved. The single-orbital tight-binding model used here allows us to identify it unambiguously since one can derive for this model a closed expression for the Kubo-Landauer formula [73]. The conductance in a spin channel σ is given by

$$\Gamma^{\sigma} = \frac{4e^2}{h} \sum_{\boldsymbol{k}_{\parallel}} \frac{\sin^2(k_{\perp}a)\,\mathrm{Im}\,g_L^{\sigma}\,\mathrm{Im}\,g_R^{\sigma}}{|\sin(N+1)k_{\perp}a - (g_L^{\sigma}+g_R^{\sigma})\sin(Nk_{\perp}a) + g_L^{\sigma}g_R^{\sigma}\sin(N-1)k_{\perp}a\,|^2} \tag{61}$$

where $k_{\perp}$ is the wave vector perpendicular to the layers, a is the lattice constant, g_L^{σ}, g_R^{σ} are the one-electron surface Green functions for the trilayer cut between the atomic planes L and R, and all energies are again measured in units of the hopping t ($t = 1$).

The key feature of the total conductance clearly seen in Eq. (61) is that all the partial conductances $\Gamma^{\sigma}(\boldsymbol{k}_{\parallel})$ in individual $\boldsymbol{k}_{\parallel}$ channels are periodic functions of the spacer layer thickness N. It follows that the total conductances (and, therefore, GMR) must oscillate as a function of the spacer thickness. It should be noted that the periodic behavior of the conductance, seen explicitly in Eq. (61), is just a consequence of the general periodic (quasi-periodic) property of one-electron Green functions for a layer system discussed in Sections 3 and 4.

Since the conductance $\Gamma^{\sigma}(\boldsymbol{k}_{\parallel})$ is a periodic function of the spacer thickness it can be expanded in a Fourier series

$$\Gamma^{\sigma} = \sum_{\boldsymbol{k}_{\parallel}} \sum_{r} c_r^{\sigma}(\boldsymbol{k}_{\parallel}) \exp(2irNk_{\perp}(\boldsymbol{k}_{\parallel}, E_F)a \tag{62}$$

where $c_r^{\sigma}(\boldsymbol{k}_{\parallel})$ is the Fourier coefficient of $\Gamma^{\sigma}(\boldsymbol{k}_{\parallel})$. This allows us to isolate the oscillatory part of the conductance and apply to it asymptotic expansion methods valid for large spacer thickness. This approach was already explained in Section 4 in the case of oscillatory exchange coupling. It is based on the following argument. For large spacer thickness N, the imaginary exponential in Eq. (62) oscillates rapidly as a function of $\boldsymbol{k}_{\parallel}$ and, therefore, the contributions to the sum in Eq. (62) from different $\boldsymbol{k}_{\parallel}$ tend to cancel. There are only two situations in which the cancellation does not occur: (1) in the vicinity of a point $\boldsymbol{k}_{\parallel}^{0}$ at which the phase of $k_{\perp}(\boldsymbol{k}_{\parallel}, E_F)$ is stationary; (2) in the vicinity of a boundary at which the sum over $\boldsymbol{k}_{\parallel}$ terminates abruptly. It follows that one needs to identify all the stationary points $\boldsymbol{k}_{\parallel}^{0}$ of $k_{\perp}(\boldsymbol{k}_{\parallel}, E_F)$ and

all the cutoff points $\boldsymbol{k}_\parallel^{cp}$ at which $\Gamma^\sigma(\boldsymbol{k}_\parallel)$ vanishes. All possible oscillation periods of the conductance (GMR) are then given by $\pi/k_\perp(\boldsymbol{k}_\parallel^0, E_f)$ and $\pi/k_\perp(\boldsymbol{k}_\parallel^{cp}, E_f)$.

Case 1 requires no special discussion since the periods coming from the stationary points of $k_\perp(\boldsymbol{k}_\parallel, E_F)$ are exactly the same as the periods of the oscillatory exchange coupling obtained in Section 4. Case 2, however, is new and has no analog in the theory of oscillatory exchange coupling. It follows from Eq. (61) that the boundary in the $\boldsymbol{k}_\parallel$ space at which $\Gamma^\sigma(\boldsymbol{k}_\parallel)$ vanishes is defined by

$$\operatorname{Im} g_L^\sigma = 0 \quad \text{and/or} \quad \operatorname{Im} g_R^\sigma = 0 \tag{63}$$

where "or" applies in the antiferromagnetic configuration. This has a simple physical interpretation. In the CPP geometry the conductance in a channel $\boldsymbol{k}_\parallel$ vanishes when carriers at E_F with that particular value of $\boldsymbol{k}_\parallel$ become totally reflected from either the left or right ferromagnet. That happens when the state $(E_f, \boldsymbol{k}_\parallel)$ lies outside the band of the left (right) ferromagnet, in which case the corresponding spectral density $\operatorname{Im} g_L^\sigma (\operatorname{Im} g_R^\sigma)$ vanishes.

While it is straightforward to identify the oscillation periods, asymptotic expansions that give the amplitudes of the corresponding oscillations are more difficult for the cutoff periods. In the case of oscillations arising from stationary points of the spacer layer Fermi surface, the application of a stationary phase approximation proceeds along the same lines as that for the oscillatory exchange coupling (Section 4). The amplitude of such oscillations [73] is determined by the curvature of the spacer Fermi surface and by the mismatch between the ferromagnet and spacer bands at the stationary point $\boldsymbol{k}_\parallel^0$. The oscillation amplitude decreases with the spacer layer thickness as $\propto 1/N$. This is slower than for the oscillatory exchange coupling because, in the latter case, there is an additional factor $1/N$ which arises from the energy integration. This is absent for GMR since transport takes place at the Fermi level.

Asymptotic expansions for the cutoff periods are much more difficult and have been done only for a parabolic band model of the magnetic trilayer [73]. They show that the period is determined by the depth of the spacer potential well in the minority-spin band, assuming again perfect matching in the majority-spin band, and the amplitude decays as $\propto 1/N^{3/2}$.

Although oscillations of the CPP GMR have not yet been seen experimentally, they are an important feature of perpendicular transport since their observation would be direct evidence for phase coherence of the wave functions across the whole multilayer (i.e., a clear test of the ballistic limit).

The conventional analysis of current CPP GMR experiments [70] seems to indicate that they are mostly in the diffusive regime to which the simple resistor model of Section 2 applies. However, this conclusion is somewhat controversial. There is no doubt that quantum reflections from interfaces that can only be treated realistically by theories based on the Kubo-Landauer formula make an important, if not the most important, contribution to the CPP GMR. This is clear from the results of Itoh et al. [72] and Kudrnovsky et al. [71], which show that interfacial roughness (impurities) not only do not give rise to CPP GMR but, in fact, reduce the value of GMR due to reflections from perfect interfaces. However, the problem is that the contribution of reflections from perfect interfaces to the CPP GMR can also be modeled phenomenologically by a resistor model, provided phase coherence from one layer to the next is broken, which happens in the presence of an even quite weak disorder [74]. Therefore, the fact that a simple resistor model describes well most experimental results does not mean that CPP GMR is governed by interfacial roughness, as is assumed in the classical treatment based on the Boltzmann equation. Detailed discussion of both theoretical and experimental results on CPP GMR is given in a comprehensive survey article by Gijs and Bauer [14].

6. TUNNELING MAGNETORESISTANCE

Tunneling from normal metals, including ferromagnets, to superconductors is a well-developed field and there is an excellent review given by Meservey and Tedrow [75] covering this subject. Tunneling between two ferromagnetic electrodes and its dependence on the relative orientation of the magnetizations of the left and right electrodes, i.e., the tunneling magnetoresistance (TMR) was first observed by Julliere [76] and Maekawa and Gäfvert [77].

Julliere was also first to estimate the magnitude of the tunneling magnetoresistance using the classical theory of tunneling which had been formalized earlier by Bardeen [78]. However, these early experiments gave the TMR of only a few percent and were often not reproducible. Also the effect quickly disappeared with increasing temperature and under an applied bias higher than a few millivolts. Over the next twenty years no real progress has been made until 1995 when Moodera et al. showed [79] that a large TMR over 10% could be obtained consistently and reproducibly at room temperature with junctions based on Al_2O_3 barrier. Similar results were obtained by Miyazaki and Tezuka [80]. We show in Fig. 28 more recent results of Moodera et al. [81] for a $Co/Al_2O_3/Ni_{80}Fe_{20}$ junction which give a TMR of some 27% at low temperature and 20% at room temperature. The results shown in Fig. 28 can be understood as follows. At high magnetic fields H the two ferromagnetic films have their magnetizations M (indicated by arrows) aligned in the direction of H. Upon reversing the field, M_{NiFe} (with a lower coercive field) aligns itself in the new field direction, whereas M_{Co} (with a higher coercive field) remains magnetized in the original H direction, creating antiparallel arrangement of the two magnetic layer moments. As H is further increased, M_{Co} also aligns in the new field direction, restoring parallel orientation of the two magnetic moments. It can be seen that the resistance of the junction is lower for the parallel configuration of the magnetic layer moments than for the antiparallel configuration and this gives rise to a TMR. Denoting the conductance of the tunneling junction in the antiparallel (antiferromagnetic) configuration by Γ_{AF} and the conductance in the parallel (ferromagnetic) configuration by Γ_{FM}, we can define the "pessimistic" TMR ratio by

$$TMR = \frac{\Gamma_{AF}^{-1} - \Gamma_{FM}^{-1}}{\Gamma_{AF}^{-1}} \tag{64}$$

The values of the TMR quoted above were obtained using this pessimistic definition. The optimistic magnetoresistance ratio, we defined in Section 2 for the GMR, is also used in the literature.

The reasons why all the earlier attempts at obtaining a high TMR were unsuccessful are essentially due to growth problems and are discussed in detail in a recent review [82]. Moodera's demonstration of a reproducible high TMR was followed by a hectic experimental activity (see Ref. [82]) but, until very recently, virtually all experiments were made on junctions with an Al_2O_3 barrier. In spite of this great effort only a very modest improvement in TMR has been achieved and the current record for junctions based on an Al_2O_3 barrier stands at about 60–70% (optimistic TMR ratio). The TMR experiments can be quite successfuly interpreted using Julliere's model and it is easy to understand within this model why there is an upper limit on the TMR of junctions based on allumina barriers. We shall,

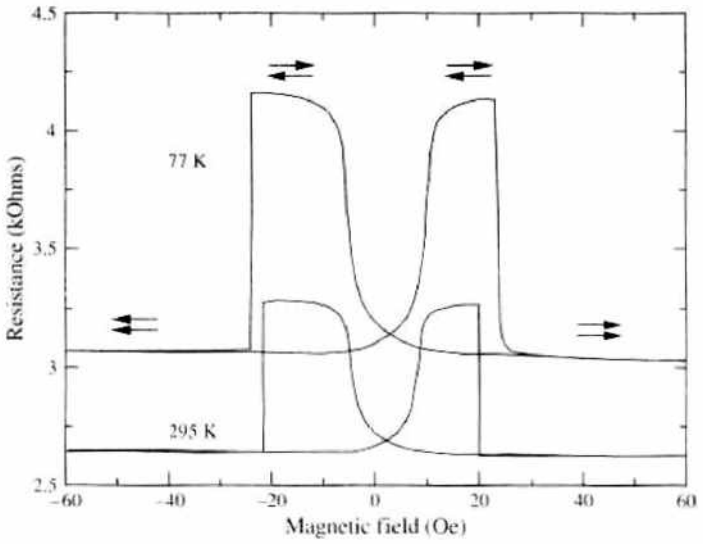


Figure 28. Resistance of a magnetic tunneling junction as a function of the applied magnetic field.

therefore, first derive Julliere's formula and then discuss its application to alumina-based junctions.

Consider tunneling between two normal nonmagnetic metals separated by a thin insulating barrier. In the absence of an applied voltage, the Fermi levels of the two electrodes must be equal (no current flows). An applied voltage V_b shifts the energy levels in one of the electrodes relative to the other by eV_b. This amounts to an energy shift eV_b between the two Fermi levels that gives rise to a tunneling current. Using the classical theory of tunneling [78], one evaluates the current $I(V_b, \omega)$ flowing at a given energy ω between the left (L) and right (R) electrodes from Fermi's golden rule. This gives [75]

$$I(V_b, \omega) \propto |\, t(\omega)\,|^2\, D_L(\omega - eV_b) D_R(\omega)[f(\omega - eV_b) - f(\omega)]d\omega \tag{65}$$

where $D_L(\omega)$, $D_R(\omega)$ are the densities of states of the left and right electrodes, $|t(\omega)|^2$ is the square of the tunneling matrix element, and $f(\omega)$ is the Fermi function. Strictly speaking, Eq. (65) applies only to a one-dimensional system [78]. However, it is usually assumed [75, 76] that $D_L(\omega)$, $D_R(\omega)$ are the total (three-dimensional) densities of states of the left and right electrodes. It is further assumed [75] that the tunneling matrix element is independent of the energy over the relevant energy range $\approx eV_b$. The total tunneling current is then obtained by integrating Eq. (65) with respect to the energy

$$I(V_b, \omega) \propto |\, t\,|^2 \int_{-\infty}^{\infty} D_L(\omega - eV_b) D_R(\omega)[f(\omega - eV_b) - f(\omega)]d\omega \tag{66}$$

In the low-bias regime, we can divide Eq. (66) by eV_b and take the limit $V_b \to 0$. Because at low temperatures $\lim_{eV_b \to 0}[f(\omega - eV_b) - f(\omega)]/eV_b = \delta(\omega - E_F)$, where E_F is the Fermi energy, we find

$$\frac{dI}{dV_b} \propto |\, t\,|^2\, D_L(E_F) D_R(E_F) \tag{67}$$

where $dI/dV_b = \Gamma$ is the zero-bias conductance of the tunneling junction. It is clear from Eq. (67) that in the low-bias regime, $I \propto V_b$, that is, the junction displays an Ohmic behavior. Equation (67) was applied by Julliere [76] to discuss the magnitude of the observed tunneling magnetoresistance. In the case of magnetic electrodes, the densities of states for carriers with spin parallel and antiparallel to the magnetization are different because the energy bands of up- and down-spin carriers are split by the exchange interaction. This is illustrated in Fig. 8 of Section 2.

To take into account the dependence of the tunneling current on the relative orientation of the magnetic moments of the two electrodes, Julliere made an additional assumption that the spin of carriers is conserved in tunneling. As discussed in Section 2, spin-flip scattering is usually an inelastic process and, therefore, this is a good approximation at low temperatures. It then follows that tunneling of up- and down-spin carriers are two quite independent processes, i.e., the tunneling current flows in the up- and down-spin channels as if in two wires connected in parallel. We have already used such a two-current model in Section 2 to interpret the closely related giant magnetoresistance effect. Using Eq. (67), we then find that the conductance in the antiferromagnetic configuration $\Gamma_{AF} \propto D_L^{\uparrow} D_R^{\downarrow} + D_L^{\downarrow} D_R^{\uparrow}$, and the conductance in the ferromagnetic configuration is given by $\Gamma_{FM} \propto D_L^{\uparrow} D_R^{\uparrow} + D_L^{\downarrow} D_R^{\downarrow}$. Introducing two parameters P_1 and P_2 characterizing the left and right electrodes $P_1 = [D_L^{\uparrow}(E_F) - D_L^{\downarrow}(E_F)]/[D_L^{\uparrow}(E_F) + D_L^{\downarrow}(E_F)]$ and $P_2 = [D_R^{\uparrow}(E_F) - D_R^{\downarrow}(E_F)]/[D_R^{\uparrow}(E_F) + D_R^{\downarrow}(E_F)]$, we can rewrite the TMR ratio (64) in the form most commonly used to analyze experimental results (Julliere's formula),

$$TMR = \frac{2P_1 P_2}{1 + P_1 P_2} \tag{68}$$

The parameters P_1 and P_2 can be determined independently from measurements of the tunneling current in FM-I-S junctions, in which one of the electrodes is a ferromagnet and the other is a superconductor [75]. An analysis of such experiments based on Eq. (66) in which $D_R(\omega)$ is now the density of states of the superconductor yields the values of $D_f^{\uparrow}(E_F)$,

$D_L^{\downarrow}(E_F)$ and, hence, P for each particular ferromagnet. Since it follows from Eq. (67) that the tunneling current of up-spin carriers in an FM-I-S junction is proportional to $D_L^{\uparrow}(E_F)$ and that of down-spin carriers is proportional to $D_L^{\downarrow}(E_F)$, the parameter P can be interpreted as the relative spin polarization of carriers tunneling from a ferromagnet. It is assumed here that $D^{\sigma}(\omega)$ is approximately independent of energy and can be taken as $D^{\sigma}(E_F)$ outside the integral in Eq. (66). The measured values of P for Fe, Co, and Ni are [82] $P_{\text{Fe}} \approx 40\%$, $P_{\text{Co}} \approx 35\%$, and $P_{\text{Ni}} \approx 23\%$. When these values of P are substituted in Eq. (68), they give a good estimate of the observed TMR in junctions based on Fe, Co, Ni with an allumina barrier. However, there is one serious problem with Julliere's formula. It gives a wrong sign of the spin polarization. Examination of the densities of states of Fe, Co, and Ni (see Fig. 8 of Section 2 for Co, Fe) shows that the spin polarization for Co and Ni deduced from the density of states should have negative sign (the density of states of the minority-spin carriers at E_F is higher than that of the majority-spin carriers). However, the observed spin polarizations deduced from measurements of the tunneling current in FM-I-S junctions are positive for Co and Ni (as well as for Fe). We shall return to this problem later.

It is easy to understand the apparent success of Julliere's formula in predicting the right magnitude of TMR. It is entirely due to factoring of the tunneling current into a product of factors depending separately on the properties of the left and right electrodes. Because these properties remain the same for FM-I-S and FM-I-FM junctions employing the same insulating barriers, the polarizations obtained from FM-I-S junctions must be consistent with the TMR ratio for FM-I-FM junctions based on the same ferromagnets. The real reason why Julliere's formula is valid for junctions with allumina barriers (apart from the wrong sign of P) is that all such barriers are amorphous and tunneling is thus incoherent, that is, the wave vector of carriers $\boldsymbol{k}_{\parallel}$ parallel to the layers is not conserved. It follows that there is no need to match the carrier wave functions across the whole junction and Julliere's treatment of tunneling is thus adequate. Using Julliere's formula, it is easy to understand why only a very modest improvement of the TMR ratio has been achieved over the last 10 years. The failure to improve the TMR ratio is due, entirely, to the fact that virtually all experiments were performed on junctions with amorphous allumina barriers. Since it follows from Julliere's formula that the TMR ratio for a given combination of the left and right ferromagnetic electrodes is determined completely by the spin polarizations of the two ferromagnets, there is an upper limit on the TMR determined by the densities of states of up- and down-spin carriers at the Fermi level. Since the densities of states are fixed for conventional transition metal ferromagnets, the scope for improving TMR in junctions with amorphous barriers is very limited. Similarly, realistic modeling of an amorphous barrier is virtually impossible and there is, therefore, very little that can be achieved theoretically for incoherent tunneling. On the other hand, for a perfectly ordered epitaxial junction, tunneling is determined by the matching of wave functions for each $\boldsymbol{k}_{\parallel}$ across the whole junction. This can be controlled predictively through the choice of materials for the electrodes and the barrier and also by the choice of crystal orientation of the junction. Epitaxial junctions with coherent tunneling are, therefore, much more interesting. Moreover, we shall see that their TMR can be computed rigorously using a fully realistic band structure of all the components of the junction.

6.1. Simple Models of Coherent TMR

The need for matching the wave functions across the junction was first recognized by Slonczewski [57]. He described the ferromagnet by two simple parabolic bands (one for up-spin, the other for down-spin) shifted rigidly with respect to one another by an amount Δ (the exchange splitting). Assuming that $\boldsymbol{k}_{\parallel}$ is conserved, he then solved the Schrödinger equation for the wave functions of up- and down-spin carriers tunneling across a rectangular barrier and determined the current from the current operator. The principal result of Slonczewski's calculation is that the polarization P of tunneling carriers now depends on the height of the barrier U through an imaginary wave vector $i\kappa$ in the barrier defined by $\hbar\kappa = [2m(U - E_F)]^{1/2}$

$$P = \frac{k^{\uparrow} - k^{\downarrow}}{k^{\uparrow} + k^{\downarrow}} \frac{\kappa^2 - k^{\uparrow}k^{\downarrow}}{\kappa^2 + k^{\uparrow}k^{\downarrow}} \tag{69}$$

Here, $k^{\uparrow}$, $k^{\downarrow}$ are the Fermi wave vectors in the up- and down-spin bands. Using the result $k^{\uparrow} \propto D^{\uparrow}(E_F)$, $k^{\downarrow} \propto D^{\downarrow}(E_F)$, which holds for parabolic bands, it is easy to see that the first factor $(k^{\uparrow} - k^{\downarrow})/(k^{\uparrow} + k^{\downarrow})$ is the polarization obtained in the classical theory of tunneling but the second factor $A = (\kappa^2 - k^{\uparrow}k^{\downarrow})/(\kappa^2 + k^{\uparrow}k^{\downarrow})$ is new. Since κ ranges from 0 (low barrier) to ∞ (high barrier), we have $-1 < A < 1$. It follows that, for a high barrier, the polarization P given by Eq. (69) reduces to Julliere's result but P can even change sign when the barrier is low.

While Slonczewski's model treats the ferromagnet/barrier interface more realistically than the classical theory of tunneling, the method of matching the wave functions is not easily generalizable beyond a simple parabolic band. It is for this reason that virtually all recent theories of TMR are based on the general Kubo-Landauer formalism described in Section 3. We shall first illustrate its application to TMR for a single-orbital tight-binding Hamiltonian on a simple cubic lattice with nearest-neighbor hopping t^{bulk}. Using a tight-binding description, one has two options how to model a tunneling junction. Harrison proposed [17] to model tunneling by turning off the tight-binding hopping integral between two neighboring atomic planes. This is a good approximation to tunneling across a vacuum gap. The other option is to consider tunneling between two ferromagnetic electrodes through a barrier which consists of N atomic planes of an insulator with an on-site potential V_{ins} chosen so that the Fermi level E_F lies outside its band of allowed energies. We shall first discuss briefly TMR due to tunneling across a vacuum gap. We assume that the electrodes are parallel to a (001) plane, and the surface atomic plane of the left electrode (atomic plane L) is connected to the surface plane of the right electrode (plane R) by a nearest-neighbor hopping integral t_{LR}. Because the hopping integral t_{LR} is determined by the overlap of the wave functions in the left (L) and right (R) surface planes, we can model tunneling by decreasing t_{LR} gradually from t_{bulk} (absence of vacuum gap) to zero (infinitely wide gap). It follows from Eq. (27) of Section 3 that the total conductance in a spin channel σ takes the form

$$\Gamma^{\sigma} = \frac{4e^2 t_{\text{LR}}^2}{h} \sum_{\boldsymbol{k}_{\parallel}} \frac{\text{Im}\, g_L^{\sigma}(E_F, \boldsymbol{k}_{\parallel})\, \text{Im}\, g_R^{\sigma}(E_F, \boldsymbol{k}_{\parallel})}{|1 - t_{\text{LR}}^2 g_L^{\sigma}(E_F, \boldsymbol{k}_{\parallel}) g_R^{\sigma}(E_F, \boldsymbol{k}_{\parallel})|^2} \tag{70}$$

where $g_L^{\sigma}(E_F, \boldsymbol{k}_{\parallel})$ and $g_R^{\sigma}(E_F, \boldsymbol{k}_{\parallel})$ are the surface Green functions of the left- and right-isolated electrodes. To calculate the TMR ratio, it is necessary to determine from Eq. (70) the conductances $\Gamma^{\uparrow}$ and $\Gamma^{\downarrow}$ of up- and down-spin carriers in the ferromagnetic and antiferromagnetic configurations of the magnetic electrodes.

Equation (70) has a simple physical interpretation. Since $-(1/\pi)\text{Im}\, g_L^{\sigma}(E_F, \boldsymbol{k}_{\parallel})$ and $-(1/\pi)\text{Im}\, g_R^{\sigma}(E_F, \boldsymbol{k}_{\parallel})$ are the one-dimensional surface densities of states (spectral densities) in a channel $(\sigma, \boldsymbol{k}_{\parallel})$ for the isolated left and right electrodes, the tunneling current in every channel $(\sigma, \boldsymbol{k}_{\parallel})$ is proportional to the product of the one-dimensional densities of states of the two electrodes, but the product is scaled by the denominator in Eq. (70). As in Slonczewski's model, the denominator in Eq. (70) describes the mutual interaction of the two electrodes due to an overlap of the carrier wave functions.

A close link between the rigorous Keldysh-Landauer theory and the conventional classical theory of tunneling is now obvious. When the vacuum gap is large $t_{\text{LR}} \approx 0$, the denominator in Eq. (70) can be set equal to unity. The conductance in each $(\sigma, \boldsymbol{k}_{\parallel})$ channel is then proportional to the product of the one-dimensional densities of states of the left and right electrodes. The total conductance is the convolution of the one-dimensional densities of states over $\boldsymbol{k}_{\parallel}$. This is the result of Bardeen's theory of tunneling. Moreover, since the multiplicative factor t_{LR}^2 in the numerator of Eq. (70) cancels out in the TMR ratio, the tunneling magnetoresistance becomes independent of the gap width. This again is in agreement with the classical theory of tunneling. We stress, however, that Julliere's formula is different because it involves three-dimensional densities of states. We shall see that to deduce the Julliere's formula from the Kubo-Landauer formula one needs to consider the effect of disorder that breaks the conservation of $\boldsymbol{k}_{\parallel}$.

We shall now discuss tunneling through an insulating barrier using the same single-orbital tight-binding Hamiltonian as for tunneling through vacuum gap. Such a model has been used to determine the dependences of the TMR on the height and width of the insulating

barrier [83] and also to investigate the effect on TMR of disorder at the ferromagnet/ insulator interface [84] and of impurities in the insulator [85].

In the case of coherent tunneling through a high barrier, the Keldysh-Landauer formula takes the following simple form [83]

$$\Gamma^{\sigma} \approx \left(\frac{4e^2}{h}\right) e^{-2\kappa_0 aN} \sum_{\boldsymbol{k}_\parallel} \frac{\operatorname{Im} g_L^{\sigma} \operatorname{Im} g_R^{\sigma}}{|1 - (g_L^{\sigma} + g_R^{\sigma}) e^{-\kappa_0 a} + g_L^{\sigma} g_R^{\sigma} e^{-2\kappa_0 a}|^2} \tag{71}$$

where $g_L^{\sigma}(E_F, \boldsymbol{k}_\parallel)$ and $g_R^{\sigma}(E_F, \boldsymbol{k}_\parallel)$ are again the surface Green's functions of the isolated left and right electrodes, κ_0 is the value of the imaginary wave vector κ in the barrier averaged over the two-dimensional Brillouin zone, a is the lattice constant, and all energies are measured in units of the bulk hopping t_{bulk}. Equation (71) holds in the limit $\kappa_0 a \gg 1$, that is, when electrons in the barrier are strongly attenuated over distances of several lattice constants.

The structure of Eq. (71) is virtually identical to that of Eq. (70) for tunneling across a vacuum gap. The multiplicative factor $e^{-2\kappa_0 aN}$ again cancels out from the TMR ratio and since the denominator in Eq. (71) tends to unity in the limit of a high barrier $e^{-2\kappa_0 aN} \ll 1$, the TMR ratio approaches the same saturation value as for tunneling across a vacuum gap in the limit of a wide gap $t_{LR}/t^{\text{bulk}} \ll 1$. It can be seen from Eq. (71) that the TMR ratio in the high-barrier limit $\kappa_0 a \gg 1$ becomes independent of the barrier height and width. Neither is true when the barrier is low. The dependence of of the optimistic ratio R_{TMR} on V_{ins}/W (W is the band width) determined numerically [83] from Eq. (71) is shown in Fig. 29 for three thicknesses of the insulating barrier $N = 1$, 3, and 5 atomic planes. The values of the ferromagnet parameters were chosen to mimic a junction with Co electrodes which will be discussed later using a fully realistic band structure of Co. It can be seen from Fig. 29 that the R_{TMR} increases with increasing barrier height and reaches a saturation value for barrier heights V_{ins} of the order of the band width (saturation is reached most rapidly for the narrow barrier $N = 1$).

The dependence of the TMR ratio on the barrier width N (measured in atomic planes) is shown in Fig. 30 for three heights of the tunneling barrier: $V_{\text{ins}}/W = 1.0$, 2.0, and also for a very low barrier $V_{\text{ins}}/W = 0.58$ (E_F just outside the insulator band). As expected from Eq. (71), the dependence of TMR on N is weak for a high potential barrier ($V_{\text{ins}}/W = 2.0$) but the TMR ratio decreases rapidly with N when the insulating barrier is very low ($V_{\text{ins}}/W = 0.58$).

6.2. Influence of the Electrode Band Structure on TMR

Although Julliere's formula and Keldysh-Landauer theories based on a one-band model provide useful insight they can predict neither the magnitude of TMR nor the correct sign of the

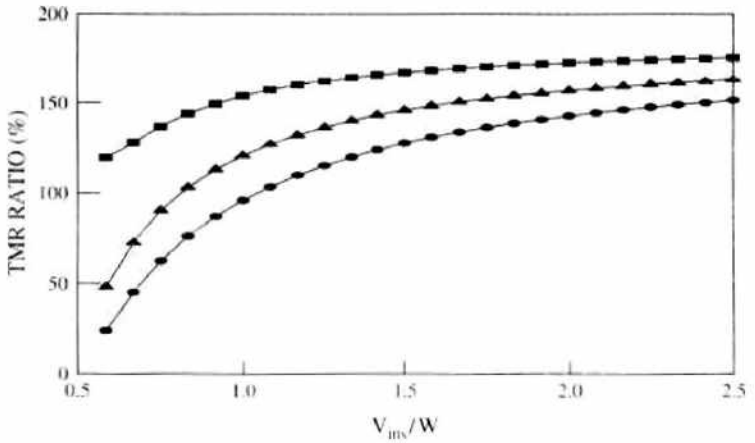


Figure 29. Dependence of the tunneling magnetoresistance on the height V_{ins} of a barrier for barriers whose thicknesses are one (triangles), three (squares), and five (circles) atomic planes. Single-orbital tight-binding band with band width W.

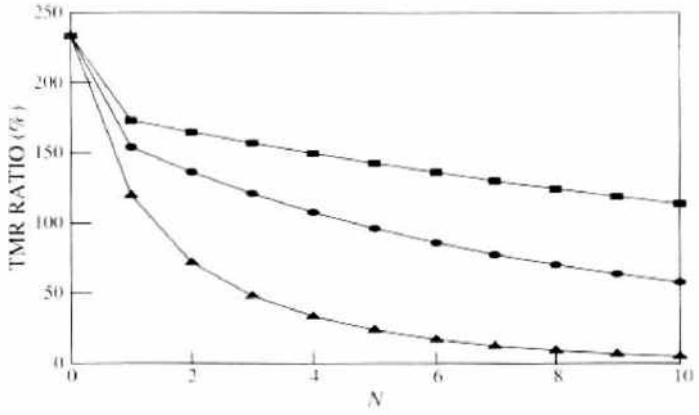


Figure 30. Dependence of the tunneling magnetoresistance on the number of atomic planes in a barrier for three heights of the barrier V_{ins} measured in units of the band width W: $V_{ins}/W = 2.0$ (squares); $V_{ins}/W = 1.0$ (circles); $V_{ins}/W = 1.0$ (triangles).

spin polarization. Key to this problem lies in the multiorbital band structure of the ferromagnetic electrodes. It was pointed out by Stearns [86] long ago that it is conduction rather than d electrons that tunnel from transition metal ferromagnets. One should not, therefore, look at the total density of states which is dominated by d electrons but only at those portions of the Fermi surface that are of s-p character. She then argued that such portions of the Fermi surface lead to the correct sign of the tunneling current. Numerical evaluation of the Kubo-Landauer formula using fully realistic band structure of the ferromagnetic electrodes allows us to test this idea. There are three such calculations dealing with tunneling between Co electrodes across a vacuum gap [83], tunneling between Fe electrodes through a simple step barrier [87], and tunneling from Fe and Co to an s-band through a barrier modelled by two s-bands separated by a gap [88].

The calculation of tunneling between Co electrodes across a vacuum gap [83] is based on numerical evaluation of the Kubo-Landauer formula (27) of Section 3 using a tight-binding parameterization of Co bands fitted to an *ab initio* band structure [64]. As already discussed, one can model within the tight-binding scheme tunneling across a vacuum gap by turning off gradually the tight-binding hopping matrix between the Co electrodes [83]. The conductances $\Gamma^{\uparrow}$, $\Gamma^{\downarrow}$ of the majority- and minority-spin carriers in the ferromagnetic and antiferromagnetic configurations of the Co(001) junction were determined in Ref. [83] as functions of electron hopping across the vacuum gap. We have used these data to plot in Fig. 31 the dependence of the spin polarization of tunneling electrons $P = (\Gamma^{\uparrow} - \Gamma^{\downarrow})/(\Gamma^{\uparrow} + \Gamma^{\downarrow})$ on the width of the vacuum gap. The width of the gap is characterized by the reciprocal of hopping t between s orbitals measured in units of bulk hopping in Co. For a small vacuum gap of the order of the lattice constant ($1/t \approx 1$), the conductance is dominated by d electrons and P has the "wrong" sign $P < 0$ consistent with the total density of states argument of the classical theory of tunneling [76]. However, there is a rapid crossover to $P > 0$ as the width of the gap increases. It can be seen from Fig. 31 that the calculated P for Co(001) junction not only has the correct sign in the tunneling regime $1/t \gg 1$ but its magnitude 30–40% is in excellent agreement with the observed [82] $P \approx 35\%$. The corresponding value of the calculated optimistic TMR ratio is ≈65%, which is again in good agreement with experiment.

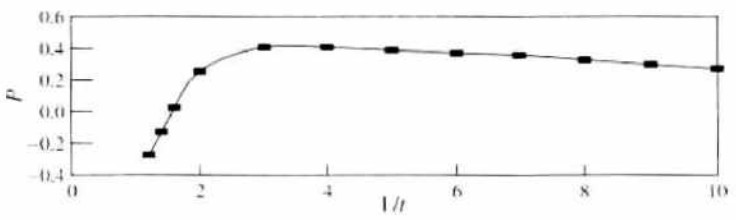


Figure 31. Dependence of the spin polarization of electrons tunneling across a vacuum gap between two Co electrodes on the reciprocal of electron hopping t between the electrodes (width of the gap).

The crossover from negative to positive P occurs because the overlap of d-orbitals decreases with increasing gap much faster than that of s-orbitals and it is, therefore, s electrons that ultimately determine the conductance in the tunneling regime. The same pattern emerges from the tight-binding calculations [88] of the conductance of Co and Fe due to tunneling to an s-band provided only sp bonding at the metal/insulator interface is included. The calculated values of $P \approx 35\%$ and $P \approx 45\%$ for Co and Fe are again in excellent agreement with the observed results [82]. One may, therefore, conclude that direct tunneling between d orbitals of the ferromagnets through Al_2O_3 barrier must be weak. The calculation of tunneling between Fe electrodes through a simple step barrier [87], based on layer Korringa-Kohn-Rostoker method, also gives the correct sign of P, but the calculated values of P are much higher than observed.

6.3. Effect of Disorder on TMR—Toward a Generalized Julliere's Formula

All the theories discussed above assume coherent tunneling, that is, conservation of the momentum $k_\parallel$ parallel to the junction. This is not satisfied for Al_2O_3 barriers which are amorphous. The theories of noncoherent tunneling fall into two categories. Either a simplified treatment of disorder is combined with a realistic band structure or a simple (one-band) model is used but disorder is treated realistically. The simplest treatment of disorder in the barrier is provided by the Julliere's formula, which assumes that tunneling from any occupied state of the left electrode to any unoccupied state in the right electrode is equally probable. This is a reasonable model of disorder but the problem with the Julliere's formula is that it leads to an incorrect sign of the tunneling current. It also fails to describe TMR in junctions with a nonmagnetic metallic interlayer. To find out how the Julliere's formula can be corrected, it is necessary to examine the approximations that are made to derive it from the exact Kubo-Landauer formula (27) of Section 3. There are three approximations involved:

1. It is assumed that in the tunneling regime $\mathbf{t} \approx 0$ (electron hopping between the electrodes is weak), the Keldysh-Landauer formula can be linearized, i.e., $\mathbf{T}_\sigma = \mathbf{t}(\mathbf{I} - \mathbf{g}_R^\sigma \mathbf{t}^\dagger \mathbf{g}_L^\sigma \mathbf{t})^{-1} \to \mathbf{t}$.
2. Only tunneling between the same orbitals is considered and assumed to be equally probable, i.e., the hopping matrix $\mathbf{t}$ is replaced by $t_0\mathbf{I}$, where $\mathbf{I}$ is a unit matrix in the orbital space and t_0 is a single tunneling matrix element independent of $k_\parallel$.
3. Complete loss of coherence across the barrier (vacuum gap) is imposed; that is, it is assumed that a state $k_\parallel$ tunnels with an equal probability to any other state $k_\parallel'$.

With these approximations, the Keldysh-Landauer formula takes the form

$$\Gamma^\sigma = \frac{4e^2}{hN_\parallel}|t_0|^2 \left[\sum_{k_\parallel} \mathrm{Tr}\,\mathrm{Im}\,\mathbf{g}_R^\sigma(E_F, k_\parallel)\right]\left[\sum_{k_\parallel'} \mathrm{Tr}\,\mathrm{Im}\,\mathbf{g}_L^\sigma(E_F, k_\parallel')\right] \tag{72}$$

where $N_\parallel$ is the number of atoms in the plane of the junction. Since the expressions in the brackets are (up to a factor $1/\pi$) the total densities of states of the right and left electrodes, Eq. (72) reduces to the usual expression for the conductance obtained in the classical theory of tunneling [76].

It will be seen in the next section that the linearization (approximation (1)) of the Kubo formula is the reason for the failure of the Julliere's formula to exclude from the tunneling current the spurious contribution of quantum well states that are formed in junctions with a nonmagnetic metallic interlayer.

The approximation (2) is responsible for the incorrect sign of the spin polarization P predicted by Eq. (72) for Fe, Co, and Ni. This is because the formula (72), based on the total DOS, allocates equal weights to tunneling via d states and s-p states. In reality, tunneling via s-p states dominates.

The approximation (3) is useful since it provides the simplest way of dealing with loss of coherence in tunneling (non-conservation of $k_\parallel$).

Based on this analysis, it is clear how the Julliere's formula should be corrected to eliminate the aforementioned problems. First of all, bound (quantum well) states, which do not contribute to transport of charge, must be omitted from the sum over $k_\parallel$ and $k_\parallel'$ in Eq. (72).

The second approximation cannot be made since it leads to an incorrect sign of the tunneling current. It is, therefore, necessary to keep the dependence of the hopping matrix on the orbital indices.

Finally, complete loss of coherence implies that hopping between the electrodes is a constant matrix (independent of the wave vector), which can be approximated by $\mathbf{t}(0)$, where $\mathbf{t}(0)$ is the value of the diagonal hopping matrix element for $k_\parallel = 0$. This is reasonable since the perpendicular tunneling with $k_\parallel = 0$ is expected to dominate.

The generalization of the Julliere's formula (72) incorporating all the corrections discussed above takes the form

$$\Gamma^{\sigma} = \frac{4e^2}{hN_\parallel} \mathrm{Tr}\left[\sum_{k_\parallel}{}' \mathbf{t}(0)\, \mathrm{Im}\, \mathbf{G}_R^{\sigma}(E_F, k_\parallel)\right]\left[\sum_{k_\parallel'}{}' \mathbf{t}^{\dagger}(0)\, \mathrm{Im}\, \mathbf{g}_L^{\sigma}(E_F, k_\parallel')\right] \tag{73}$$

where the prime indicates that all the quantum well states are excluded and the trace is again over all the orbital indices.

The structure of Eq. (73) is very similar to that of the Julliere's formula. However, there are two important differences. All the nonpropagating (quantum well) states are removed from the DOSs of the left and right electrodes and the DOS of each electrode is multiplied by the hopping (tunneling) matrix $\mathbf{t}$. The latter means that the trace over the orbital indices can no longer be factorized as in the original Julliere's formula (72). The generalized Julliere's formula (73) provides the simplest physically plausible description of tunneling in the presence of disorder [89].

Rigorous studies of the effect of disorder on tunneling, based on a single-orbital tight-binding model and the Kubo formula [85], show that, in addition to a mixing of $k_\parallel$ channels, disorder also leads to resonant tunneling via localized electronic states which are formed in the barrier in the presence of impurities or defects [90–92]. Resonant tunneling results [85] in quasi-one-dimensional high-conductance channels, which dominate tunneling when disorder is high and the barrier is thick. It follows that the tunneling current is determined not only by the intrinsic properties of the ferromagnet, such as DOS for a given spin, but also by the type and degree of disorder in the barrier. In spite of that, the numerical results of Ref. [85] indicate that the TMR ratio is approximately given by the Julliere's formula. This suggests that, even in the presence of resonant tunneling, a generalized Julliere's formula (73) should be a reasonable approximation.

6.4. TMR of a Junction with a Nonmagnetic Interlayer

The tunneling junction with a thin nonmagnetic metallic interlayer, such as Cu, Ag, or Au inserted between one of the ferromagnetic electrodes and the insulating barrier, is an interesting system because it allows us to decide unambiguously whether tunneling in a particular junction is coherent or incoherent. In the case of incoherent tunneling, the conventional Julliere's formula, which applies to a totally disordered junction, predicts that there should be no TMR since the density of states of the nonmagnetic interlayer adjacent to the barrier is spin independent. However, early experiments [93] indicated a small nonzero TMR for a very thin interlayer but the observed TMR decayed rapidly to zero with increasing interlayer thickness. A nonzero TMR for coherent tunneling was predicted theoretically by Vedyaev et al. [94] and Zhang and Levy [95] using a simple parabolic band model. Mathon and Umerski [89] applied the Kubo-Landauer formula (27) of Section 3 to calculate the dependence of the TMR ratio on the thickness of the Cu interlayer in a Co junction with vacuum gap. Their results are reproduced in Fig. 32. It can be seen that TMR is nonzero and the optimistic TMR ratio oscillates as a function of the Cu interlayer about an average value of about 15%. For comparison, the results obtained from the generalized Julliere's formula (73) are also shown in Fig. 32 (broken line). In order to understand why a nonzero TMR is obtained for coherent tunneling in a Co/Cu/vacuum/Co junction, it is instructive to examine the band structures of Co and Cu shown in Fig. 9 of Section 2. As already discussed in Section 2, there is a very good match between the majority-spin bands of Co and those of Cu. It follows that majority-spin carriers cross easily the Co/Cu interface and participate

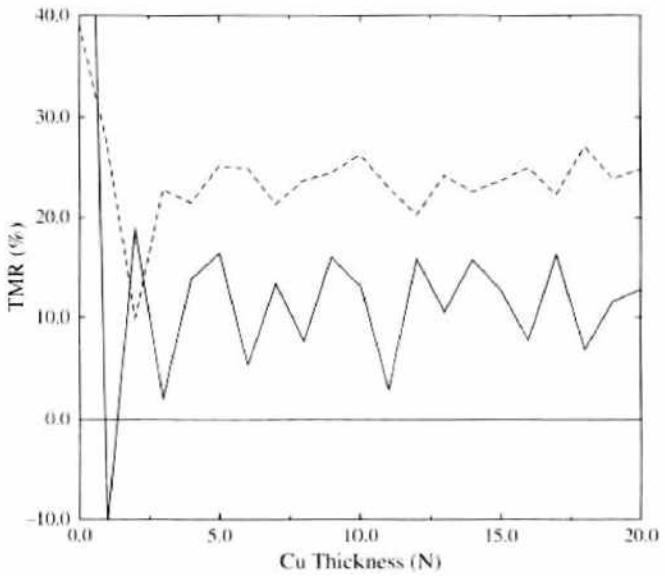


Figure 32. Dependence of the TMR ratio of a Co/Cu_N/vacuum/Co junction on the thickness of the Cu interlayer. Results denoted by a solid line were obtained from the Kubo-Landauer formula (3.15); broken line denote the results obtained from Eq. (73).

in tunneling as if there were no intervening Cu layer, that is, as far as majority-spin carriers are concerned, the Co electrode extends effectively right up to vacuum gap. On the other hand, there is a poor match between the Cu bands and the minority-spin bands in Co, which results in formation of down-spin quantum well states in the Cu interlayer [48, 96]. These are bound states fully localized in the Cu layer and, therefore, they do not contribute to transport of charge. Since quantum well states occur only in the down-spin channel and the up-spin channel is unaffected by the Cu interlayer, it is clear that there is a spin asymmetry in transport of charge (i.e., nonzero TMR). It is important to realize that, although there is a spin asymmetry in transport through the Cu interlayer, the Cu layer itself remains paramagnetic. This is because quantum well states contribute to the ordinary density of states, which means that the total numbers of up- and down-spin carriers in the Cu layer remain equal.

The origin of large-amplitude oscillations of TMR is the same as for CPP GMR that was already discussed in Section 5. The Cu interlayer represents an effective potential well for minority-spin carriers and those carriers that do not get trapped in the well undergo interference above the well which results in oscillations of the conductance. As in the case of CPP GMR, we expect two types of oscillation period. One originates from extremal points of the Cu Fermi surface and the other from cut off points of the conductance at the edge of the quantum well. All such oscillations combine to produce a rather complicated oscillatory behavior seen in Fig. 32.

The dependence of the TMR on the thickness of a Cu interlayer was studied experimentally by Yuasa et al. [97] for a Co/Cu/Al_2O_3Py junction. Their results, which are reproduced in Fig. 33, reveal well-defined oscillations of the TMR ratio about zero average with a single oscillation period corresponding to the spanning vector at the belly extremum of the Cu Fermi surface. The observation of quantum oscillations of TMR is very important since it demonstrates for the first time coherence in transport in a magnetic tunneling junction. However, there are two important aspects in which the experimental results differ from the theoretical predictions of Mathon and Umerski [89]. Firstly, only a single oscillation period is observed while there are several periods predicted for tunneling across a vacuum gap [89]. The second problem is that the observed TMR oscillates about zero average rather than about a finite background predicted theoretically.

Why only a single oscillation period is observed is simple to understand. In fact, there is a qualitative difference between tunneling across a vacuum gap and tunneling through an insulating barrier. In the tight-binding model of tunneling across a vacuum gap, turning off the tight-binding hopping integrals across the gap affects equally all the $k_\parallel$ channels. It follows

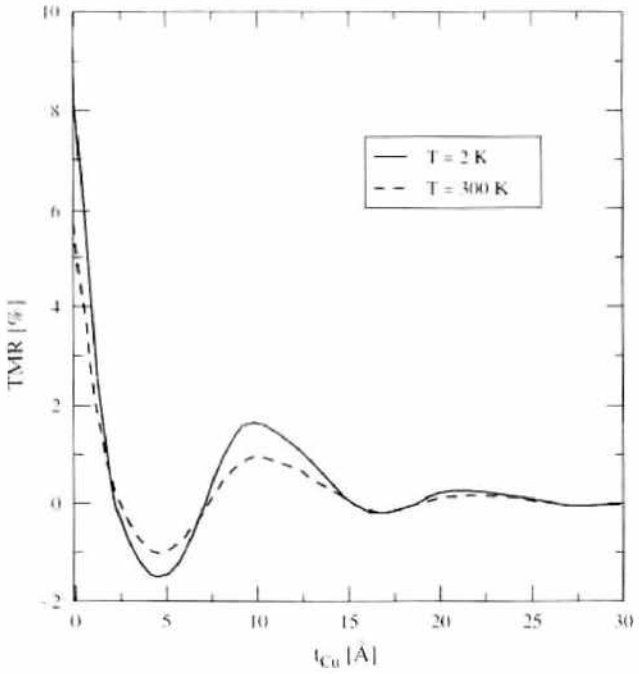


Figure 33. TMR ratio of a Co/Cu/Al_2O_3/Py junction as a function of Cu thickness.

that the contributions to oscillations of the conductance from all the extrema of the Cu Fermi surface and from all the cut off points have approximately the same weight and multiple oscillation periods thus occur. However, in the case of an insulating barrier, carriers with $k_{\parallel} \approx 0$ have a much higher tunneling probability since the corresponding perpendicular energy has the largest value. This explains why all the other oscillation periods that occur at $k_{\parallel} \neq 0$ are strongly suppressed and tunneling with $k_{\parallel} \approx 0$ (Cu Fermi surface belly) dominates [98].

The reason why the observed TMR oscillates about zero average is more subtle. The calculated results for tunneling across a vacuum gap [89] apply to a fully ordered junction with coherent tunneling. On the other hand, the junction studied experimentally [97] is only partially ordered since the Al_2O_3 barrier is amorphous. The only part of the junction with phase coherence is the Co/Cu bilayer. Since the barrier (whether amorphous or not) represents a high potential step and the Co/Cu interface is another potential step for minority-spin carriers, quantum interference of such carriers in the resulting quantum well must occur. This explains why oscillations of TMR are expected even for an amorphous barrier. However, to understand why oscillations of TMR are about zero average, one has to consider explicitly the effect of disorder in the barrier. This was done by Itoh et al. [98] using a single-orbital tight-binding model and coherent potential approximation to treat disorder in the barrier. They found that indeed the disorder results in oscillations of TMR about zero average. Their results can be understood using a simple physical argument based on the relative sizes of the majority- and minority-spin Fermi surfaces. The CPA calculation of Itoh et al. [98] shows that the average conductances of majority- and minority-spin carriers in the ferromagnetic configuration of the junction are virtually unaffected by disorder in the barrier. Similarly, the average tunneling conductance of minority-spin carriers in the antiferromagnetic configuration is also only very weakly influenced by the disorder. However, it is found that the average conductance of the majority-spin carriers in the antiferromagnetic configuration increases with disorder, and that results in zero average TMR. This can be understood as follows. In an ordered junction, conservation of $k_{\parallel}$ means that only carriers tunneling from the portion of the larger majority Fermi surface which coincides with the smaller minority Fermi surface can tunnel. However, disorder in the barrier breaks the conservation of $k_{\parallel}$, which has the consequence that carriers from the whole cross section of the larger majority Fermi surface can tunnel into the smaller minority Fermi surface of the right electrode. Hence the corresponding conductance increases.

Although growing a magnetic tunneling junction that is at least partially ordered is a significant progress, full understanding of spin-dependent tunneling is possible only for a

fully ordered junction with a crystalline barrier. More significantly, we can not only model fully ordered junction realistically but we can even design junctions with better performance through the appropriate choice of the electrode and barrier materials. This is the subject of the next section.

6.5. Fully Realistic Modeling of Epitaxial Magnetic Tunneling Junctions

Although planar junctions with a vacuum gap can be modeled realistically they cannot be realized experimentally. On the other hand, little theoretical progress has been made for junctions with an insulating barrier since systems for which tunneling could be regarded as coherent were lacking. The situation has changed radically with the recent demonstration [99, 100] of tunneling in an epitaxial Fe/MgO/Fe(001) junction. To the first approximation (neglecting defects), tunneling should be coherent and Fe/MgO/Fe(001) is, therefore, an ideal system to be studied theoretically. Soon after the experimental demonstration of an epitaxial growth of Fe/MgO/Fe(001), two calculations were made of the coherent TMR in this system [101, 102]. The calculation of the Butler group [101] is based on direct matching of the wave functions across the whole junction whereas the calculation of Mathon and Umerski [102] employs the Kubo-Landauer formula. Both calculations give very similar results but since the calculation of Mathon and Umerski conforms to the Kubo-Landauer Green function formalism adopted in this review we shall describe their approach.

It is known experimentally [103] that thin epitaxial bcc Fe(001) films grow pseudomorphically on rocksalt MgO(001) substrate so that the Fe atoms sit above the O ions. The Fe lattice is, therefore, rotated by 45° relative to the MgO lattice. LEED studies [104] show that the Fe-O distance is almost exactly equal to the distance between the neighboring MgO atomic planes. This picture is confirmed by first-principle calculations of Li and Freeman [105]. There is only a small lattice mismatch of about 3.5% between the Fe-Fe and O-O in-plane distances. Li and Freeman [105] further show that the electron population at the MgO interface plane is virtually the same as for the clean MgO surface and the Fe interface plane also behaves like a free Fe surface.

On the basis of these results, one can neglect the small lattice mismatch between Fe and MgO and assume that the whole Fe/MgO/Fe(001) junction grows epitaxially. The band structure of the electrodes was described in [102] by tight-binding bands fitted to the *ab initio* band structure of bcc Fe [64] and that of the barrier by tight-binding bands fitted to the band structure of bulk MgO [106]. The on-site potentials in the Fe interface plane were adjusted self-consistently to reproduce the correct surface moment of Fe [105]. No adjustments of the surface potentials of MgO were found to be necessary. Hoppings up to third nearest neighbors were used. The band gap for the band structure of bulk MgO used in [102] is 7.6 eV, which is in a good agreement with the height of the tunneling barrier of 3.6 eV obtained by Wulfhekel et al. [99]. The fact that the observed tunneling barrier is about a half of the band gap suggests that the Fermi level of the junction lies close to the middle of the gap. This is in very good agreement with the calculated results of Li and Freeman [105] who place the Fermi level 3.5 eV above the top of the valence band of MgO. One can, therefore, use this value to align the tight-binding bands of Fe and MgO. Finally, the tight-binding hopping integrals between Fe and MgO were determined by Harrison's method [17].

Once the band structure of the Fe/MgO/Fe(001) junction is determined, evaluation of the conductance proceeds using the general Kubo-Landauer formula Eq. (27) of Section 3. One only requires the surface one-electron Green functions at the two neighboring atomic planes, labelled L and R, separated by a cleavage plane. Equation (27) is the computationally most efficient way of calculating the conductance since only the diagonal (in the plane index i) elements of the left and right surface Green functions of the cut junction need to be evaluated. The surface Green functions g_L^σ, g_R^σ are determined from the surface Green's function g_s of a semi-infinite Fe electrode using the Dyson equation. g_s itself is calculated by the generalized Möbius transformation method described in Section 3, which allows us to determine g_s quite accurately for an imaginary part of the energy ϵ as small as 10^{-12} Ry.

The numerical evaluation of Eq. (27) for a tunneling junction is not straightforward. The exact Green's function in the MgO barrier should decay exponentially. However, a small

imaginary part ϵ of the energy results in a propagating component, which leads to a spurious "metallic-like" ballistic conductance independent of the barrier thickness. For a thick barrier, this spurious contribution eventually becomes dominant for any nonzero ϵ. This problem can be eliminated by reformulating Eq. (27) in terms of the off-diagonal elements of the Green's function connecting the surfaces of the left and right electrodes [27]. However, it is computationally more demanding to calculate the off-diagonal components of the Green's function than the diagonal ones in Eq. (27).

The second problem is that the partial conductance $\Gamma^{\sigma}(\boldsymbol{k}_{\parallel})$ exhibits very sharp peaks in certain regions of the two-dimensional Brillouin zone (2D BZ). Since such peaks make a significant contribution to the total conductance, an extremely fine mesh of $\boldsymbol{k}_{\parallel}$ points is required. It turns out that up to $\approx 10^6$ $\boldsymbol{k}_{\parallel}$ points in the irreducible segment of the 2D BZ are needed to achieve convergence. Moreover, with a fine mesh of $\boldsymbol{k}_{\parallel}$ points one also needs a very small ϵ for numerical stability.

Given the requirement of a very fine mesh of $\boldsymbol{k}_{\parallel}$ points, it is best to use the computationally most efficient Eq. (27). To minimize the effect of a finite ϵ (10^{-12} Ry), the conductances were calculated by cutting the junction in the middle of the MgO barrier. The most stringent test that the error due to a finite ϵ is negligible is to check that all the conductances decrease exponentially in the limit of a thick barrier. This was satisfied for MgO barriers as thick as 20 atomic planes.

The dependence of the pessimistic TMR ratio on the thickness of the MgO barrier obtained in Ref. [102] is reproduced in Fig. 34(a). The majority-spin $\Gamma^{\uparrow}_{\rm FM}$ and minority-spin $\Gamma^{\downarrow}_{\rm FM}$ conductances in the ferromagnetic configuration of the junction and the conductance $\Gamma_{\rm AF}$ of electrons of either spin in the antiferromagnetic configuration are plotted against the MgO thickness on a logarithmic scale in Fig. 34(b). The TMR ratio oscillates initially with MgO thickness but after about seven atomic planes of MgO stabilizes and increases only slowly reaching a very high value of 0.92 for 20 atomic planes of MgO. This corresponds to the optimistic ratio of some 1200%. The behavior of the individual conductances is more informative. Firstly, it is clear from Fig. 34(b) that the majority-spin conductance is always higher than the minority-spin conductance. It follows that the calculated spin polarization of the tunneling current is positive, as found experimentally for junctions based on Al_2O_3 barrier. It is also clear that after some 10 atomic planes of MgO the junction reaches an asymptotic regime with all the conductances decreasing exponentially with MgO thickness. However, the slope of $\Gamma^{\uparrow}_{\rm FM}$ is somewhat smaller than that of $\Gamma^{\downarrow}_{\rm FM}$ and $\Gamma_{\rm AF}$. This indicates that, even for 20 atomic planes of MgO, the decay of the conductances in these three channels is not controlled by the same exponential factor.

To clarify the rather unusual behavior of the Fe/MgO/Fe junction, we show in Figs. 35 and 36 the $\boldsymbol{k}_{\parallel}$-dependence of the partial conductances $\Gamma^{\sigma}_{\rm FM}(\boldsymbol{k}_{\parallel})$ and $\Gamma_{\rm AF}(\boldsymbol{k}_{\parallel})$ in the 2D BZ. The results shown in Fig. 35 are for four atomic planes of MgO and those in Fig. 36 for

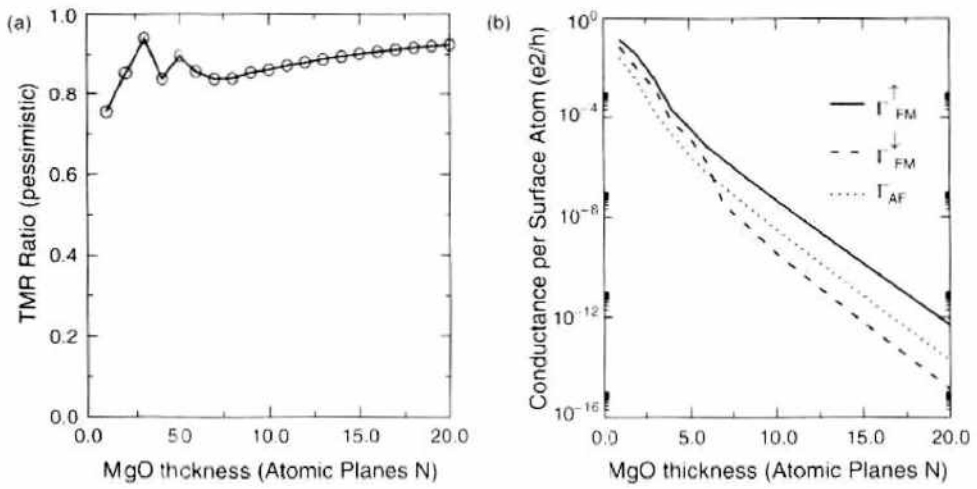


Figure 34. (a) Dependence of the pessimistic TMR ratio $R_{\rm TMR}$ of an Fe/MgO/Fe(001) junction on MgO thickness. (b) Dependencies of the total conductances $\Gamma^{\uparrow}_{\rm FM}$, $\Gamma^{\downarrow}_{\rm FM}$, and $\Gamma_{\rm AF}$ on MgO thickness.

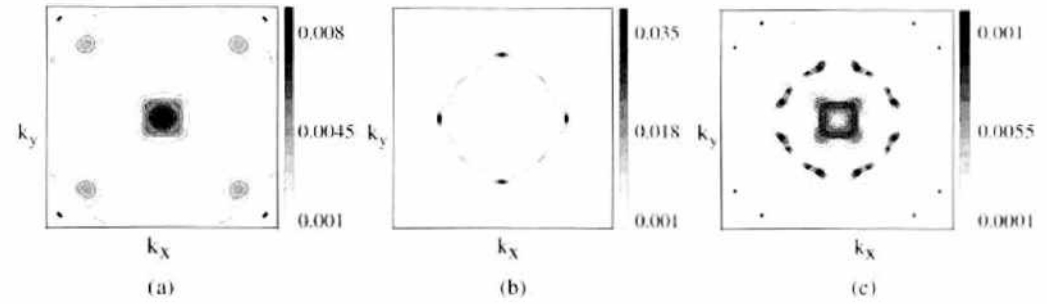


Figure 35. Distribution of the partial conductances in the two-dimensional Brillouin zone for an Fe/MgO/Fe(001) junction with four atomic planes of MgO: (a) $\Gamma^{\uparrow}_{FM}(\boldsymbol{k}_{\parallel})$; (b) $\Gamma^{\downarrow}_{FM}(\boldsymbol{k}_{\parallel})$; (c) $\Gamma_{AF}(\boldsymbol{k}_{\parallel})$.

10 planes. These thicknesses were chosen because they correspond to the transition from a preasymptotic regime to the asymptotic regime.

The conductance $\Gamma^{\uparrow}_{FM}$ shown in Fig. 35(a) has the expected maximum at $\boldsymbol{k}_{\parallel} = 0$ but also four subsidiary maxima along the $k_x = k_y$ lines in the 2D BZ. On the other hand, $\Gamma^{\downarrow}_{FM}$ [Fig. 35(b)] is virtually zero at the Γ point and most conduction goes through a 'ring' well removed from $\boldsymbol{k}_{\parallel} = 0$. Finally, Γ_{AF} [Fig. 35(c)] has maxima along two concentric 'rings' in the 2D BZ but a minimum at $\boldsymbol{k}_{\parallel} = 0$. For a thicker MgO barrier (10 atomic planes), the junction moves closer to the expected asymptotic regime. The conductance $\Gamma^{\uparrow}_{FM}$ [Fig. 36(a)] is now dominated by the Γ point. Similarly, Γ_{AF} [Fig. 35(c)] and $\Gamma^{\downarrow}_{FM}$ [Fig. 36(b)] are determined by the inner "ring," which is also very close to the Γ point. However, both the conductances Γ_{AF} and $\Gamma^{\downarrow}_{FM}$ have a minimum at the Γ point.

The calculated dependence of the TMR on the thickness of MgO can be understood qualitatively in terms of the surface spectral densities $(-1/\pi)\mathrm{Im}\,\mathbf{G}^{\sigma}_{s}(E_F, \boldsymbol{k}_{\parallel})$ of Fe(001) and the complex Fermi surface (FS) of MgO. They are reproduced in Fig. 37. We first show in Fig. 37(a) the smallest decay constant $\mathrm{Im}\,k_{\perp}(\boldsymbol{k}_{\parallel})$ for electrons in the MgO barrier (the lowest sheet of the complex MgO FS). Perpendicular tunneling $\boldsymbol{k}_{\parallel} = 0$ is clearly favored, but there are four subsidiary minima of $\mathrm{Im}\,k_{\perp}(\boldsymbol{k}_{\parallel})$ along the $k_x = k_y$ lines. These are responsible for the four subsidiary maxima of $\Gamma^{\uparrow}_{FM}$ seen in Fig. 35(a). The other factor contributing to the maxima is that the surface spectral density of the majority-spin electrons [Fig. 37(b)] is distributed over the whole two-dimensional BZ. On the other hand, the spectral density of the minority-spin electrons [Fig. 37(c)] is concentrated along the large ring seen already in Fig. 35(b) and there is hardly any density at the Γ point. This factor alone explains the behavior of $\Gamma^{\downarrow}_{FM}$ seen in Fig. 35. The behavior of Γ_{AF} is determined by a superposition of the pictures for the $\uparrow$- and $\downarrow$-spin spectral densities. As the thickness of MgO increases, the contributions from the parts of the two-dimensional BZ further away from the Γ point are weakened and that explains the transition of the junction from the preasymptotic to the asymptotic regime.

It remains to clarify why the MgO junction does not reach the expected asymptotic regime in which all the conductances decay with the same $\mathrm{Im}\,k_{\perp}(\boldsymbol{k}_{\parallel} = 0)$ (perpendicular tunneling). Figure 36 shows that there is virtually no tunneling at the Γ point in the minority-spin channel and that explains why $\Gamma_{AF}(\boldsymbol{k}_{\parallel})$ and $\Gamma^{\downarrow}_{FM}(\boldsymbol{k}_{\parallel})$ decay faster with

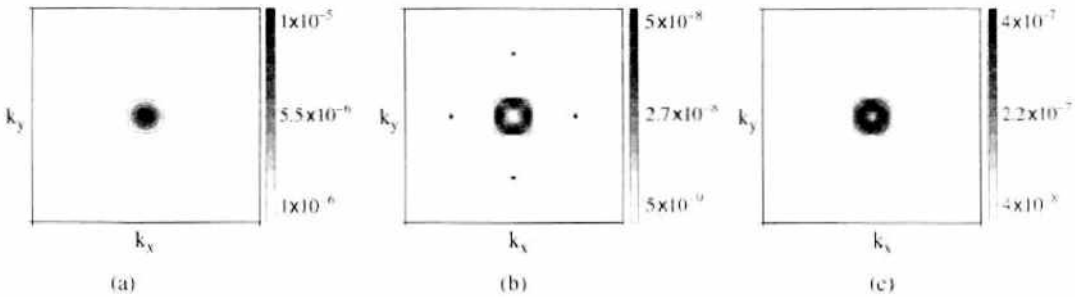


Figure 36. Distribution of the partial conductances in the two-dimensional Brillouin zone for an Fe/MgO/Fe(001) junction with eight atomic planes of MgO: (a) $\Gamma^{\uparrow}_{FM}(\boldsymbol{k}_{\parallel})$; (b) $\Gamma^{\downarrow}_{FM}(\boldsymbol{k}_{\parallel})$; (c) $\Gamma_{AF}(\boldsymbol{k}_{\parallel})$.

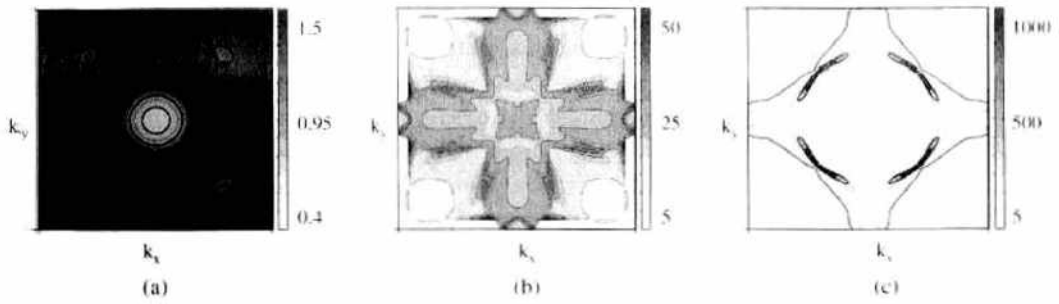


Figure 37. (a) The smallest decay constant $\mathrm{Im}\,k_z(k_\parallel)$ of electrons in the MgO barrier (the lowest sheet of the complex MgO Fermi surface). (b) The majority-spin surface spectral density of Fe(001). (c) The minority-spin surface spectral density of Fe(001).

MgO thickness than $\Gamma^{\uparrow}_{\mathrm{FM}}$. Neither the minority-spin spectral density nor the MgO complex FS can explain the presence of a 'hole' of the conductances $\Gamma_{\mathrm{AF}}(k_\parallel)$ and $\Gamma^{\uparrow}_{\mathrm{FM}}(k_\parallel)$ at the Γ point.

The reason the conductance in the minority-spin channel is so low at the Γ point can be best understood using the approach of Butler et al. [101] which is based on direct matching of electron wave functions. Their results for the decay at the Γ point of the local density of states of majority- and minority-spin electrons in the ferromagnetic configuration of the junction are reproduced in Fig. 38(a) (majority-spin electrons) and Fig. 38(b) (minority-spin electrons). It can be seen that there is a slowly decaying state of Δ_{spd} symmetry present in the majority-spin band but this state is absent in the minority-spin band. It follows that the conductance of majority-spin carriers decays much more slowly with MgO thickness than that of minority spin carriers and also than that of carriers of either spin orientation in the antiferromagnetic configuration.

Early attempts to verify experimentally the very high TMR ratio predicted theoretically were plagued by oxidation of the iron electrodes [107]. However, two groups [108, 109] succeeded recently in overcoming this problem and the measured optimistic TMR ratios $\approx 250\%$ are beginning to approach the theoretical values of $\approx 1000\%$. We reproduce in Fig. 39 the results of Yuasa et al. [108], which show quite clearly that the observed TMR ratio for Fe/MgO/Fe(001) increases with increasing thickness of MgO, as predicted theoretically [101, 102]. Small oscillations of the TMR ratio seen in Fig. 39 do not appear in the calculated TMR and their precise origin still needs to be clarified.

The observation of very high TMR ratios for Fe/MgO/Fe(001) system is very significant since it is the first convincing demonstration of coherrent tunneling in a junction with metallic electrodes. It marks the beginning of a new era in spintronics. Not only junctions with such high TMR ratios are required for a new generation of reading heads and in the development of MRAM, but they are also fundamentally very important since a meaningful comparison between experiment and fully predictive theories of the type described above is now possible.

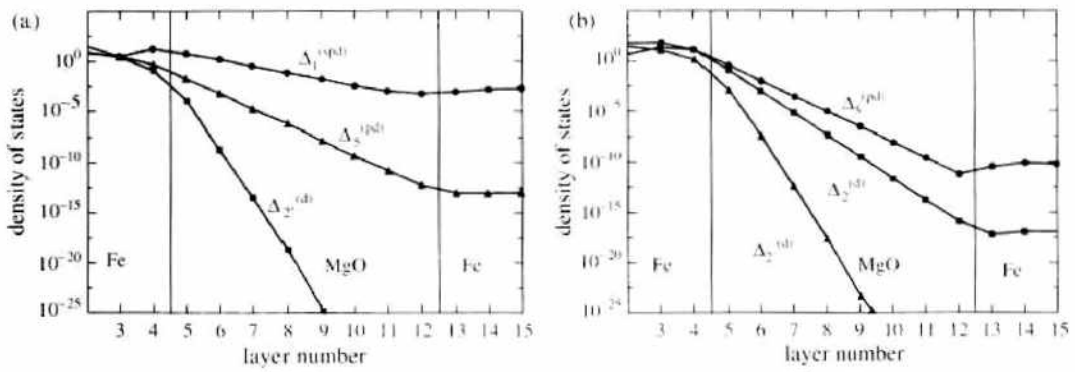


Figure 38. Dependence of the local density of states of majority-spin electrons (a) and minority-spin electrons (b) in the ferromagnetic configuration on the atomic layer number in the Fe/MgO/Fe(001) junction.

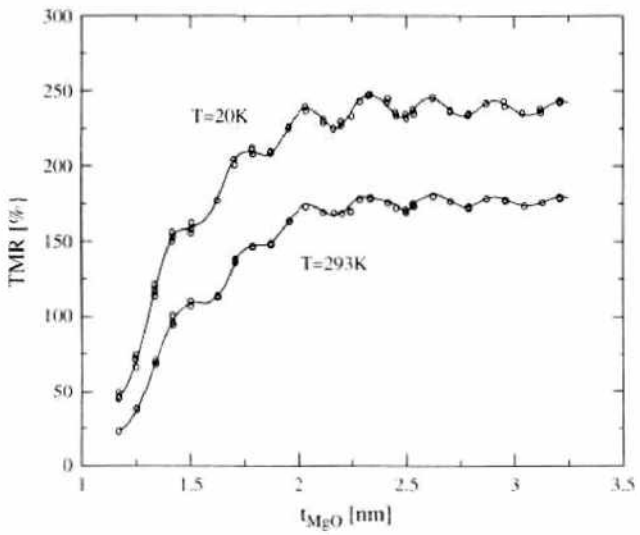


Figure 39. Dependence of the TMR ratio of Fe/MgO/Fe(001) junction on MgO thickness (after Yuasa et al. [108]).

7. CURRENT-INDUCED SWITCHING OF MAGNETIZATION

As discussed in Section 2, it is now well established that by altering the magnetic configuration of a magnetic multilayer we can influence the charge current flowing in it. It was pointed out by Slonczewski [110] that, conversely, by passing a strong charge current one can alter the magnetic state of a magnetic multilayer. Early related theoretical work is due to Berger [111]. The layer structure in which this effect is expected to occur is shown schematically in Fig. 40, where **p** and **m** are unit vectors in the direction of the magnetization. The structure consists of a thick (semi-infinite) left magnetic layer (polarizing magnet), a nonmagnetic metallic spacer layer, a thin second magnet (switching magnet) and a semi-infinite lead. It is assumed that the magnetization of the polarizing magnet is pinned (for example by a strong anisotropy field) in a particular direction θ. Slonczewski argued that the left magnet will then spin polarize the current passing through it and the resultant spin current flowing through the spacer can be absorbed by the switching magnet. The rate of change of the total spin, given by the difference between the spin current entering the switching magnet and that leaving the magnet, is equal to the torque exerted on the magnetic moment of the switching magnet. If the charge current, and the associated spin current, is strong enough the spin-transfer torque can cause total reversal of the switching magnet moment. This effect is called current-induced switching of magnetization.

On the most elementary level, one can simply assume that the polarizing magnet produces a spin current that gets partially or fully absorbed by the switching magnet and explore the consequences of the resultant torque acting on the magnetization of the switching magnet. This can be done using a phenomenological Landau-Lifshitz (LL) equation with an appropriate spin-transfer torque term. To treat correctly the dynamics of current-induced switching of magnetization, it is also necessary to include in the LL equation the usual Gilbert damping term. We shall refer to the LL equation with the Gilbert damping term as LLG equation and

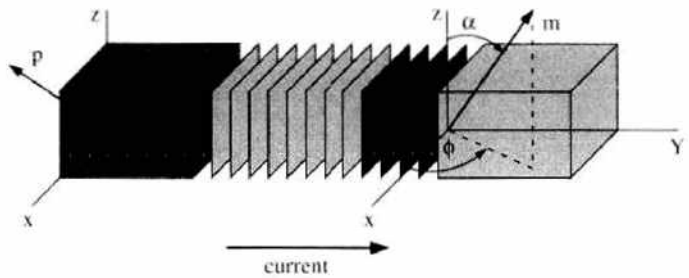


Figure 40. Schematic picture of a magnetic layer structure for current-induced switching (magnetic layers are darker, non-magnetic layers lighter).

begin our review of the current-induced switching of magnetization with this phenomenological treatment. We shall also use this phenomenological description to discuss experiments on current-induced switching of magnetization.

However, the phenomenological approach leaves many questions unanswered. In particular, to understand and optimize the switching effect, we need to know the magnitude and direction of the spin-transfer torque for any specific combination of nonmagnetic and magnetic layer materials. We also require the dependences of the spin-transfer torque on the thicknesses of both the nonmagnetic and magnetic layers. Finally, to describe correctly the switching effect, we also need to know the detailed dependence of the spin-transfer torque on the angle between the magnetizations of the polarizing and switching magnets. To answer all these questions we need to calculate microscopically the spin current entering and leaving the switching magnet, that is, the torque acting on it. The most direct microscopic approach is the original calculation of Slonczewski [110] for a simple parabolic band model of a magnetic multilayer. He calculated the spin current (torque) from the one-electron wave functions assuming that the magnetizations of the polarizing and switching magnets are kept at a given fixed angle. This type of calculation corresponds to a scattering experiment. An incoming electron with a given spin orientation (determined by the polarizing magnet) is scattered off an exchange field of the switching magnet which is not parallel to the spin orientation of the incident electron. Calculations based on this idea are designed to tell us how much of the spin angular momentum of the incident electron is absorbed by the switching magnet. The reaction of the switching magnet to the absorbed spin angular momentum is ignored at this stage and is determined separately in a second independent calculation using the phenomenological LLG equation. In the way originally described by Slonczewski the method was applicable without any approximations only to ferromagnets with a very large exchange splitting of up and down-spin bands. However, we shall show that this restriction can be relaxed using the transfer matrix method to match the electron wave functions across all the interfaces.

The principal limitation of this method is that it is not easily generalizable to a realistic band structure. On a more fundamental level, evaluating the spin current directly from one-electron wave functions requires some justification since one also has to link the current obtained from the spin current operator to an applied bias. In the analogous problem of charge transport this was done by Landauer [112]. He showed quite rigorously that the conductance of a system sandwiched between two reservoirs is given by its total quantum mechanical transmission coefficient. It is quite straightforward to generalize Landauer's method to transmission of spin [113, 114]. However, care has to be taken when applying such a Landauer-like formula to a magnetic layer system. The Landauer formula for ordinary charge current gives us the conductance of the whole system. This is all that is needed since the charge current is conserved and, therefore, can be calculated anywhere in the structure. This is the reason why in the Landauer method one evaluates only the total transmission coefficient for the whole structure. However, in the case of a magnetic layer structure, the spin current is not conserved since it may get partially or even fully absorbed by the magnetic layers. It follows that to calculate the spin-transfer torques acting on various parts of a layer structure, we require the local spin currents entering and leaving each magnetic layer. Rigorous justification of a generalized Landauer formula to such a situation is still lacking. However, we shall show that the generalized Landauer formula gives results for local spin currents that are in complete agreement with the results obtained from the rigorous Keldysh formalism (Section 3).

Finally, we shall describe the application of the Keldysh formalism to current-induced switching of magnetization. In this approach, one no longer regards the calculation of the spin current as a scattering problem. One adopts instead the point of view that, for any given applied bias (charge current), the magnetization of the switching magnet will reach a steady state in which the spin-transfer torque acting on the switching magnet is exactly compensated by the torques due to anisotropy and applied magnetic fields. Hence we set ourselves the task of calculating the local steady-state spin current. The method for calculating such a steady-state current was developed by Keldysh [26, 27]. It is assumed that, initially, the multilayer is separated by a cleavage plane into two independent left-hand and right-hand parts

so that charge carriers cannot move between them. When a bias is applied to such a system no current flows and the system remains in equilibrium but with unequal chemical potentials μ_L and μ_R for the left- and right-hand parts. Next, carrier hopping across the cleavage plane is adiabatically turned on and the system evolves toward a steady state in which both the charge and spin currents flow. The Keldysh method [26] described in Section 3 provides a rigorous prescription for calculating all the steady-state properties of the connected nonequilibrium system from the known properties of the equilibrium cleaved system. In particular, to calculate the spin-transfer torque, it is only necessary to pass the cleavage plane immediately before and after the switching magnet and the difference between the incoming and outgoing spin currents obtained by the Keldysh method is the required torque. It is straightforward to implement the method for a fully realistic band structure [115].

While the Keldysh method gives us all the possible steady states and the corresponding spin-transfer torques of a magnetic multilayer under an applied bias, there remains a question of the dynamical stability of the steady state. If one is merely using the Keldysh method to determine the spin-transfer torque for any particular orientation of the switching magnet magnetization, the stability of the corresponding steady state is immaterial. However, more generally, we can regard the current-induced switching of magnetization as a loss of stability of the steady state. One can argue that, when an applied bias (charge current) reaches a critical value, a steady state that the system has reached via a sequence of other steady states becomes unstable and the system seeks out a new steady state in which the magnetization is switched to the opposite direction [115]. The stability of steady states cannot be decided within the microscopic Keldysh formalism but this poses no problem since the stability of any particular state can be decided using the phenomenological LLG equation. Provided the microscopically determined spin-transfer torques are used as an input into the LLG equation, this approach should yield an essentially first-principle modelling of current-induced switching of magnetization.

7.1. Phenomenological Treatment of Current-Induced Switching of Magnetization

In this section, we explore the consequences of the spin-transfer torque acting on a switching magnet using a phenomenological Landau-Lifshitz equation with Gilbert damping (LLG equation). This is essentially a generalization of the approach used originally by Slonczewski [110] and Sun [116]. We assume that there is a polarizing magnet whose magnetization is pinned in the (x, z)-plane in the direction of a unit vector $\mathbf{p}$ which is at a general fixed angle θ to the z-axis, as shown in Fig. 40. The pinning of the magnetization of the polarizing magnet can be due to its large coercitivity (thick magnet) or a strong uniaxial anisotropy. The role of the polarizing magnet is to produce a stream of spin-polarized electrons, i.e., spin current that is going to exert a torque on the magnetization of the switching magnet whose magnetization lies in the general direction of a unit vector $\mathbf{m}$. The orientation of the vector $\mathbf{m}$ is defined by the polar angles α, ϕ shown in Fig. 40. There is a thin nonmagnetic metallic layer inserted between the two magnets whose role is merely to separate magnetically the two magnetic layers and allow a strong charge current to pass. The total thickness of the whole trilayer sandwiched between two nonmagnetic leads must be smaller than the spin diffusion length l_{sf} so that there are no spin flips due to impurities or spin-orbit coupling. A typical junction in which current-induced switching is studied experimentally [117] is shown schematically in Fig. 41. The thickness of the polarizing magnet is 40 nm, that of the switching magnet 2.5 nm and the nonmagnetic spacer is 6 nm thick. The materials most commonly used for the two magnets and the spacer are cobalt and copper, respectively. The junction cross section is approximately oval-shaped and its diameter is only 130 nm. A small diameter is necessary so that the torque due to the Oersted field generated by a charge current of 10^{7-8} A/cm^2, required for current-induced switching, is much smaller than the spin-transfer torque we are interested in.

It is useful at this stage to review briefly the experimental set-up which is used to study the current-induced switching so that we can define the remit of the phenomenological treatment of the effect. The aim of most experiments is to determine the orientation of the switching

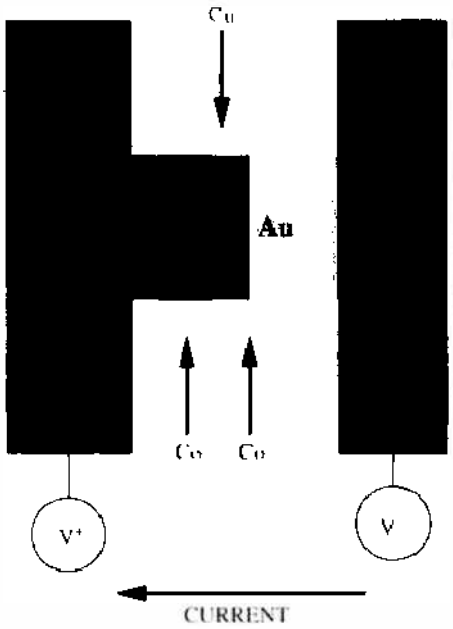


Figure 41. Schematic picture of a junction in which current-induced switching is studied experimentally.

magnet moment as a function of the current (applied bias) in the junction. Discontinuities in such a dependence indicate sudden jumps of the magnetization direction (i.e., current-induced switching). The orientation of the switching magnet moment **m** relative to that of the polarizing magnet **p**, which is fixed, is determined by measuring the resistance of the junction. Because of the GMR effect, the resistance of the junction is higher when the magnetizations of the two magnets are antiparallel than when they are parallel. In other words, what is observed are hysteresis loops of resistance versus current. A typical experimental hystersis loop of this type [118] is reproduced in Fig. 42. It can be seen from Fig. 42 that, for any given current, the switching magnet moment is stationary (the junction resistance has a well defined value); i.e., the system is in a steady state. This holds everywhere on the hystersis loop except for the two discontinuities where current-induced switching occurs. It follows that to interpret experiments that exhibit such hysteretic behavior [117], the first task of the theory is to determine from the LLG equation all the possible steady states and then investigate their dynamical stability. At the point of instability, the system seeks out a new steady state, that is, discontinuous transition to a new steady state with the switched magnetization occurs. We have tacitly assumed that there is always a stable steady state available for the system to jump to. There is now experimental evidence that this is not always the case. In the absence

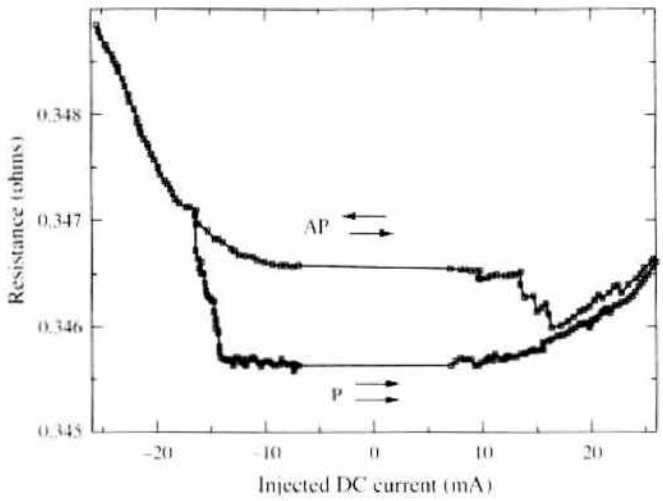


Figure 42. Resistance versus current hysteresis loop (after Grollier et al. [118]).

of any stable steady state the switching magnet moment must remain permanently in a time-dependent state. This interesting case is implicit in the phenomenological LLG treatment and we shall discuss it in detail later.

In describing the switching magnet by a unique unit vector **m**, we assume that it remains uniformly magnetized during the switching process. This is only strictly true when the exchange stiffness of the switching magnet is infinitely large. It is generally a good approximation as long as the switching magnet can be regarded as small enough to remain single domain so that switching occurs purely by rotation of the magnetization as in the Stoner-Wohlfarth theory [119] of field switching. This seems to be the case in many experiments [117].

Before we can apply the LLG equation to study the time evolution of the unit vector **m** in the direction of the magnetization of the switching magnet, we need to determine all the contributions to the torque acting on the switching magnet. First, there is a torque due to the uniaxial and easy plane (shape) anisotropies. The shape anisotropy torque arises because the switching magnet is a thin layer typically only a few nanometers thick. The uniaxial anisotropy is believed to be due to a rather irregular cross section of the switching magnet [117]. We take the uniaxial anisotropy axis of the switching magnet to be parallel to the z axis of the coordinate system shown in Fig. 40. Since the switching magnet lies in the (x, z)-plane, we can write the total anisotropy field as

$$\boldsymbol{H}_A = \boldsymbol{H}_u + \boldsymbol{H}_p \tag{74}$$

where $\boldsymbol{H}_u$ and $\boldsymbol{H}_p$ are given by

$$\boldsymbol{H}_u = H_{u0}(\boldsymbol{m} \cdot \boldsymbol{e}_z)\boldsymbol{e}_z \tag{75}$$

$$\boldsymbol{H}_p = -H_{p0}(\boldsymbol{m} \cdot \boldsymbol{e}_y)\boldsymbol{e}_y \tag{76}$$

Here, $\boldsymbol{e}_x$, $\boldsymbol{e}_y$, $\boldsymbol{e}_z$ are unit vectors in the direction of the axes shown in Fig. 40. If we write the energy of the switching magnet in the anisotropy field as $-\boldsymbol{H}_A\langle \boldsymbol{S}_{\text{tot}}\rangle$, where $\langle \boldsymbol{S}_{\text{tot}}\rangle$ is the total spin angular momentum of the switching magnet, then H_{u0}, H_{p0} which measure the strengths of the uniaxial and easy-plane anisotropies have dimensions of frequency. These quantities may be converted to a field in tesla by multiplying them by $\hbar/2\mu_B = 5.69 \times 10^{-12}$.

The next problem is to choose the correct phenomenological form of the spin-transfer torque $\boldsymbol{T}^{s-t}$. Without loss of generality, the total spin-transfer torque $\boldsymbol{T}^{s-t}$ may be written as the sum of the two components in the directions of the vectors $\mathbf{m} \times \mathbf{p}$ and $\mathbf{m} \times (\mathbf{p} \times \mathbf{m})$, where **p** is a unit vector in the direction of the magnetization of the polarizing magnet. Thus

$$\boldsymbol{T}^{s-t} = \boldsymbol{T}_\perp + \boldsymbol{T}_\parallel \tag{77}$$

where

$$\boldsymbol{T}_\perp = (g^{ex} + g_\perp eV_b)(\boldsymbol{m} \times \boldsymbol{p}) \tag{78}$$

$$\boldsymbol{T}_\parallel = g_\parallel eV_b \boldsymbol{m} \times (\boldsymbol{p} \times \boldsymbol{m}) \tag{79}$$

V_b is the applied bias and e is the electronic charge (a negative quantity). The bias-independent term g^{ex} in the perpendicular component of the torque corresponds to the usual zero-bias oscillatory exchange coupling of two magnetic layers separated by a non-magnetic spacer [115]. The amplitude of this oscillatory torque tends to zero with increasing thickness of the spacer layer (see Section 4), and, in what follows, we shall assume that the spacer is thick enough for this term to be negligible. We assume here implicitly that the terms proportional to the bias remain finite for an arbitrary thickness of the nonmagnetic spacer layer. This is confirmed by microscopic calculations [110, 115] in the ballistic limit. The spin-transfer torque depends only on the angle ψ between the magnetizations of the polarizing and switching magnets **p**, **m** but not on the orientations of **p** and **m** relative to the coordinate axes. It follows that the coefficients $g_\perp$ and $g_\parallel$ we have introduced in Eqs. (78) and (79) are functions only of ψ. The modulus of both vector products in Eqs. (78) and (79) is equal to $\sin\psi$ and microscopic calculations [110, 115] show that the $\sin\psi$ factor accounts

for most of the angular dependence of $\boldsymbol{T}_\perp$ and $\boldsymbol{T}_\parallel$. In fact, to a good approximation $g_\perp$ and $g_\parallel$ may be regarded as constant parameters which fully determine the phenomenological spin-transfer torque.

We are now ready to study the time evolution of the unit vector $\mathbf{m}$ in the direction of the switching magnet moment. The LLG equation takes the usual form

$$\frac{d\boldsymbol{m}}{dt} + \gamma \boldsymbol{m} \times \frac{d\boldsymbol{m}}{dt} = \boldsymbol{\Gamma} \tag{80}$$

where the reduced total torque $\boldsymbol{\Gamma}$ acting on the switching magnet is given by

$$\boldsymbol{\Gamma} = \frac{[(\boldsymbol{H}_A + \boldsymbol{H}_{\text{ext}}) \times \langle \boldsymbol{S} \rangle + \boldsymbol{T}_\perp + \boldsymbol{T}_\parallel]}{|\langle \boldsymbol{S} \rangle|} \tag{81}$$

Here, $\boldsymbol{H}_{\text{ext}}$ is an external field, in the same frequency units as $\boldsymbol{H}_A$, and γ is the Gilbert damping parameter. Following Sun [116], Eq. (80) may be written more conveniently as

$$(1 + \gamma^2)\frac{d\boldsymbol{m}}{dt} = \boldsymbol{\Gamma} - \gamma \boldsymbol{m} \times \boldsymbol{\Gamma} \tag{82}$$

It is also useful to measure the strengths of all the torques in units of the strength of the uniaxial anisotropy [116]. We shall, therefore, write the total reduced torque Γ in the form

$$\boldsymbol{\Gamma} = H_{u0}\left\{(\boldsymbol{m} \cdot \boldsymbol{e}_z)\boldsymbol{m} \times \boldsymbol{e}_z - \boldsymbol{h}_p(\boldsymbol{m} \cdot \boldsymbol{e}_y)\boldsymbol{m} \times \boldsymbol{e}_y + v_\parallel(\psi)\boldsymbol{m} \times (\boldsymbol{p} \times \boldsymbol{m}) + [v_\perp(\psi) + h_{\text{ext}}]\boldsymbol{m} \times \boldsymbol{p}\right\} \tag{83}$$

where the relative strength of the easy plane anisotropy $h_p = H_{p0}/H_{u0}$ and $v_\parallel(\psi) = vg_\parallel(\psi)$, $v_\perp(\psi) = vg_\perp(\psi)$ measure the strengths of the torques $\boldsymbol{T}_\parallel$ and $\boldsymbol{T}_\perp$. The reduced bias is defined by $v = eV_b/(|\langle \boldsymbol{S}_{\text{tot}} \rangle| H_{u0})$ and has the opposite sign from the bias voltage since e is negative. Thus positive v implies a flow of electrons from the polarizing to the switching magnet. The last contribution to the torque in Eq. (83) is due to the external field H_{ext} with $h_{\text{ext}} = H_{\text{ext}}/H_{u0}$. The external field is taken in the direction of the magnetization of the polarizing magnet, as is the case in most experimental situations.

It follows from Eq. (80) that in a steady state $\boldsymbol{\Gamma} = 0$. We shall defer the discussion of how steady-state solutions $\boldsymbol{m} = \boldsymbol{m}_0 = (\sin\alpha_0 \cos\phi_0, \sin\alpha_0 \sin\phi_0, \cos\alpha_0)$ of this equation are obtained in general. (The polar angles α, ϕ are defined in Fig. 40.) We shall first consider some cases of experimental importance where the steady-state solutions are trivial and the important physics is concerned entirely with their stability. To discuss stability, we linearize Eq. (82), using Eq. (83), about a steady state solution $\boldsymbol{m} = \boldsymbol{m}_0$. Thus

$$\boldsymbol{m} = \boldsymbol{m}_0 + \xi \boldsymbol{e}_\alpha + \eta \boldsymbol{e}_\phi \tag{84}$$

where $\boldsymbol{e}_\alpha$, $\boldsymbol{e}_\phi$ are unit vectors in the direction $\boldsymbol{m}$ moves when α and ϕ are increased independently. The linearized equation may be written in the form

$$\frac{d\xi}{d\tau} = A\xi + B\eta; \quad \frac{d\eta}{d\tau} = C\xi + D\eta \tag{85}$$

Following Sun [116], we have introduced the natural dimensionless time variable $\tau = tH_{u0}/(1 + \gamma^2)$. The conditions for the steady state to be stable are

$$F = A + D \le 0; \quad G = AD - BC \ge 0 \tag{86}$$

excluding $F = G = 0$ [120]. For simplicity, we give these conditions explicitly only for the case $v'_\parallel(\psi_0) = v'_\perp(\psi_0) = 0$, with $\psi_0 = \cos^{-1}(\boldsymbol{p}\boldsymbol{m}_0)$, or $\boldsymbol{m}_0 = \pm\boldsymbol{p}$. The case $\boldsymbol{m}_0 = \pm\boldsymbol{p}$ is very common experimentally as is discussed below. The stability condition $G \ge 0$ may be written

$$\begin{aligned} &Q^2 v_\parallel^2 + (Qh + \cos 2\alpha_0)(Qh + \cos^2\alpha_0) + h_p\{Qh(1 - 3\sin^2\phi_0 \sin^2\alpha_0) \\ &\quad + \cos 2\alpha_0(1 - 2\sin^2\alpha_0 \sin^2\phi_0)\} - h_p^2 \sin^2\alpha_0 \sin^2\phi_0(1 - 2\sin^2\phi_0 \sin^2\alpha_0) \ge 0 \end{aligned} \tag{87}$$

where $v_\parallel = v_\parallel(\psi_0)$, $h = v_\perp(\psi_0) + h_{\text{ext}}$ and $Q = \cos\psi_0$. The condition $F \le 0$ takes the form

$$-2(v_\parallel + \gamma h)Q - \gamma(\cos 2\alpha_0 + \cos^2\alpha_0) - \gamma h_p(1 - 3\sin^2\phi_0 \sin^2\alpha_0) \le 0 \tag{88}$$

We now discuss several interesting examples, the first of those relating to experiments of Grollier et al. [118] and others. In these experiments, the magnetization of the polarizing magnet, the uniaxial anisotropy axis and the external filed are all collinear (along the in-plane z axis in our convention). In this case, the equation $\Gamma = 0$, with Γ given by Eq. (83), shows immediately that possible steady states are given by $\boldsymbol{m}_0 = \pm\boldsymbol{p}(\alpha_0 = 0, \pi)$, corresponding to the switching magnet moment along the z axis. These are the only solutions when $h_p = 0$. For $h_p \neq 0$ other steady-state solutions may exist but in the parameter regime which has been investigated they are always unstable [115]. We shall assume this is always the case and concentrate on the solutions $\boldsymbol{m}_0 = \pm\boldsymbol{p}$. In the state of parallel magnetization (P) $\boldsymbol{m}_0 = \boldsymbol{p}$, we have $v_\parallel = vg_\parallel(0)$, $h = vg_\perp(0) + h_{\rm ext}$, $\alpha_0 = 0$ and $Q = 1$. The stability conditions (87) and (88) become

$$[g_\parallel(0)]^2 v^2 + (vg_\perp(0) + h_{\rm ext} + 1)^2 + h_p[vg_\perp(0) + h_{\rm ext} + 1] \geq 0 \tag{89}$$

$$g_\parallel(0)v + \gamma\left[vg_\perp(0) + h_{\rm ext} + 1 + \frac{1}{2}h_p\right] \geq 0 \tag{90}$$

In the state of antiparallel magnetization (AP) $\boldsymbol{m}_0 = -\boldsymbol{p}$, we have $v_\parallel = vg_\parallel(\pi)$, $h = vg_\perp(\pi) + h_{\rm ext}$, $\alpha_0 = \pi$ and $Q = -1$. The stability conditions for the AP state are thus

$$[g_\parallel(\pi)]^2 v^2 + (-vg_\perp(\pi) - h_{\rm ext} + 1)^2 + h_p[-vg_\perp(\pi) - h_{\rm ext} + 1] \geq 0 \tag{91}$$

$$g_\parallel(\pi)v + \gamma\left[vg_\perp(\pi) + h_{\rm ext} - 1 - \frac{1}{2}h_p\right] \leq 0 \tag{92}$$

In the regime of low external field ($h_{\rm ext} \approx 1$, i.e., $H_{\rm ext} \approx H_{k0}$) we have $H_p \gg h_{\rm ext}$ ($h_p \approx 100$). Eqs. (89) and (91) may then be approximated by

$$vg_\perp(0) + h_{\rm ext} + 1 > 0 \tag{93}$$

$$vg_\perp(\pi) + h_{\rm ext} - 1 < 0 \tag{94}$$

Equation (93) corresponds to P stability and (94) to AP stability. It is convenient to define scalar quantities $T_\perp$, $T_\parallel$ by $T_\perp = g_\perp(\psi)\sin\psi$, $T_\parallel = g_\parallel(\psi)\sin\psi$, these being scalar components of spin-transfer torque in units of eV_b (cf. Eqs. (78) and (79)). Then $g(0) = [dT/d\psi]_{\psi=0}$ and $g(\pi) = -[dT/d\psi]_{\psi=\pi}$. Model calculations [115] show that both $g_\perp$ and $g_\parallel$ can be of either sign, although positive values are more common. Also there is no general rule about the relative magnitude of $g_i(0)$ and $g_i(\pi)$.

We now illustrate the consequences of the above stability conditions by considering two limiting cases. We first consider the case $g_\perp(\psi) = 0$, $g_\parallel(\psi) > 0$, as assumed by Grollier et al. [118] in the analysis of their data. In Fig. 43, we plot the regions of P and AP stability, deduced from Eqs. (90), (92)–(94), in the $v - h_{\rm ext}$ plane. Grollier et al. plot current instead of bias, but this should not change the form of the figure. Theirs is rather more complicated, owing to a less transparent stability analysis with unnecessary approximations. The only approximations made above, to obtain Eqs. (93) and (94), can easily be removed, which results in the critical field lines $h_{\rm ext} = \pm 1$ acquiring a very slight curvature given by $h_{\rm ext} \approx 1 + [vg_\parallel(\pi)]^2/h_p$ and $h_{\rm ext} \approx -1 - [vg_\parallel(0)]^2/h_p$. The critical biases in the figure are given by

$$v_{AP\to P} = \gamma\frac{[1 + \frac{1}{2}h_p - h_{\rm ext}]}{g_\parallel(\pi)}, \quad v_{P\to AP} = -\gamma\frac{[1 + \frac{1}{2}h_p + h_{\rm ext}]}{g_\parallel(0)} \tag{95}$$

Since the damping parameter γ is small ($\gamma \approx 0.01$) the downward slope of the critical bias lines is also small. From Fig. 43, we can deduce the behavior of resistance versus bias in the external field regimes $|h_{\rm ext}| < 1$ and $|h_{\rm ext}| > 1$.

Consider first the case $|h_{\rm ext}| < 1$. Suppose we start in the AP state with a bias $v = 0$, which is gradually increased to $v_{AP\to P}$. At this point the AP state becomes unstable and the system switches to the P state as v increases further. On reducing v the hysteresis loop is completed via a switch back to the AP state at the negative bias $v_{P\to AP}$. The hysteresis loop is shown in Fig. 44(a). The increase in resistance R between the P and AP states is

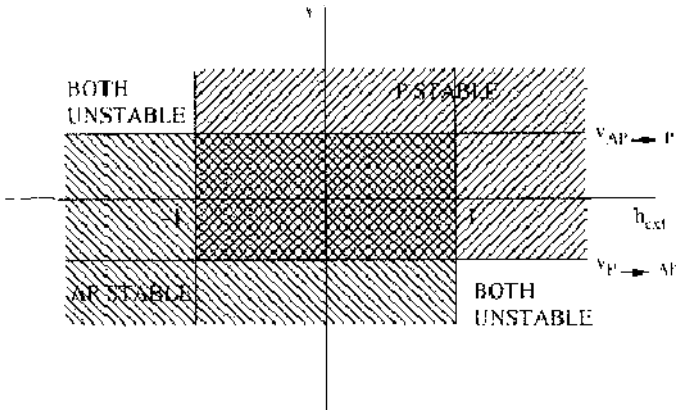


Figure 43. Bias-field stability diagram for $g_{\perp}(\psi) = 0$, $g_{\parallel}(\psi) > 0$.

the same as would be produced by varying the applied field in a GMR experiment. Now consider the case $h_{ext} < -1$. Starting again in the AP state at $v = 0$, we see from Fig. 43 that, on increasing v to $v_{AP\to P}$, the AP state becomes unstable, but there is no stable P state to switch to. This point is marked by an asterisk in Fig. 44(b). For $v > v_{AP\to P}$, the moment of the switching magnet is in a persistently time-dependent state. However, if v is now decreased below $v_{P\to AP}$, the system homes in on the stable AP state and the overall behavior is reversible (i.e., no switching and no hysteresis occur). When $h_{ext} > 1$ similar behavior, now involving the P state, occurs at negative bias, as shown in Fig. 44(b). The dashed curves in Fig. 44(b) show a hypothetical time-averaged resistance in the regions of time-dependent magnetization. As discussed later, time-resolved measurements of resistance suggest that several different types of dynamics can occur in these regions.

It is clear from Fig. 44(a) that the jump AP → P always occurs for positive reduced bias v, which corresponds to flow of electrons from the polarizing to the switching magnet. This result depends on the assumption that $g_{\parallel} > 0$; if $g_{\parallel} < 0$ it is easy to see that the sense of the hysteresis loop is reversed and the jump P → AP occurs for positive v. To our knowledge, this reverse jump has never been observed, although $g_{\perp} < 0$ can occur in principle and is predicted theoretically [115] for the Co/Cu/Co(111) system with a switching magnet consisting of a single atomic plane of Co. It follows from Eq. (95) that $|\, v_{P\to AP}/v_{AP\to P} \,| = |\, g_{\perp}(\pi)/g_{\parallel}(0) \,|$ in zero external field. Experimentally this ratio, essentially the same as the ratio of critical currents, may be considerably less than 1 (e.g., <0.5 [121]), greater than 1 (e.g., ≈2 [122]), or close to 1 [118]. Usually the field dependence of the critical current is found to be stronger than that predicted by Eq. (95) Ref. [121, 118].

We now discuss the reversible behavior shown in Fig. 44(b), which occurs for $|\, h_{ext} \,| > 1$. The transition from hysteretic to reversible behavior at critical external field seems to have been first seen in pillar structures by Katine et al. [123]. Curves similar to the lower one in Fig. 44(b) are reported with $|\, v_{P\to AP} \,|$ increasing with increasing h_{ext}, as expected from Fig. 43. Plots of the differential resistance dV/dI show a peak near the point of maximum

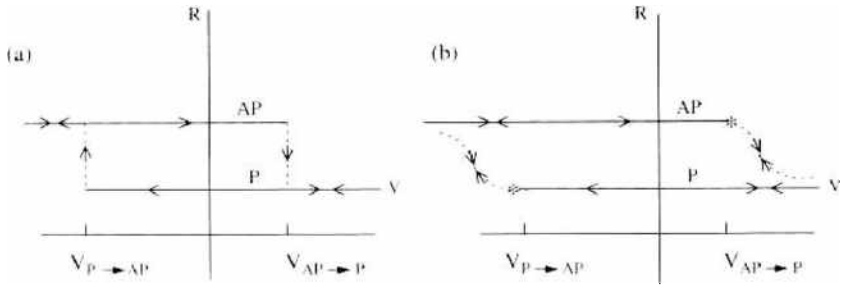


Figure 44. (a) Hysteresis loop of resistance versus bias for $|\, h_{ext} \,| < 1$. (b) Reversible behavior (no hysteresis) for $h_{ext} < -1$ (upper curve) and $h_{ext} > 1$ (lower curve). The dashed lines represent hypothetical behavior of average resistance in regions of Fig. 43 marked "both unstable" where no steady states exist.

gradient of the dashed curve. Similar behavior has been reported by several groups [124–126]. It is particularly clear in the work of Kiselev et al. [124] that the transition from hysteretic behavior (as in Fig. 44[a]) to reversible behavior with peaks in dV/dI occurs at the coercive field 600 Oe of the switching layer ($h_{ext} = 1$). The important point about the peaks in dV/dI is that for a given sign of h_{ext} they only occur for one sign of the bias. This clearly shows that this effect is due to spin-transfer and not to Oersted fields. Myers et al. [127] show a current-field stability diagram similar to the bias-field one of Fig. 43 with a critical field of 1500 Oe. They examine the time dependence of the resistance at room temperature with the field and current adjusted so that the system is in the "both unstable" region in the fourth quadrant of Fig. 43 but very close to its top left-hand corner. They observe telegraph-noise-type switching between approximately P and AP states with slow switching times in the range 0.1–10 s. Similar telegraph noise with faster switching times was observed by Urazhdin et al. [125] at current and field close to a peak in dV/dI. In the region of P and AP instability, Kiselev et al. [124] and Puffal et al. [126] report various types of dynamics of precessional type and random telegraph switching type in the microwave GHz regime. Kiselev et al. [124] propose that systems of the sort considered here might serve as nanoscale microwave sources or oscillators, tunable by current and field over a wide frequency range.

We now return to the stability conditions (90), (92)–(94) and consider the case of $g_\perp(\psi) \neq 0$ but $h_{ext} = 0$. The conditions of stability of the P state may be written approximately, remembering that $\gamma \ll 1$, $h_p \gg 1$, as

$$vg_\perp(0) > -1, \quad vg_\parallel(0) > -\frac{1}{2}\gamma h_p \tag{96}$$

The conditions for stability of the AP state are

$$vg_\perp(\pi) < 1, \quad vg_\parallel(\pi) < \frac{1}{2}\gamma h_p \tag{97}$$

In Fig. 45, we plot the regions of P and AP stability, assuming $g_\perp(0) = g_\perp(\pi) = g_\perp$ and $g_\parallel(0) = g_\parallel(\pi) = g_\parallel$ for simplicity. We also put $r = g_\perp/g_\parallel$. For $r > 0$, we find the normal hysteresis loop as in Fig. 44(a) if we plot R against $vg_\parallel$ (valid for either sign of $g_\parallel$). In Fig. 46, we plot the hysteresis loops for the case $r_c < r < 0$ and $r < r_c$, where $r_c = -2/(\gamma h_p)$ is the value of r at the point x in Fig. 45. The points labeled by asterisks have the same significance as in Fig. 44(b). If in Fig. 46(a) we increase $vg_\parallel$ beyond its value indicated by the right-hand asterisk, we move into "both unstable" region, where the magnetization direction of the switching magnet is perpetually in a time-dependent state. Thus negative r introduces

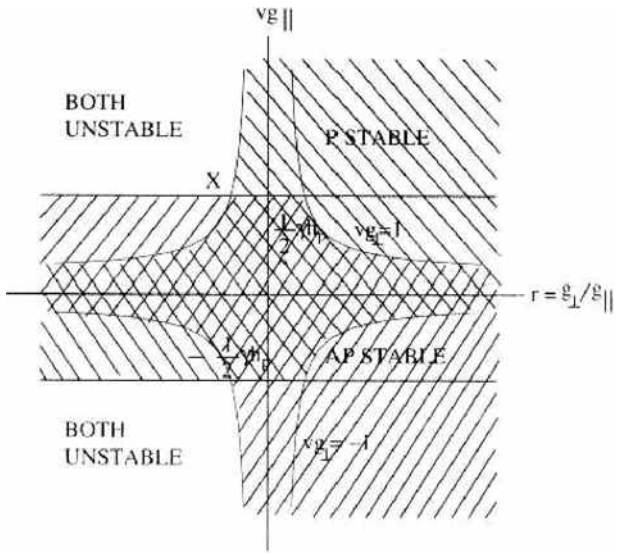


Figure 45. Stability diagram for $h_{ext} = 0$.

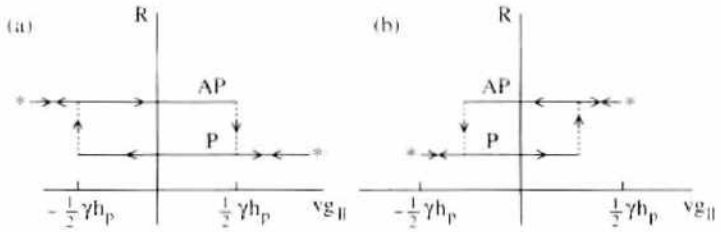


Figure 46. Hysteresis loop for (a) $r_c < r < 0$; (b) $r < r_c$.

behavior in zero applied field, which is similar to that found when the applied field exceeds the coercive field of the switching magnet for $r = 0$. This behavior was predicted by Edwards et al. [115], in particular for a Co/Cu/Co(111) system with the switching magnet consisting of a Co monolayer. Zimmler et al. [128] use methods similar to the ones described here to analyze their data on a Co/Cu/Co nanopillars and deduce that $g_\parallel > 0$, $r = g_\perp/g_\parallel \approx -0.2$. It would be interesting to carry out time-resolved resistance measurements in this system at large current density (corresponding to $vg_\perp < -1$) and zero external field.

So far we have considered the low-field regime ($H_{\text{ext}} \approx$ coercive field of switching magnet) with both magnetizations and the external field in-plane. There is another class of experiments in which a high field, greater than the demagnetizing field ($>2T$), is applied perpendicular to the plane of the layers. The magnetization of the polarizing magnet is then also perpendicular to the plane. This is the situation in the early experiments where a point contact was employed to inject high current densities into magnetic multilayers [127, 129, 130]. In this high-field regime a peak in the differential resistance dV/dI at a critical current was interpreted as the onset of current-induced excitation of spin waves in which the spin-transfer torque leads to uniform precession of the magnetization [129–131]. No hysteretic magnetization reversal was observed and it seemed that the effect of spin-polarized current on the magnetization is quite different in the low- and high-field regimes. Recently, however, Ozyilnez et al. [132] have studied Co/Cu/Co nanopillars ($\approx$100 nm in diameter) at $T = 4.2$ K for large applied fields perpendicular to the layers. They observe hysteretic magnetization reversal and interpret their results using the Landau-Lifshitz equation. We now give a similar discussion within the framework of this section.

Following Ozyilnez et al., we neglect the uniaxial anisotropy term in Eq. (83) for the reduced torque $\boldsymbol{\Gamma}$ while retaining H_{u0} as a scalar factor. Hence

$$\boldsymbol{\Gamma} = H_{u0}\{[h_{\text{ext}} + v_\perp(\psi) - h_p\cos\psi]\boldsymbol{m}\times\boldsymbol{p} + v_\parallel(\psi)\boldsymbol{m}\times(\boldsymbol{p}\times\boldsymbol{m})\} \tag{98}$$

where $\boldsymbol{p}$ is the unit vector perpendicular to the plane. When $v_\parallel(\psi) \neq 0$ the only possible steady-state solutions of $\boldsymbol{\Gamma} = 0$ are $\boldsymbol{m}_0 = \pm\boldsymbol{p}$. On linearizing Eq. (82) about $\boldsymbol{m}_0$ as before we find that the condition $G \geq 0$ is always satisfied. The second stability condition $F < 0$ becomes

$$\{v_\parallel(\psi_0) + \gamma[v_\perp(\psi_0) + h_{\text{ext}} - h_p]\}\cos\psi_0 > 0 \tag{99}$$

where $\psi_0 = \cos^{-1}(\boldsymbol{m}_0\boldsymbol{p})$. Applying this to the P state ($\psi_0 = 0$) and the AP state ($\psi_0 = \pi$), we obtain the conditions

$$v > \gamma\frac{(h_p - h_{\text{ext}})}{g(0)} \tag{100}$$

$$v < -\gamma\frac{(h_p + h_{\text{ext}})}{g(\pi)} \tag{101}$$

where the first condition applies to the P stability and the second to the AP stability. Here, $g(\psi) = g_\parallel(\psi) + \gamma g_\perp(\psi)$. The corresponding stability diagram is shown in Fig. 47, where we have assumed $g(\pi) > g(0) > 0$ for definiteness. The boundary lines cross at $h_{\text{ext}} = h_c$, where $h_c = h_p[g(\pi) + g(0)]/[g(\pi) - g(0)]$. This analysis is valid for fields larger than the demagnetizing field ($h_{\text{ext}} > h_p$), and we see from the figure that for $h > h_c$ hysteretic

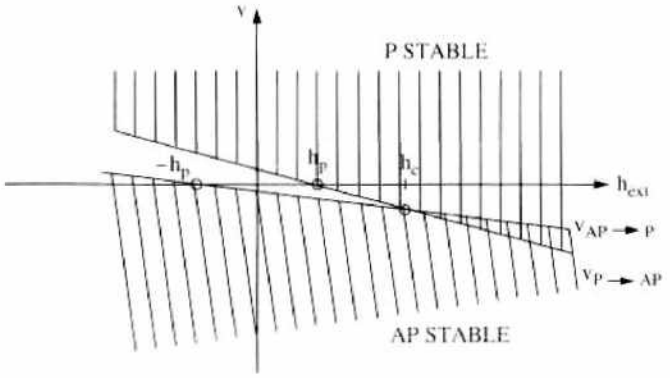


Figure 47. Bias-field stability diagram for large external field ($h_{ext} > h_p$) perpendicular to the layers.

switching occurs. This takes place for only one sign of the bias (current) and the critical biases (currents) increase linearly with h_{ext} as does the width of the hysteresis loop $| v_{P \to AP} - v_{AP \to P} |$. This accords with the observations of Ozyilnez et al. The critical currents are not larger than those in the low-field or zero-field regimes (cf. Eqs. (100) and (101) with Eq. (95)) and yet the magnetization of the switching magnet can be switched against a very large external fields.

7.2. Microscopic Calculation of the Spin-Transfer Torque

To implement the discussion of the stability of steady states for a specific system, we need a method for calculating microscopically the spin-transfer torques $T_\parallel$ and $T_\perp$ we have introduced phenomenologically. We first discuss briefly the microscopic basis for balancing spin-transfer torques with anisotropy torques to obtain a steady state in which the magnetization of the switching magnet is stationary.

We recall that the Hamiltonian of the layer structure depicted in Fig. 40 has the form

$$H = H_0 + H_{\text{int}} + H_{\text{anis}} \tag{102}$$

where H_0 is the Hamiltonian of noninteracting electrons (the hopping term in our tight-binding description of Section 3), H_{int} is an on-site interaction between electrons that leads to an exchange splitting of the bands in the ferromagnets and is neglected in the spacer and lead, and H_{anis} contains effective fields in the switching magnet corresponding to uniaxial $\boldsymbol{H}_u$ and easy-plane $\boldsymbol{H}_p$ anisotropies defined by Eqs. (75) and (76). Although we shall assume in what follows such a tight-binding model, the arguments we give here hold for any Hamiltonian which can be written in a localized basis, for example, tight-binding LMTO [24].

If $\boldsymbol{S}_n$ is the spin angular momentum operator of the atomic plane n in the switching magnet, defined in Section 3, the rate of change of $\boldsymbol{S}_n$ is given by

$$i\hbar \dot{\boldsymbol{S}}_n = [\boldsymbol{S}_n, H_0] + [\boldsymbol{S}_n, H_{\text{anis}}] \tag{103}$$

This result holds since the spin operator commutes with the interaction Hamiltonian H_{int}.

It is straightforward to show that

$$[\boldsymbol{S}_n, H_0] = i\hbar(\boldsymbol{j}_{n-1} - \boldsymbol{j}_n) \tag{104}$$

where $\boldsymbol{j}_{n-1}$ is the operator for spin angular momentum current between planes $n-1$ and n defined by Eq. (15) of Section 3, and

$$[\boldsymbol{S}_n, H_{\text{anis}}] = -i\hbar(\boldsymbol{H}_A \times \boldsymbol{S}_n) \tag{105}$$

In a steady state, the magnetization is time independent so that $\langle \dot{\boldsymbol{S}}_n \rangle = 0$. Hence

$$\langle \boldsymbol{j}_{n-1} \rangle - \langle \boldsymbol{j}_n \rangle = \boldsymbol{H}_A \times \langle \boldsymbol{S}_n \rangle \tag{106}$$

The left-hand side of Eq. (106) corresponds to the rate of transfer of spin angular momentum to plane n in the steady state. Thus Eq. (106) shows explicitly how, in the steady state, this spin-transfer torque is balanced by the torque due to anisotropy fields. The concept of spin-transfer torque was first introduced by Slonczewski [110].

It should be noted that a rigorous self-consistent solution of Eq. (106) requires balancing the spin-transfer and anisotropy torques in every atomic plane of the switching magnet [115]. There are two problems that arise in the fully self-consistent formulation. Firstly, the local magnetization direction in a steady state need not be uniform; that is, the orientation of the switching magnet moment may vary from one atomic plane to another. To make contact with the LLG phenomenological description, which assumes that the switching magnet moment is uniform, we need to show that the assumption of a uniform magnetization is a good approximation.

To clarify the argument (for detailed justification see Ref. [115]), let us first assume that instead of anisotropy fields we apply a weak external magnetic field to keep the switching magnet moment in any specified direction. It is clear that in the limit of an infinitesimal applied bias (liner-response limit) the spin-transfer torques in Eq. (106) are also infinitesimal, and we only need an infinitesimal applied field to balance them and thus stabilize the switching magnet moment in any assumed direction. Any deviations from uniformity due to spin-transfer torques acting on individual atomic planes of the switching magnet would result in strong torques of opposite sign due to the exchange stiffness of the magnet which tends to keep the magnetization uniform. Since the exchange stiffness of a typical magnet such as Co is very large and the spin-transfer torques are infinitesimal, any deviations from uniformity are negligible in the limit $V_b \to 0$.

The second problem is that one should calculate the spin-transfer torques self-consistently [115], which means that one needs to include the applied field (anisotropy fields) in the calculation of the one-electron Green functions which determine the spin current given by Eq. (24) in the Keldysh formalism of Section 3. However, as discussed above, in the limit $V_b \to 0$ the applied magnetic field is negligibly weak compared with the exchange field; therefore, the total effective field, which enters in the calculation of the Green functions, is just the exchange field. It follows that we can calculate the spin-transfer torques assuming that the switching magnet moment is uniform and also neglecting the applied field in the one-electron Green functions. These two approximations become exact in the limit $V_b \to 0$. In practice, the anisotropy fields are much weaker than $\approx 1\%$ of the exchange field and, therefore, the aforementioned approximations are very well justified. We shall, therefore, adopt them in this review and discuss calculations of the spin-transfer torques assuming that the exchange fields of the polarizing and switching magnet are at a fixed arbitrary angle ψ. For convenience, we shall assume that the switching magnet moment is fixed in the direction of the z axis ($\alpha = \phi = 0$ in Fig. 40). It follows that $\psi = \theta$, where θ is the angle between the polarizing magnet moment and the z axis of the system of coordinates in Fig. 40.

Given that the switching magnet moment is assumed to be uniform, we can sum trivially in Eq. (106) over all planes n of the switching magnet, which yields

$$\langle \boldsymbol{j}_{\text{spacer}} \rangle_2 - \langle \boldsymbol{j}_{\text{lead}} \rangle_2 = \boldsymbol{H}_A \times \sum_n \langle \boldsymbol{S}_n \rangle \tag{107}$$

where we have included only the contributions $\langle \boldsymbol{j}_{\text{spacer}} \rangle_2$ and $\langle \boldsymbol{j}_{\text{lead}} \rangle_2$ to the spin currents in the spacer and lead that are proportional to the applied bias (see Section 3). The contribution $\langle \boldsymbol{j}_{\text{lead}} \rangle_1$ vanishes and the term $\langle \boldsymbol{j}_{\text{spacer}} \rangle_1$ determines the oscillatory exchange coupling discussed in Section 4 that tends to zero in the limit of a thick spacer. The left-hand side of Eq. (107) is, therefore, the spin current absorbed by the switching magnet, that is, the total spin-transfer torque, which is required in the phenomenological LLG equation. The spin currents $\langle \boldsymbol{j}_{\text{spacer}} \rangle_2$ and $\langle \boldsymbol{j}_{\text{lead}} \rangle_2$ determined in Section 3 are given by Eq. (24) and can be readily calculated from the one-electron Green functions of the system which is first cleaved between any two neighboring atomic planes in the spacer and then between any two neighboring planes in the lead. This is a very efficient way of calculating the spin-transfer torque, which can be implemented for a fully realistic band structure [115]. However, the Keldysh method, which is formulated in terms of one-electron Green functions, is less transparent

than the original calculation of Slonczewski [110] for a simple parabolic band. Although it would be difficult to generalize his calculation to a realistic band structure, Slonczewki's method, which relies on simple matching of electron wave functions across the layer structure, is much more intuitive. We shall, therefore, first describe his original calculation and then generalize and reformulate his wave function method in the spirit of the Landauer scattering theory so that we can make direct contact with the Keldysh formalism of Section 3.

Slonczewski evaluated the spin current from one-electron wave functions. Writing a two-component wave function as $\psi = (\psi_\uparrow, \psi_\downarrow)^T$, where $\psi_{\uparrow,\downarrow}$ are the wave functions for $\uparrow$ and $\downarrow$ spin projections, we can derive an expression for the spin current by considering the rate of change of the spin angular momentum $d/dt(\psi^\dagger(1/2)\hbar\boldsymbol{\sigma}\psi)$. Since the time derivative of the wave function is determined only by the kinetic energy operator, it is straightforward to show that the relevant components $J_\parallel$ and $J_\perp$ of the spin current are given by

$$\begin{aligned} J_\parallel &= -\left(\frac{\hbar^2}{2m}\right) \mathrm{Im}\,(\psi_\uparrow^* \psi_\downarrow' - \psi_\uparrow^{*\prime} \psi_\downarrow) \\ J_\perp &= \left(\frac{\hbar^2}{2m}\right) \mathrm{Re}\,(\psi_\uparrow^* \psi_\downarrow' - \psi_\uparrow^{*\prime} \psi_\downarrow) \end{aligned} \tag{108}$$

where $\parallel$ and $\perp$ correspond to the x and y components of the spin current, referring to the system of coordinates in Fig. 40 and prime denotes differentiation with respect to y.

Electrons in the spacer and in the leads see a spin-independent potential $V_{\mathrm{sp},l}$ but the Hartree-Fock potential in the ferromagnets has a spin-dependent component $V^{\mathrm{ex}}_{p,\mathrm{sw}}$ given by

$$V_p^{\mathrm{ex}} = \left(\frac{\Delta_p}{2}\right)\begin{pmatrix} \cos\psi & \sin\psi \\ \sin\psi & -\cos\psi \end{pmatrix}, \quad V_{\mathrm{sw}}^{\mathrm{ex}} = \left(\frac{\Delta_{\mathrm{sw}}}{2}\right)\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{109}$$

where $\Delta_{p,\mathrm{sw}}$ is the exchange splitting between the majority- and minority-spin bands in the polarizing and switching magnets. To evaluate the spin current from Eq. (108), it is necessary to solve the Schrödinger equation by matching wave functions and their derivatives across all the interfaces. In general, this can only be done numerically. However, the problem can be solved analytically in the situation when the polarizing and switching magnets have infinitely large exchange splittings $\Delta_{p,\mathrm{sw}} \to \infty$ (half-metallic ferromagnet) and the potentials in the majority-spin bands and in the spacer and leads have the same value (perfect matching). This is the case considered first by Slonczewski in his original calculation [110]. He showed that the torque $T_\parallel$, which is equal to the spin current $J_\parallel$ in the spacer since the corresponding current in the lead vanishes, is given by

$$T_\parallel = \frac{\hbar}{2\,|\,e\,|} \tan\left(\frac{\psi}{2}\right) \times (\text{charge current}) \tag{110}$$

It should be noted that the torque $T_\parallel$ goes to zero for $\psi \to \pi$ since the charge current for a halfmetallic magnet contains a factor $1 + \cos\psi$. It turns out that $T_\perp$ is strictly zero. These results obtained by Slonczewski are quite rigorous but are only valid for this rather special model. It will be seen that, in general, $T_\perp$ is nonzero for ferromagnets with a finite exchange splitting and can be comparable in magnitude with $T_\parallel$. The interesting result that $T_\perp = 0$ for this model may be traced to an effective reflection symmetry of the system about a plane at the center of the spacer. Although the system with ferromagnets having different thicknesses appears asymmetric, the infinite-exchange splitting makes it equivalent to a symmetric system with semi-infinite magnets. More generally, one can show [115] that the $J_\parallel$ component of the spin current in the spacer always vanishes for a system with reflection symmetry. In general, however, the $J_\parallel$ spin current in the lead is nonzero so that $T_\perp \neq 0$ even for a symmetric system. The result $T_\perp = 0$ for the above model is, therefore, a very special one due to the artifact of a very large exchange splitting in the ferromagnets.

To determine the local spin current in the more realistic case of ferromagnets with a finite exchange splitting, it is necessary not only to match numerically the wave functions across all the interfaces but also to specify the correct boundary conditions in the left and right

leads. In the analogous problem of charge transport this was done by Landauer [112]. He showed quite rigorously that the conductance of a system sandwiched between two reservoirs is given by its total quantum mechanical transmission coefficient. Since the charge current is conserved everywhere the total transmission coefficient (conductance) gives us all the information about transport of charge. Similarly, if we are only interested in the spin current due to a single nonmagnet/magnet interface the concept of conductance remains valid and can be easily generalized to include spin dependent scattering from the local exchange potential. One then defines so called mixing conductance [114]. However, since spin current is not conserved, we need to calculate it locally within the structure. In that case, the knowledge of the transmission coefficient for an isolated interface is not sufficient since we also need to take into account all the internal multiple reflections of carriers from all the interfaces. Their contribution to the local spin current is essential.

To determine the local spin current without any approximations, the Landauer method needs to be applied to the whole magnetic layer structure. This was done by in Ref. [113]. To explain the method, we consider, in the spirit of Landauer, two magnets separated by a nonmagnetic spacer layer and assume that the whole structure is sandwiched between two semi-infinite nonmagnetic reservoirs (leads). An infinitesimal bias V_b is applied between the left and right reservoirs so that steady states in the reservoirs with electron distributions $f(\omega-\mu_L)$ and $f(\omega-\mu_R)$ are maintained. As in Section 3, $\mu_L-\mu_R=eV_b$. Since we assume that V_b is infinitesimal, the one-electron states of the system can be calculated from the Schrödinger equation neglecting the effect of V_b. We take the global spin quantization axis to be the z axis and classify thus the electron spin projections as $\uparrow$ or $\downarrow$.

Electrons of either spin orientation ($\uparrow$, $\downarrow$) are incident on the magnetic trilayer both from the left and right reservoirs. To determine the spin current from Eq. (108), we therefore need to solve four independent one-electron scattering problems. First problem corresponds to an $\uparrow$-spin electron incident from the left, which is partially reflected both to the $\uparrow$-spin and $\downarrow$-spin channels in the left reservoir and partially transmitted to the $\uparrow$-spin and $\downarrow$-spin channels in the right reservoir. Similarly, a $\downarrow$-spin electron incident from the left reservoir is reflected and transmitted to both $\uparrow$- and $\downarrow$-spin channels in the left and right reservoirs, respectively. Finally, both $\uparrow$- and $\downarrow$-spin electrons incident from the right reservoir are similarly reflected and transmitted to both spin channels in the left and right reservoirs, respectively. If $j_i^L(\sigma,\omega,\boldsymbol{k}_\parallel)$ is the spin current component i ($i=\parallel,\perp$) due to an electron of spin σ incident from the left reservoir with an energy ω and parallel wave vector $\boldsymbol{k}_\parallel$, and $j_i^R(\sigma,\omega,\boldsymbol{k}_\parallel)$ is the corresponding contribution due to an electron incident from the right reservoir, we can write the total spin current in any part of the structure by summing over the contributions from all the electrons incident on the magnetic structure

$$J_i^{\text{tot}}=\sum_{\boldsymbol{k}_\parallel}\int d\omega\{f(\omega-\mu_L)[J_i^L(\uparrow,\omega,\boldsymbol{k}_\parallel)+J_i^L(\downarrow,\omega,\boldsymbol{k}_\parallel)]+f(\omega-\mu_R)[J_i^R(\uparrow,\omega,\boldsymbol{k}_\parallel)+J_i^R(\downarrow,\omega,\boldsymbol{k}_\parallel)]\} \tag{111}$$

where $J_i^{L(R)}=j_i^{L(R)}(dk_\perp/d\omega)$ was used to convert the sum over $k_\perp$ into an energy integral. Using the identities $f_R=(1/2)[f_R+f_L]+(1/2)[f_R-f_L]$ and $f_L-(1/2)[f_R+f_L]-(1/2)[f_R-f_L]$, where f_L and f_R denote the Fermi functions in the left and right reservoirs, we can finally write the total spin current as

$$J_i^{\text{tot}}=\sum_{\boldsymbol{k}_\parallel,\sigma}\int d\omega\{[J_i^L(\sigma,\omega,\boldsymbol{k}_\parallel)+J_i^R(\sigma,\omega,\boldsymbol{k}_\parallel)](f_L+f_R)+[J_i^L(\sigma,\omega,\boldsymbol{k}_\parallel)-J_i^R(\sigma,\omega,\boldsymbol{k}_\parallel)](f_L-f_R) \tag{112}$$

The first term in Eq. (112), which is proportional to the sum of the two Fermi functions, is equivalent to the term given by Eq. (23) of Section 3, and the term proportional to the difference between the two Fermi functions is equivalent to Eq. (24) for the spin current derived in Section 3 using the general Keldysh formalism. It is straightforward to evaluate the spin current from Eq. (112) for the parabolic band model using, for example, the transfer matrix method to determine the electron wave functions. However, it would be difficult to implement such a calculation for a realistic band structure and, to our knowledge, no such

calculation has ever been attempted. Nevertheless, the Landauer formula (112) is very useful for discussing some general properties of the spin current.

In the absence of bias, only the terms proportional to the sum of the Fermi functions remain. It is clear on physical grounds that $J_{\parallel}^{tot} = 0$ both in the spacer and in the leads. We also have $J_{\perp}^{tot} = 0$ in the leads but $J_{\perp}^{tot} \neq 0$ in the spacer. The term $J_{\perp}^{tot}$ in the spacer determines the oscillatory exchange coupling between two ferromagnets discussed in Section 4. It is easy to verify that the results for the oscillatory exchange coupling obtained from Eq. (112) are in complete agreement with previous parabolic-band model calculations of this effect [58, 59]. We note that the result $J_i^{tot} = 0$ in the leads in the absence of bias shows explicitly that the contribution of the spin current transmitted through the magnetic structure must be exactly compensated by the contribution due to electrons reflected from the structure. Direct numerical evaluation of Eq. (112) for a parabolic band model confirms this result.

Since the oscillatory exchange coupling tends to zero for a thick spacer layer, we shall now concentrate on the bias-induced term, which is proportional to the difference between the Fermi functions and remains nonzero [110, 115] even for an infinitely thick spacer (in the ballistic limit). This is indeed the term that determines the spin-transfer torques responsible for the current-induced switching of the magnetization discussed in the phenomenological section. In the linear-response limit we are considering, it follows from Eq. (112) that the spin current in any part of the structure is proportional to the bias V_b. Consequently, upon changing the direction of the bias (current), the spin-transfer torque vector changes its direction but not its magnitude. As already discussed in the phenomenological section, the proportionality of the torque to the bias (current) is reflected in the shape of the observed hysteresis loops which are approximately symmetric (in the absence of an external field). This test shows that a recent calculation of the spin-transfer torque [133], which claims that the torque is proportional to the square of the charge current, cannot be correct. Another property that is already implicit in the definition of the spin current (108) is that the spin current is a continuous function of the coordinates. This follows because the electron wave functions in the aforementioned scattering problem are obtained by matching the functions and their derivatives at all the interfaces. Stiles and Zangwill [134] argue in their calculation of the spin current across a nonmagnet/magnet interface that there is a discontinuity in the spin current at the interface. This conclusion is incorrect and the results that follow from such a calculation are thus suspect.

Although full implementation of the parabolic band model using Eq. (112) gives qualitatively correct results, the model itself is too simple to predict the correct magnitudes, relative sign and angular dependencies of the spin transfer torques $T_{\parallel}$, $T_{\perp}$ for specific systems. We therefore conclude this section by giving examples of realistic calculations [115] of the spin-transfer torques for Co/Cu/Co(111) system based on the Keldysh Green function formalism of Section 3.

Referring to Fig. 40, the system considered by Edwards et al. [115] consists of a semi-infinite slab of Co (polarizing magnet), the spacer of 20 atomic planes of Cu, the switching magnet containing M atomic planes of Co with $M = 1$ and 2, and the lead which is semi-infinite Cu. The spacer thickness of 20 atomic planes of Cu was chosen so that the contribution of the oscillatory exchange coupling term is so small that it can be neglected. The spin currents in the right lead and in the spacer were determined from Eq. (24) of Section 3 using the same tight-binding parameterization of an *ab initio* band structure of Co and Cu that was discussed in Section 4. Figures 48(a) and 48(b) shows the angular dependences of $T_{\parallel}$, $T_{\perp}$ for the cases $M = 1$, and $M = 2$, respectively. For the monolayer switching magnet, the torques $T_{\perp}$ and $T_{\parallel}$ are equal in magnitude and they have the opposite sign. However, for $M = 2$, the torques have the same sign and $T_{\perp}$ is somewhat smaller than $T_{\parallel}$. A negative sign of the ratio of the two torque components has important and unexpected consequences for hysteresis loops as already discussed in the phenomenological section. It can be seen that the angular dependence of both $T_{\perp}$ and $T_{\parallel}$ is dominated by a $\sin\psi$ factor but distortions from this dependence are clearly visible. In particular, the slopes of the angular dependences of the torques $T_{\perp}$ and $T_{\parallel}$ at $\psi = 0$ and $\psi = \pi$ are quite different. As pointed out in the phenomenological section, this may be important in the discussion of the stability of steady states.

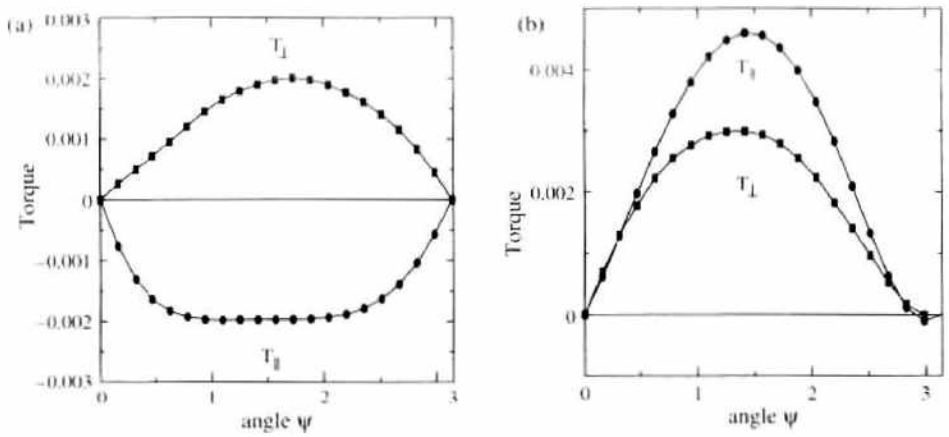


Figure 48. Dependence of the spin-transfer torque $T_\parallel$ and $T_\perp$ for Co/Cu/Co(111) on the angle ϕ. The torques per surface atom are in units of eV_b. (a) $M = 1$ and (b) $M = 2$ monolayers of Co in the switching magnet.

In Fig. 49, we reproduce the dependence of $T_\perp$ and $T_\parallel$ on the thickness of the Co switching magnet. It can be seen that the out-of-plane torque $T_\perp$ becomes smaller than $T_\parallel$ for thicker switching magnets. This is the expected behavior since the polarizing magnet is semi-infinite Co, so that as the switching Co magnet becomes thicker we approach the limit of a symmetric junction for which the $T_\perp$ component of the spin current vanishes and the corresponding component in the lead is usually small. However, $T_\perp$ is by no means negligible (27% of $T_\parallel$) even for a typical experimental thickness of the switching Co layer of 10 atomic planes. It is also interesting that beyond the monolayer thickness, the ratio of the two torques is positive with the exception of $M = 4$.

The microscopically calculated spin-transfer torques for Co/Cu/Co(111) were used by Edwards et al. [115] as an input into the phenomenological LLG equation. The LLG equation was first solved numerically to determine all the steady states and then the stability discussion outlined in the phenomenological section was applied to determine the critical bias for which instabilities occur. Finally, the ballistic resistance of the structure was evaluated from the real-space Kubo-Landauer formula (27) of Section 3 at every point of the steady state path. Such a calculation for the realistic Co/Cu system then gives hysteresis loops of the resistance versus bias which can be compared with the observed hystersis loops. The LLG equation was solved including a strong easy-plane anisotropy with $h_p = 100$. If we take $H_{a0} = 1.86 \times 10^9$ sec^{-1} corresponding to an uniaxial anisotropy field of about 0.01 T, this value of h_p corresponds to the shape anisotropy for a magnetization of 1.6×10^6 A/m, similar to that of Co [116]. Also a realistic value [116] of the Gilbert damping parameter

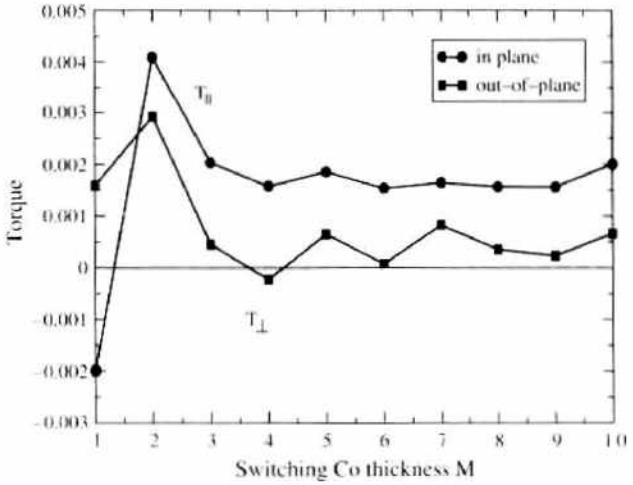


Figure 49. Dependence of the spin-transfer torque $T_\parallel$ and $T_\perp$ for Co/Cu/Co(111) on the thickness of the switching magnet M for $\phi = \pi/3$. The torques are in units of eV_b.

$\gamma = 0.01$ was used. Finally, referring to the geometry of Fig. 40, two different values of the angle θ were employed in these calculations: $\theta = 2$ rad and $\theta = 3$ rad, the latter value being close to the value of π which is realized in most experiments.

We first reproduce in Fig. 50 the hysteresis loops for the case of Co switching magnet consisting of two atomic planes. We recall that the ratio $r = T_\perp / T_\perp \approx 0.65$ deduced from Fig. 48 is positive in this case. Figure 50(a) shows the hysteresis loop for $\theta = 2$ rad and Fig. 50(b) that for $\theta = 3$ rad. The hysteresis loop for $\theta = 3$ rad shown in Fig. 50(b) is an illustration of the stability scenario with $r > 0$ discussed in the phenomenological section. It is rather interesting that the critical bias for switching is $\approx$0.2 mV both for $\theta = 2$ and $\theta = 3$ radians. When this bias is converted to the current density using the calculated ballistic resistance of the junction, it is found [115] that the critical current for switching is $\approx 10^7$ A/cm^2, which is in very good agreement with experiment [117].

The hysteresis loops for the case of the Co switching magnet consisting of a single atomic plane are reproduced in Fig. 51. The values of h_p, γ, H_{a0}, and θ are the same as in the previous example. However, the ratio $r \approx -1$ is now negative and the hysteresis loops in Fig. 51 illustrate the interesting behavior discussed in the phenomenological section when the system subjected to a bias higher than a critical bias moves to the "both unstable" region shown in Fig. 45. As in Fig. 46, the points on the hysteresis loops shown in Fig. 51 corresponding to the critical bias are labeled by asterisks.

The calculations based on the Keldysh formalism of Section 3 (or on the equivalent generalized Landauer formula [112]) we have described are truly microscopic and provide rigorous steady-state spin-transfer torques in the ballistic limit. There are also several semiphenomenological calculations of the spin-transfer torque [135, 136] and we shall now briefly discuss how they are related to the steady-state Keldysh approach adopted here. Heide et al. [135] argued that spin-polarized electrons accumulate in the spacer at the spacer/switching magnet interface. That means that there is a spin angular momentum in the spacer at an angle to the switching magnet moment. This results in an effective exchange interaction between the accumulated spin and the switching magnet moment. The argument here is quite analogous to that used in the derivation of the well-known RKKY interaction (see, e.g., Ref. [137]). The effective exchange interaction between two moments at an angle is clearly associated with a spin-transfer torque. Heide et al. argued that such an effect results only in a component $T_\perp$ of the torque. The microscopic calculations [115] described here confirm that a spin accumulation in the spacer occurs and that an exchange interaction-type term $T_\perp$ exists. However, what is missing in the argument of Heide et al. is that the spin accumulated in the spacer has not only an in-plane component, which gives rise to $T_\perp$, but also an out-of-plane component, which gives rise to the important $T_\parallel$ component of the spin-transfer torque. Moreover, Heide et al. [135] provide no prescription for calculating the magnitude of their $T_\perp$ microscopically.

Zhang et al. [136] argue that there is an accumulation of spin in the switching magnet with components perpendicular to the switching magnet moment. This would necessarily result in a precession of the accumulated spin about the exchange field in the switching magnet. Such a precession clearly cannot take place in a steady state since the total local torque

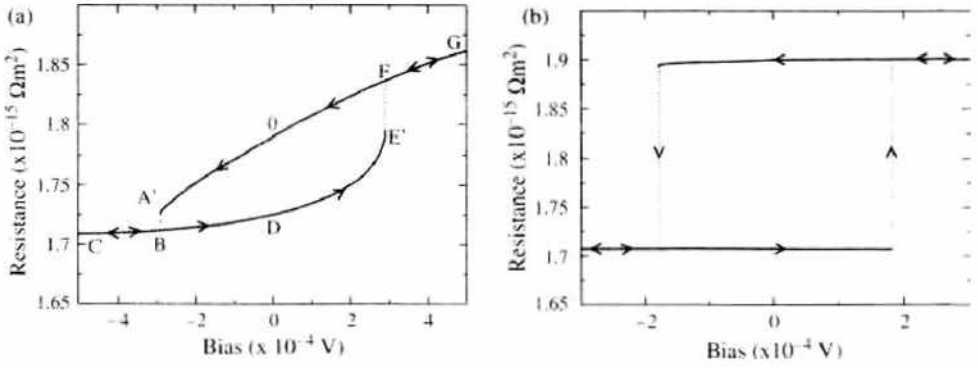


Figure 50. Resistance of the Co/Cu/Co(111) junction as a function of applied bias with $M = 2$ monolayers of Co in the switching magnet. (a) is for $\theta = 2$ radians and (b) is for $\theta = 3$ radians.

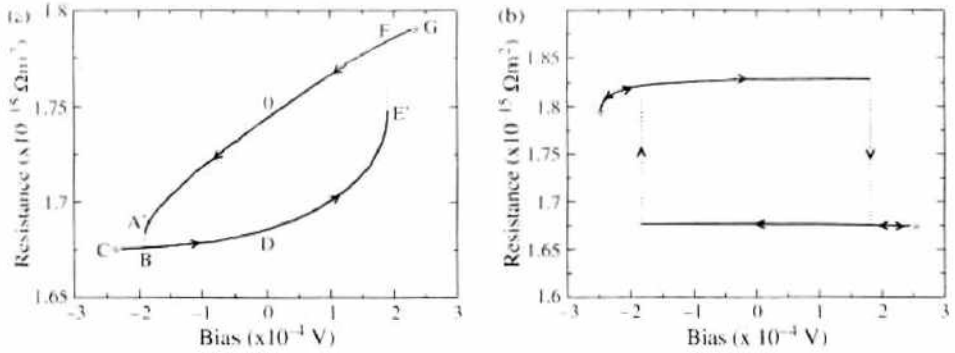


Figure 51. Resistance of the Co/Cu/Co(111) junction as a function of applied current, with $M = 1$ monolayer of Co in the switching magnet. (a) is for $\theta = 2$ radians and (b) is for $\theta = 3$ radians.

acting on the local spin in each atomic plane of the switching magnet is zero (see Eq. [106]). We conclude that the concept of steady-state spin accumulation in the switching magnet is unphysical and the calculation of Zhang et al. thus cannot give the correct steady-state spin-transfer torque.

Spin accumulation in the switching magnet could be a valid concept when the system is not in a steady state. As discussed in the phenomenological section, the switching magnet is not in a steady state for a set of parameters corresponding to the "both unstable" region shown in Fig. 46. In that situation, the steady-state formalism for calculating the spin-transfer torque described here cannot be used. However, to discuss the time-dependent motion of the switching magnet moment quantitatively we need to know the time-dependent spin-transfer torques. This is clearly one of the important problems that needs to be addressed. Another interesting outstanding problem is the effect of impurities/interfacial roughness on current-induced switching. In principle, imperfections could be included in the Keldysh formula (24) using the lateral supercell method but, to our knowledge, it has not been done.

We conclude that the current-induced switching of magnetization is one of the most topical and rapidly developing area of spintronics. In particular, there is rich new physics associated with the observed nonhysteretic (time-dependent) motion of the switching magnet moment. While the steady-state (hysteretic) regime can now be modeled by essentially first-principle calculations of the type described here, rigorous microscopic modeling of the nonhysteretic regime is currently nonexistent, but is urgently required.

REFERENCES

1. A. Fert and I. A. Campbell, *J. Phys. F* 6, 849 (1976).
2. P. Grünberg, R. Schreiber, Y. Pang, M. B. Brodsky, and H. Sower, *Phys. Rev. Lett.* 57, 2442 (1986); M. N. Baibich, J. M. Broto, A. Fert, Van Dau Nguyen, F. Petroff, P. Etienne, G. Creuset, A. Friederich, and J. Chazelas, *Phys. Rev. Lett.* 61, 2472 (1988).
3. J. S. Moodera, L. R. Kinder, T. M. Wong, and R. Meservey, *Phys. Rev. Lett.* 74, 3273 (1995); T. Miyazaki and N. Tezuka, *J. Magn. Magn. Mater.* 139, L231 (1995).
4. J. C. Slonczewski, *J. Magn. Magn. Mater.* 159, L1 (1996).
5. S. S. P. Parkin, Xin Jiang, C. Kaiser, A. Panchula, K. Roche, and M. Samant, *Proceedings of the IEEE* 91, 661 (2003); W. H. Butler and A. Gupta, *Nature Mater.* 3, 845847 (2004); S. S. P. Parkin, C. Kaiser, A. Panchula, P. M. Rice, B. Hughes, M. Samant, and SEE-Hun Yang, *Nature Mater.* 3, 862 (2004).
6. S. S. P. Parkin, N. More, and K. P. Roche, *Phys. Rev. Lett.* 64, 2304 (1990); D. M. Edwards, J. Mathon, R. B. Muniz, and M. S. Phan, *Phys. Rev. Lett.* 67, 493 (1991); P. Bruno and C. Chappert, *Phys. Rev. Lett.* 67, 1602 (1991).
7. P. Grünberg, R. Schreiber, Y. Pang, M. B. Brodsky, and H. Sower, *Phys. Rev. Lett.* 57, 2442 (1986).
8. M. N. Baibich, J. M. Broto, A. Fert, Van Dau Nguyen, F. Petroff, P. Etienne, G. Creuset, A. Friederich, and J. Chazelas, *Phys. Rev. Lett.* 61, 2472 (1988).
9. S. S. P. Parkin, N. More, and K. P. Roche, *Phys. Rev. Lett.* 64, 2304 (1990).
10. R. Schad, C. D. Potter, P. Belien, G. Verbanck, V. V. Moshchalkov, and Y. Bruynseraede, *Appl. Phys. Lett.* 64, 3500 (1994).
11. W. P. Pratt, Jr., S. F. Lee, J. M. Slaughter, R. Loloee, P. A. Schroeder, and J. Bass, *Phys. Rev. Lett.* 66, 3060 (1991); M. A. M. Gijs, S. K. J. Lenczowski, and J. B. Giesbers, *Phys. Rev. Lett.* 70, 3343 (1993).
12. T. Valet and A. Fert, *Phys. Rev. B* 48, 7099 (1993); A. Fert, J. L. Duvail, and T. Valet, *Phys. Rev. B* 52, 6513 (1995).

13. N. F. Mott, *Adv. Phys.* 13, 325 (1964).
14. M. A. M. Gijs and G. E. W. Bauer, *Adv. Phys.* 46, 285 (1997).
15. R. E. Camley and J. Barnas, *Phys. Rev. Lett.* 63, 664 (1989); J. Barnas, A. Fuss, R. E. Camley, P. Grünberg, and W. Zinn, *Phys. Rev. B* 42, 8110 (1990).
16. D. M. Edwards, J. Mathon, and R. B. Muniz, *IEEE Trans. Magn.* 27, 3548 (1991).
17. W. A. Harrison, "Solid State Theory." Dover Publications, New York, 1979.
18. D. M. Edwards, J. Mathon, R. B. Muniz, and S. S. P. Parkin, *J. Magn. Magn. Mater.* 114, 252 (1992).
19. P. A. Schroeder, J. Bass, P. Holody, S. F. Lee, W. P. Pratt, Jr., and Q. Yang, in "Materials Research Society Proceedings." Pittsburgh, 1993, Vol. 313, p. 47.
20. J. Bass, P. A. Schroeder, W. P. Pratt, Jr., S. F. Lee, Q. Yang, P. Holody, L. L. Henry, and R. Loloee, *Mater. Sci. Eng. B* 31, 77 (1995); W. P. Pratt, Jr., S. F. Lee, Q. Yang, P. Holody, R. Loloee, P. A. Schroeder, and J. Bass, *J. Appl. Phys.* 73, 5326 (1993).
21. S. Zhang and P. M. Levy, *J. Appl. Phys.* 69, 4786 (1991); P. M. Levy, *Solid State Phys.* 47, 367 (1994); P. M. Levy and S. Zhang, *J. Magn. Magn. Mater.* 151, 315 (1995).
22. X. G. Zhang and W. H. Butler, *Phys. Rev. B* 51, 10085 (1995); W. H. Butler, X. G. Zhang, D. M. C. Nicholson, and J. M. MacLaren, *Phys. Rev. B* 52, 13399 (1995); P. Zahn, I. Mertig, M. Richter, and H. Eschrig, *Phys. Rev. Lett.* 75, 2996 (1995).
23. E. Y. Tsymbal and P. G. Pettifor, *Phys. Rev. B* 54, 15314 (1996).
24. I. Turek, V. Drchal, J. Kudrnovsky, M. Sob, and P. Weinberger, "Electronic Structure of Disordered Alloys, Surfaces and Interfaces." Kluwer, Dordrecht, 1997.
25. J. Mathon, A. Umerski, and M. Villeret, *Phys. Rev. B* 55, 14378 (1997).
26. L. V. Keldysh, *Sov. Phys. JETP* 20, 1018 (1965).
27. C. Caroli, R. Combescot, P. Nozieres, and D. Saint-James, *J. Phys. C* 4, 916 (1971).
28. D. M. Edwards, in "Exotic States in Quantum Nanostructures" (S. Sarkar, Ed.). Kluwer Academic Press, Dordrecht, 2002.
29. G. D. Mahan, "Many Particle Physics." 2nd Edn., Plenum Press, New York, 1990.
30. R. Landauer, *IBM J. Res. Dev.* 32, 306 (1988); A. D. Stone and A. Szafer, *IBM J. Res. Dev.* 32, 384 (1988).
31. A. Umerski, *Phys. Rev. B* 55, 5266 (1997).
32. D. H. Lee and J. D. Joannopulos, *Phys. Rev. B* 23, 4988 (1981); J. Mathon and S. B. Ahmad, *Phys. Rev. B* 37, 660 (1988).
33. S. Demokritov, J. A. Wolf, and P. Grüunberg, *Europhys. Lett.* 15, 881 (1991).
34. J. Unguris, R. J. Celotta, and D. T. Pierce, *Phys. Rev. Lett.* 67, 140 (1991).
35. J. J. de Miguel, A. Cebollada, J. M. Gallego, R. Miranda, C. M. Schneider, P. Schuster, and J. Kirschner, *J. Magn. Magn. Mater.* 93, 1 (1991).
36. M. T. Johnson, R. Coehoorn, J. J. de Vries, N. W. E. McGee, J. aan de Stegge, and P. J. H. Bloemen, *Phys. Rev. Lett.* 69 (1992).
37. S. S. P. Parkin, R. F. Marks, R. F. C. Farrow, G. R. Harp, Q. H. Lam, and R. J. Savoy, *Phys. Rev. B* 46, 9262 (1992).
38. P. J. H. Bloemen, M. T. Johnson, J. aan de Stegge, and W. J. M. Jonge, *J. Magn. Magn. Mater.* 46, L315 (1992).
39. M. A. Howson, B. J. Hickey, J. Xu, D. Grieg, and N. Wiser, *Phys. Rev. B* 48, 1322 (1993).
40. S. N. Okuno and K. Inomata, *Phys. Rev. Lett.* 70, 1711 (1993).
41. P. J. H. Bloemen, M. T. Johnson, M. T. H. van Vorst, R. Coehoorn, J. J. de Vries, R. Jungblut, J. aan de Stegge, A. Reinders, and W. J. M. de Jonge, *Phys. Rev. Lett.* 72, 764 (1994).
42. D. H. Mosca, F. Petroff, A. Fert, P. A. Schroeder, W. P. Pratt, Jr., and R. Laloe, *J. Magn. Magn. Mater.* 94, L1 (1991).
43. S. S. P. Parkin, R. Bhadra, and K. P. Roche, *Phys. Rev. Lett.* 66, 2152 (1991).
44. D. M. Edwards and J. Mathon, *J. Magn. Magn. Mater.* 93, 85 (1991).
45. D. M. Edwards, J. Mathon, R. B. Muniz, and M. S. Phan, *Phys. Rev. Lett.* 67, 493 (1991).
46. D. M. Edwards, J. Mathon, R. B. Muniz, and M. S. Phan, *J. Phys.* 3, 4941 (1991).
47. J. Mathon, M. Villeret, and D. M. Edwards, *J. Phys.* 4, 9873 (1992).
48. J. Mathon, M. Villeret, R. B. Muniz, J. d'Albuquerque e Castro, and D. M. Edwards, *Phys. Rev. Lett.* 74, 3696 (1995).
49. J. Mathon, M. Villeret, A. Umerski, R. B. Muniz, J. d'Albuquerque e Castro, and D. M. Edwards., *Phys. Rev. B* 56, 11797 (1997).
50. J. d'Albuquerque e Castro, J. Mathon, M. Villeret, and A. Umerski, *Phys. Rev. B* 53, 13306 (1996).
51. C. Kittel, in "Solid State Physics" (F. Seitz, D. Turnbull, and H. Ehrenreich, Eds.), Vol. 22, p. 1. Academic Press, New York, 1968.
52. P. Bruno and C. Chappert, *Phys. Rev. Lett.* 67, 1602 (1991).
53. P. Bruno, *Phys. Rev. B* 52, 411 (1995).
54. M. D. Stiles, *Phys. Rev. B* 44, 9331 (1993).
55. J. E. Ortega, F. J. Himpsel, G. J. Mankey, and R. F. Willis, *Phys. Rev. B* 47, 1540 (1993).
56. K. Garrison, Y. Chang, and P. D. Johnson, *Phys. Rev. Lett.* 71, 2801 (1993).
57. J. C. Slonczewski, *Phys. Rev. B* 39, 6995 (1989).
58. K. B. Hathaway and J. R. Cullen, *J. Magn. Magn. Mater.* 104–107, 1840 (1992).
59. R. P. Erickson, K. B. Hathaway, and J. R. Cullen, *Phys. Rev. B* 47, 2626 (1993).
60. D. M. Edwards, A. M. Robinson, and J. Mathon, *J. Magn. Magn. Mater.* 140–144, 517 (1995).
61. J. Kudrnovsky, V. Drchal, I. Turek, and P. Weinberger, *Phys. Rev. B* 50, 16105 (1994).

62. J. Kudrnovsky, V. Drchal, I. Turek, M. Sob, and P. Weinberger, *Phys. Rev. B* 53, 5225 (1996).
63. V. Drchal, J. Kudrnovsky, I. Turek, and P. Weinberger, *Phys. Rev. B* 53, 15036 (1996).
64. D. A. Papaconstantopoulos, "Handbook of The Band Structure of Elemental Solids," Plenum, New York, 1986.
65. V. L. Moruzzi, J. F. Janak, and A. R. Williams, "Calculated Electronic Properties of Metals," Pergamon, Oxford, 1978.
66. P. Lang, L. Nordström, R. Zeller, and P. H. Dederichs, *Phys. Rev. Lett.* 50, 1927 (1993).
67. L. Nordström, P. Lang, R. Zeller, and P. H. Dederichs, *Phys. Rev. B* 50, 13058 (1994).
68. W. Weber, R. Allenspach, and A. Bischof, *Europhys. Lett.* 31, 431 (1995).
69. K. M. Schep, P. J. Kelly, and G. E. W. Bauer, *Phys. Rev. Lett.* 74, 586 (1995); *J. Magn. Magn. Mater.* 156, 385 (1996).
70. W. P. Pratt, Jr., S. F. Lee, P. Holody, Q. Yang, R. Loloee, J. Bass, and P. A. Schroeder, *J. Appl. Phys.* 73, 5326 (1993).
71. J. Kudrnovsky, V. Drchal, C. Blaas, P. Weinberger, I. Turek, and P. Bruno, *Phys. Rev. B* 62, 15084 (2000).
72. H. Itoh, J. Inoue, and S. Maekawa, *Phys. Rev. B* 51, 342 (1995).
73. J. Mathon, M. Villeret, and H. Itoh, *Phys. Rev. B* 52, R6983 (1995).
74. J. Mathon, *Phys. Rev. B* 55, 960 (1997).
75. R. Meservey and P. M. Tedrow, *Phys. Rep.* 238, 173 (1994).
76. M. Juliere, *Phys. Lett. A* 54, 225 (1975).
77. S. Maekawa and U. Gäfvert, *IEEE Trans. Magn.* MAG-18, 707 (1982).
78. J. Bardeen, *Phys. Rev. Lett.* 6, 57 (1961).
79. J. S. Moodera, L. R. Kinder, T. M. Wong, and R. Meservey, *Phys. Rev. Lett.* 74, 3273 (1995).
80. T. Miyazaki and N. Tezuka, *J. Magn. Magn. Mater.* 139, L231 (1995).
81. J. S. Moodera, J. Nowak, and R. J. M. van de Veerdonk, *Phys. Rev. Lett.* 80, 2941 (1998).
82. J. S. Moodera and J. Mathon, *J. Magn. Magn. Mater.* 200, 248 (1999).
83. J. Mathon, *Phys. Rev. B* 56, 11810 (1997).
84. H. Itoh, A. Shibata, T. Kumazaki, J. Inoue, and S. Maekawa, *J. Phys. Soc. Jpn.* 68, 1632 (1999).
85. E. Y. Tsymbal and D. G. Pettifor, *Phys. Rev. B* 58, 432 (1998).
86. M. B. Stearns, *J. Magn. Magn. Mater.* 5, 167 (1977).
87. J. M. MacLaren, X.-G. Zhang, and W. H. Butler, *Phys. Rev. B* 56, 11827 (1997).
88. E. Y. Tsymbal and D. G. Pettifor, *J. Phys.* 9, L411 (1997).
89. J. Mathon and A. Umerski, *Phys. Rev. B* 60, 1117 (1999).
90. A. M. Bratkovsky, *Phys. Rev. B* 56, 2344 (1997).
91. A. M. Bratkovsky, *Appl. Phys. Lett.* 72, 2334 (1998).
92. J. Halbritter, *Surf. Sci.* 122, 80 (1982).
93. J. S. Moodera, J. Nowak, L. R. Kinder, P. M. Tedrow, R. J. M. van de Veerdonk, B. A. Smits, M. van Kampen, H. J. M. Swagten, and W. J. M. de Jonge, *Phys. Rev. Lett.* 83, 3029 (1999).
94. V. Vedyaev, N. Ryzhanova, C. Lacroix, L. Giacomoni, and B. Dieny, *Europhys. Lett.* 39, 219 (1997).
95. S. Zhang and P. M. Levy, *Phys. Rev. Lett.* 81, 5660 (1998).
96. J. E. Ortega, F. J. Himpsel, G. J. Mankey, and R. F. Willis, *Phys. Rev. B* 47, 1540 (1993); P. Segovia, E. G. Michel, and J. E. Ortega, *Phys. Rev. Lett.* 77, 3455 (1996).
97. S. Yuasa, T. Nagahama, and Y. Suzuki, *Science* 297, 234 (2002).
98. H. Itoh, J. Inoue, A. Umerski, and J. Mathon, *Phys. Rev. B* 68, 174421 (2003).
99. W. Wulfhekel, B. Heinrich, M. Klaua, T. Monchesky, F. Zavaliche, R. Urban, and J. Kirschner, *Appl. Phys. Lett.* 78, 509 (2001).
100. M. Bowen, V. Cros, F. Petroff, A. Fert, C. Martnez Boubeta, J. L. Costa-Krmer, J. V. Anguita, A. Cebollada, F. Briones, J. M. de Teresa, L. Morelln, M. R. Ibarra, F. Gell, F. Peir, and A. Cornet, *Appl. Phys. Lett.* 79, 1655 (2001).
101. W. H. Butler, X. G. Zhang, T. C. Schulthess, and J. M. MacLaren, *Phys. Rev. B* 63, 054416 (2001).
102. J. Mathon and A. Umerski, *Phys. Rev. B* 63, 220403(R) (2001).
103. T. Kanaji, T. Kagotani, and S. Nagata, *Thin Solid Films* 32, 217 (1976).
104. T. Urano and T. Kanaji, *J. Phys. Soc. Jpn.* 57, 3043 (1988).
105. Chun Li and A. J. Freeman, *Phys. Rev. B* 43, 780 (1991).
106. Lee and Wong, *J. Phys. Soc. Jpn.* 45, 895 (1978).
107. H. L. Meyerheim, R. Popescu, J. Kirschner, N. Jedrecy, M. Sauvage-Simkin, B. Heinrich, and R. Pinchaux, *Phys. Rev. Lett.* 87, 076102 (2001).
108. S. Yuasa, T. Nagahama, A. Fukushima, Y. Suzuki, and K. Ando, *Nature Mater.* 3, 868 (2004).
109. S. S. P. Parkin, C. Kaiser, A. Panchula, P. M. Rice, B. Hughes, M. Samant, and See-Hun Yang, *Nature Mater.* 3, 862 (2004).
110. J. C. Slonczewski, *J. Magn. Magn. Mater.* 159, L1 (1996); *J. Magn. Magn. Mater.* 195, L261 (1999); *J. Magn. Magn. Mater.* 247, 324 (2002).
111. L. Berger, *J. Appl. Phys.* 49, 2156 (1978); *Phys. Rev. B* 54, 9353 (1996).
112. R. Landauer, *IBM J. Res. Dev.* 32, 306 (1988).
113. X. Waintal, E. B. Myers, P. W. Brouwer, and D. C. Ralph, *Phys. Rev. B* 62, 12317 (2000).
114. A. Brataas, Y. V. Nazarov, and G. E. W. Bauer, *Phys. Rev. Lett.* 84, 2481 (2000).
115. D. M. Edwards, F. Federici, J. Mathon, and A. Umerski, *Phys. Rev. B* 71, 134501 (2005).
116. J. Z. Sun, *Phys. Rev. B* 62, 570 (2000).

117. M. Tsoi, A. G. M. Jansen, J. Bass, W. C. Chiang, M. Seck, V. Tsoi, and P. Wyder, *Phys. Rev. Lett.* 80, 4281 (1998); M. Tsoi, A. G. M. Jansen, J. Bass, W. C. Chiang, V. Tsoi, and P. Wyder, *Nature (London)* 406, 46 (2000); E. B. Myers, D. C. Ralph, J. A. Katine, R. N. Louie, and R. A. Buhrman, *Science* 285, 867 (1999); J. A. Katine, F. J. Albert, R. A. Buhrman, E. B. Meyers, and D. C. Ralph, *Phys. Rev. Lett.* 84, 3149 (2000); F. J. Albert, J. A. Katine, R. A. Buhrman, and D. C. Ralph, *Appl. Phys. Lett.* 77, 3809 (2000); J. Grollier, V. Cros, A. Hamzic, J. M. George, H. Jaffres, A. Fert, G. Faini, J. Ben Youssef, and H. Legall, *Appl. Phys. Lett.* 78, 3663 (2001).

118. J. Grollier, V. Cros, A. Hamzic, J. M. George, H. Jaffres, A. Fert, G. Faini, J. Ben Youssef, and H. Legall, *Appl. Phys. Lett.* 78, 3663 (2001); *Phys. Rev. B* 67, 174402 (2003).

119. E. C. Stoner and E. P. Wohlfarth, *Philos. Trans. R. Soc. A* 240, 599 (1948).

120. D. W. Jordan and P. Smith, "Nonlinear Ordinary Differential Equations." Clarendon Press, Oxford, 1977.

121. F. J. Albert, J. A. Katine, R. A. Buhrman, and D. C. Ralph, *Appl. Phys. Lett.* 77, 3809 (2000).

122. F. J. Albert, N. C. Emley, E. B. Myers, D. C. Ralph, and R. A. Buhrman, *Phys. Rev. Lett.* 89, 226802 (2002).

123. J. A. Katine, F. J. Albert, R. A. Buhrman, E. B. Meyers, and D. C. Ralph, *Phys. Rev. Lett.* 84, 3149 (2000).

124. S. I. Kiselev, J. C. Sankey, I. N. Krivorotov, N. C. Emley, R. J. Schoelkopf, R. A. Buhrman, and D. C. Ralph, *Nature (London)* 425, 380 (2003).

125. S. Urazhdin, N. O. Birge, W. P. Pratt, Jr., and J. Bass, *Phys. Rev. Lett.* 91, 146803 (2003).

126. M. R. Pufall, W. H. Rippard, S. Kaka, S. E. Russek, T. J. Silva, J. Katine, and M. Carey, *Phys. Rev. B* 69, 214409 (2004).

127. E. B. Myers, F. J. Albert, J. C. Sankey, E. Bonet, R. A. Buhrman, and D. C. Ralph, *Phys. Rev. Lett.* 89, 196801 (2002).

128. M. A. Zimmler, B. Özyilmaz, W. Chen, A. D. Kent, J. Z. Sun, M. J. Rooks, and R. H. Koch, *Phys. Rev. B* 70, 184438 (2004).

129. M. Tsoi, A. G. M. Jansen, J. Bass, W. C. Chiang, M. Seck, V. Tsoi, and P. Wyder, *Phys. Rev. Lett.* 80, 4281 (1998).

130. M. Tsoi, A. G. M. Jansen, J. Bass, W. C. Chiang, V. Tsoi, and P. Wyder, *Nature (London)* 406, 46 (2000).

131. J. C. Slonczewski, *J. Magn. Magn. Mater.* 195, L261 (1999).

132. B. Özyilmaz, A. D. Kent, D. Monsma, J. Z. Sun, M. J. Rooks, and R. H. Koch, *Phys. Rev. Lett.* 91, 067203 (2003).

133. P. Weinberger, A. Vernes, B. L. Gyorffy, and L. Szunyogh, *Phys. Rev. B* 70, 094401 (2004).

134. M. D. Stiles and A. Zangwill, *Phys. Rev. B* 65, 014407 (2002).

135. C. Heide, P. E. Zilberman, and R. J. Elliott, *Phys. Rev. B* 63, 064424 (2001).

136. S. Zhang, P. M. Levy, and A. Fert, *Phys. Rev. Lett.* 88, 236601 (2002).

137. C. Kittel, "Quantum Theory of Solids." Wiley, New York, 1963.

Index

A

ISBN: 1-58883-049-7

Handbook of Theoretical and Computational Nanotechnology
Edited by Michael Rieth and Wolfram Schommers
Volume 7: Pages (747–769)

B

C

D

E

F

G

H

I

J

K

L

M

N

O

P

Q

R

S

T

U

V

W

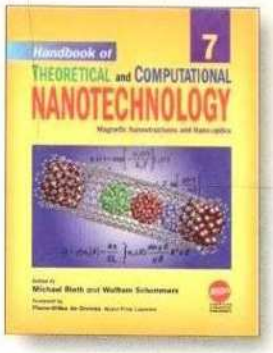
X

Y

Z

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 1-58883-049-7
90000
9 781588 830494

ASP **AMERICAN SCIENTIFIC PUBLISHERS**
Los Angeles, California, USA

Printed in the United States of America