

Introduction to Complex Mediums for Optics and Electromagnetics

Introduction to Complex Mediums for Optics and Electromagnetics

Editors: Werner S. Wiegelhofer • Akhlesh Lakhtakia

SPIE
PRESS

Bellingham, Washington USA

Library of Congress Cataloging-in-Publication Data

Introduction to complex mediums for optics and electromagnetics / [edited] by Werner S. Weiglhofer, Akhlesh Lakhtakia.

p. cm. — (SPIE Press monograph ; PM123)

Includes bibliographical references and index.

ISBN 0-8194-4947-4 (hardcover)

1. Optical materials. 2. Electromagnetism—Materials. 3. Nanostructure materials. 4. Composite materials. I. Weiglhofer, Werner S. II. Lakhtakia, A. (Akhlesh), 1957- III. Series.

QC374.I63 2003

620.1'1295—dc21

2003042373

Published by

SPIE—The International Society for Optical Engineering

P.O. Box 10

Bellingham, Washington 98227-0010 USA

Phone: (1) 360.676.3290

Fax: (1) 360.647.1445

Email: spie@spie.org

Web: www.spie.org

Copyright © 2003 The Society of Photo-Optical Instrumentation Engineers

All rights reserved. No part of this publication may be reproduced or distributed in any form or by any means without written permission of the publisher.

Printed in the United States of America.

About the cover: The advances made in the last 150 years in the field of electromagnetics are connected to the ideas seeded by the 24 people depicted. These stalwart scientists are A.M. Ampere, J.C. Bose, C.A. Coulomb, A. Einstein, B. Franklin, A.J. Fresnel (top row); K.F. Gauss, O. Heaviside, H. von Helmholtz, H. Hertz, J. Larmor, M. von Laue (second row); H.A. Lorentz, G. Marconi, J.C. Maxwell, A.A. Michelson, H.C. Oersted, C.V. Raman (third row); W.H. Bragg, W.L. Bragg, J.W. Strutt (Rayleigh), J.J. Thomson, A. Volta, L. Onsager (bottom row).

Dedicated to pristine mountains and cheerful childhoods

Contents

<i>Foreword</i>	xxi
<i>Preface</i>	xxv
<i>List of Contributors</i>	xxxix

Part I: General

Separating Field and Constitutive Equations in Electromagnetic Theory	3
<i>Evert J. Post</i>	
The beginnings	4
Georgi's rationalization	5
Georgi version of Minkowski electrodynamics	7
SR(3)'s suffocating hold on field theories	11
Mathematical specifics	14
Transformation of tensors	14
Differential forms and de Rham cohomology	18
Constitutive specifics	19
Conclusion	22
Acknowledgments	23
References	24
Constitutive Characterization of Simple and Complex Mediums	27
<i>Werner S. Weiglhofer</i>	
Introduction: the curtain rises	28
Basics: the Maxwell equations	30
Space and time	30
Space and frequency	31
Setting the stage: constitutive relations	32
Exploring the stage: simple mediums	34
The classical vacuum as reference medium	34
Homogeneous isotropic dielectric-magnetic mediums	35
A plethora of complex mediums	37
Beyond isotropy	37
Constitutive relations: generalities	41
Linear mediums: bianisotropy	43
Beyond homogeneity	46
Nonlinear mediums	47

Regulating the stage: symmetries and constraints	49
General remarks	49
Reciprocity	50
Losslessness	50
A structural constraint	50
Symmetries: biaxial bianisotropic mediums	51
Preparing the stage: homogenization	53
Faraday chiral mediums	53
Concluding remarks	55
References	55
 Isotropic Chiral Materials	 63
<i>Craig F. Bohren</i>	
Introduction	64
Polarization: the simple truth	65
Circular birefringence and circular dichroism	67
A digression on vectors	70
Electromagnetic fields in a chiral material	72
Essential reading	76
References	76
 Point Group Symmetries	 79
<i>Daniel B. Litvin</i>	
Point groups	80
Physical property tensors	82
Tensor distinction of domains in ferroic crystals	83
Global tensor distinction	85
Domain pair tensor distinction	86
Domain pair symmetry and twinning groups	88
Completely transposable twinning groups	90
Domain tensors and tensor invariants	92
Domain average engineering of ferroics	94
Conclusions	96
Appendix A: Point group symbols	96
Appendix B: Form of tensors	97
References	98
 Part II: Nonlinear Optical Materials	
 Nonlinear Optics Using Semiconductor Quantum Wells	 105
<i>John M. Arnold</i>	
Introduction	106
Theoretical nonlinear optics	108
Quantum wells	110

Second-order quasi-phase-matching	113
Third-order nonlinearity	116
Conclusions	118
Acknowledgments	118
References	118
Organic Thin-Film Photorefractive Materials	121
<i>Partha P. Banerjee</i>	
Introduction	122
Photorefractive polymers	123
Engineering photorefractive polymers	124
Nonlinear optical polymer hosts	124
Charge-transporting polymer hosts	125
Fully functionalized polymers	126
Wave mixing in photorefractive polymers	127
Real-time edge enhancement	131
Edge-enhanced correlation	133
Conclusion	136
Acknowledgments	137
References	137
Optical Energy Harvesting Materials	141
<i>David L. Andrews</i>	
Introduction	142
Precepts from photobiology	143
Resonance energy transfer	145
Dendrimers	149
Rare-earth materials for energy pooling	151
Energy pooling in multichromophore arrays	155
The future of energy pooling	157
Acknowledgments	158
References	158
Part III: Magnetic Materials	
Magnetoelectric Effects in Insulating Magnetic Materials	167
<i>Hans Schmid</i>	
Introduction	168
Thermodynamic potential	169
Linear and bilinear magnetoelectric effects	172
ME _H effects	172
ME _E effects	172
Tensor form of the linear magnetoelectric effect	172
Tensor form of the bilinear (quadratic) magnetoelectric effects	173

Measuring units	173
Examples of materials with a linear magnetoelectric effect	174
Spontaneous ferroelectric versus magnetic-field-induced polarization	176
The bilinear magnetoelectric <i>EHH</i> effect	176
The bilinear magnetoelectric <i>HEE</i> effect	177
ME _H and ME _E effect measurements	177
Optical magnetoelectric measurements	177
The piezomagnetoelectric effect <i>EHσ</i>	178
Spontaneous magnetoelectric effects and related phenomenons	178
Selected applications	181
Optical effects	182
Poling (spin reversal) of antiferromagnetic domains	183
Magnetic point group determination	183
Determination of magnetic field <i>versus</i> temperature phase diagrams	184
Determination of Néel temperatures and critical exponents	184
Magnetoelectric butterfly hysteresis loops	184
Disclosure of problems relating to symmetry property relationships	185
Magnetoelectric control of a screw spin structure	185
Determination of toroidal contributions to the magnetoelectric signal	185
Conclusions	187
Acknowledgments	188
References	188
 Magneto-optics: A Critical Review	 197
<i>Allan D. Boardman and Ming Xie</i>	
Introduction	198
Linear magneto-optics of bulk material	201
Fundamentals of planewave behavior	201
The Faraday configuration	204
The Voigt configuration	205
Envelopes in a waveguide	207
Complex planar waveguide	213
Vector solitons	216
Concluding remarks	217
Acknowledgments	219
References	219
 Static and Dynamic Magnetoelasticity	 223
<i>Graeme Dewar</i>	
Introduction	224
Magnetoelastic interaction	225
Elastic energy	225
Magnetic energy	228

Magnetoelastic energy	230
Total magnetoelastic interaction energy density	235
Static and dynamic measurements	236
Villari and ΔE effects	239
Wiedemann effect	240
Conclusion	241
References	242
 Frequency Shifts Induced by a Time-Varying Magnetoplasma Medium	245
<i>Dikshitulu K. Kalluri</i>	
Introduction	246
Frequency change due to a temporal discontinuity in the medium properties	246
Time-varying plasma medium	248
Sudden creation of an unbounded plasma medium	251
Switched plasma slab	253
Applications	254
Time-varying magnetoplasma medium	255
Basic field equations	255
Characteristic waves	256
R-wave propagation	256
Sudden creation	258
Frequency-shifting characteristics of various R waves	259
Frequency upshifting with power intensification	261
Generation of a controllable helical wiggler magnetic field	262
Conclusion	262
Acknowledgment	263
References	264
 Magnetoimpedance in Multilayered Films for Miniature Magnetic Sensors	267
<i>Larissa V. Panina and Dmitriy P. Makhnovskiy</i>	
Introduction	268
Analysis of MI in multilayer structures	269
Impedance of symmetrical three-layer film	270
Surface impedance tensor	271
Exact solution for the surface impedance tensor	272
MI in a narrow sandwich (width effect)	274
Asymmetric magnetoimpedance (AMI)	275
Dynamical AMI	276
Static AMI in a film with cross-anisotropy	277
Experimental methods	278
Film preparation and experimental results	280
MI in CoFeSiB/Cu/CoFeSiB multilayers	280
MI in NiFe/M/NiFe multilayers	281
Asymmetric MI (AMI)	282

Practical MI sensor design	286
Conclusions	288
References	289

Part IV: Composite Materials

Metamaterials: An Introduction	295

Rodger M. Walser

Introduction	296
Conventional macroscopic composites	297
Motivation for metamaterials	298
Definitions of metamaterials and metaparticles	301
Examples of metamaterials	303
Thermoelectric metamaterials	304
High-frequency magnetic metamaterials	304
Electromagnetic metamaterials	306
Conventional electromagnetic composites	306
The need for electromagnetic metamaterials	308
Metamaterial electromagnetic composites	310
Conclusions	313
Acknowledgments	314
References	314

Homogenization of Linear and Nonlinear Complex Composite Materials	317

Tom G. Mackay

Introduction	318
Preliminaries	319
Component phases	319
Depolarization and polarizability dyadics	321
Conventional approaches to homogenization	322
Maxwell Garnett formalism	322
Bruggeman formalism	322
Recent developments: incremental and differential Maxwell Garnett formalisms	323
A numerical example	324
SPFT homogenization	325
Generalities	325
Degenerate cases	326
Bilocal approximation	326
Trilocal approximation	327
Numerical results	328
Weakly nonlinear regime	330
Generalities	330
Degenerate cases	332
Bilocal approximation	333

Contents	xiii
Trilocal approximation	334
Numerical results	335
Concluding remarks	337
Appendix 1	338
Appendix 2	341
References	342
Negative Phase-Velocity Mediums	347
<i>Akhlesh Lakhtakia, Martin W. McCall and Werner S. Weiglhofer</i>	
Introduction	348
Phenomenology	350
Basic equations	350
Negative phase-velocity	351
Dispersion	352
Reflection and refraction	353
Experimental evidence	354
Ring-wire composite material	354
The crucial observation	355
Terminology	357
Research trends	357
Perfect lenses	357
Unusual composite materials	358
Planar technology	358
Complex materials	358
Concluding remarks	358
Acknowledgments	359
References	359
Scattering Theory of Photonic Crystals	365
<i>Didier Felbacq and Frédéric Zolla</i>	
Introduction	366
Scattering theory of photonic crystals	367
One-dimensional photonic crystals	367
Numerical examples	370
Defect in infinitely extended periodic medium	372
Scattering photonic crystal of finite thickness	374
Two-dimensional photonic crystals	378
Optical characterization of photonic crystals	380
Construction of the scattering matrix	380
Gamow vectors and quasi-normal modes	384
Resonant modes	385
Isolated defect	385
Photonic waveguides	386

Current problems and future directions	388
Concluding remarks	390
Acknowledgment	390
References	390

Part V: Nanostructured Materials

Optical Properties of Metal-Dielectric Films	397
<i>Andrey K. Sarychev and Vladimir M. Shalaev</i>	
Introduction	398
Generalized Ohm's law approximation and giant fluctuations of local electromagnetic fields	399
Surface plasmon polaritons	403
Resonant transmission	404
Light-induced resonant transmission	408
Extraordinary optical transmittance through nanoholes	409
Electric and magnetic resonances	411
Light circuiting in nanoholes	413
Concluding remarks	414
Acknowledgment	415
References	415
Nanostructured Thin Films	421
<i>Geoff B. Smith</i>	
Introduction	422
History and scope	422
Effective-medium models	423
Nanostructured films containing conductors: an overview	426
Thin films containing nanoparticles	429
General issues	429
Polarization eigenmodes in arrays	431
Isolated nanoparticles	432
Increasing density and clustering effects	435
Metal thin films on dielectric nanoparticles and nanostructures	438
Dense arrays, clusters touching particles	440
Conclusions	442
References	443
The Past, the Present, and the Future of Sculptured Thin Films	447
<i>Akhlesh Lakhtakia and Russell Messier</i>	
Introduction	448
From columnar to sculptured thin films	449
Columnar thin films	449
Primitive STFs with nematic morphology	455

Chiral STFs	456
Sculptured thin films	456
Electromagnetic field equations	458
Linear constitutive equations	458
Electromagnetic wave propagation	459
Structure-property relationships	460
Applications of STFs	461
Accomplishments	461
Emerging applications	465
Future research directions	467
Acknowledgments	468
References	468
 Towards Optoelectronic Applications of Chiral Sculptured Thin Films	 479
<i>Martin W. McCall</i>	
Introduction	480
Preliminaries	481
Projected Helmholtz equation	481
Bragg grating physics	483
Chiral sculptured thin films	484
Full electromagnetic analysis	486
The optical response of a CSTF to axial excitation	488
Coupled-wave techniques	491
The multireflectivity model of CSTFs	493
Applications	495
Narrow-band polarization filter	496
Multipass narrow-band filters	497
Tailored polarization-specific filters	498
Polarization routing	499
Issues for optical communications	501
Conclusion	502
References	504
 Electromagnetics of Carbon Nanotubes	 507
<i>Sergey A. Maksimenko and Gregory Ya. Slepyan</i>	
Introduction	508
Electron transport in carbon nanotubes	509
Dispersion properties of π -electrons	509
Bloch equation for π -electrons in carbon nanotubes	513
Linear electrodynamics of carbon nanotubes	515
Dynamic conductivity of carbon nanotubes	515
Effective boundary conditions for CNs	517
Surface electromagnetic waves in nanotubes	518
Edge effects in nanotubes	520

Nonlinear processes in nanotubes	524
Current density spectrum in an isolated nanotube	525
Negative differential conductivity in an isolated nanotube	528
Quantum electrodynamics of carbon nanotubes	532
The Maxwell equations for electromagnetic field operators	532
Spontaneous decay of an excited atom in the carbon nanotube	534
Conclusion	539
Acknowledgments	540
References	540

Part VI: Patterns and Statistics

Randomness in Complex Materials	549
<i>H. John Caulfield, Donald O. Henderson, and Mikhail A. Noginov</i>	
Introduction	550
Raw material for self-organization	551
Random lasing in scattering solid-state materials	552
History and the state of the art	552
Formation of a coherent mode	556
Ease of manufacturing	559
Uniformization of optical properties	563
Conclusion	564
Acknowledgments	564
References	566
Nonlinear Spatial Structures	571
<i>William J. Firth and John M. McSloy</i>	
General introduction	572
Basic models	573
Basics of pattern formation	575
Pattern formation in nonlinear optics	576
Solitonlike self-localized structures	580
Conclusions	585
Acknowledgments	585
References	585
Additional references not directly cited	588
Statistical Approaches to Scattering	591
<i>Walid Tabbara, Véronique Rannou, and Stefano Salio</i>	
Introduction	592
Elements of the statistical vocabulary	592
The statistical approach	594
Application I: Crosstalk	595

Geometrical representation of the cable	595
Electrical parameters of the cable	596
Observables and factors	598
Experiment design	598
A brief presentation of kriging	599
Transmission-line coupling	600
Results	601
Coupling to a cable	601
Coupling to a transmission line	604
Conclusion	606
Acknowledgments	606
References	606
 Elastic Orthonormal Beams and Localized Fields	 609
<i>George N. Borzdov</i>	
Introduction	610
Basic relations	612
Elastic eigenwaves in an anisotropic medium	612
Elastic eigenwaves in an isotropic medium	615
Sound eigenwaves in an ideal liquid	615
Superpositions of eigenwaves	616
Fields defined by spherical harmonics	618
Photoelasticity in an isotropic medium	620
Superpositions of longitudinal eigenwaves	621
Orthonormal beams of Type I	621
Orthonormal beams of Type II	623
Localized fields	624
Superpositions of transverse eigenwaves	627
Orthonormal beams of Type I	628
Orthonormal beams of Type II	629
Localized fields	632
Complex field structures	634
Conclusion	637
References	638
 Part VII: Measurements	
 Polarimeter for Anisotropic Optically Active Materials	 645
<i>Toru Asahi and Jinzo Kobayashi</i>	
Introduction	646
Optical activity	649
Principle of high-accuracy universal polarimeter (HAUP)	651
Original HAUP method	651
General HAUP method	656

Examples of experimental results	660
BaMnF ₄	660
Poly-L-lactic acids	661
Lysozyme crystal	663
Silver thiogallate	669
Chiral physics	669
Acknowledgment	669
References	671
 Generalized Ellipsometry	 677
<i>Mathias Schubert</i>	
Introduction	678
Experimental	679
Birefringence in stratified mediums	679
Generalized ellipsometry	680
Jones matrix presentation	681
Mueller matrix presentation	683
Light propagation in layered anisotropic mediums	684
Coherent treatment	684
Incoherent treatment	686
Generalized ellipsometry data analysis	688
A survey of birefringent material applications	690
Orthorhombic bulk minerals	690
Phonons in wurtzite-type films on sapphire substrates	692
Partially CuPt-type ordered (Al,Ga)InP ₂	695
Anisotropic refractive indexes and geometry of chiral liquid crystals	696
Sculptured thin films	699
Far-infrared magneto-optic birefringence in <i>n</i> -type GaAs	700
Conclusions	703
Acknowledgments	704
References	704
 In memoriam: Werner S. Weiglhofer	
Professor Werner S. Weiglhofer (1962–2003)	713
<i>David R. Fearn</i>	
Personal Memories of Werner S. Weiglhofer	719
<i>Tom G. Mackay</i>	
Werner S. Weiglhofer—A Personal Tribute	721
<i>Edward Spence</i>	

Contents	xix
Memories of Werner S. Weiglhofer <i>Martin W. McCall</i>	723
My Friend Werner <i>Akhlesh Lakhtakia</i>	725
Published Scientific Works of Werner S. Weiglhofer <i>Tom G. Mackay</i>	731
<i>Index</i>	749

Foreword

Richard P. McNitt

Engineers/scientists are oft exposed to the concept that the half-life of one's useful scientific knowledge is of the order of a decade, that those not *keeping up* will quickly be left behind. The contents of this book are proof, to me, that—if anything—a decade may well overstate the length of scientific half-life in the realm of electromagnetic/material interactions. I vividly recall a graduate course in physics (taught by a well-known physicist) four decades ago where we were exposed, more or less in passing, to some *second-order* effects: Peltier, Seebeck, and Thompson. As most of the students were engineers with strong interest in devices, some suggestions for utilization were quickly put forward—for instance, a possible direct-current refrigerator—but the professor noted that the effects were “small” for extant materials and thus unsatisfactory for such use. Similarly, we were given some rudimentary information about liquid crystals (possible thermometers), and birefringence of some strained materials, the latter an area of active research at that time for photoelasticians. As to electromagnetic waves, the “Maxwell equations of the time” were considered adequate, providing solutions in free space, isotropic homogenous materials, and waveguides. Most mathematical requirements were satisfied by the utilization of linear partial differential equations with constant coefficients.

This book presents many aspects of what is essentially a brave new engineering/scientist world, presenting major findings of the last decade, current research activity, speculations, and suggestions for future attack. There are powerful and general (we certainly had not been exposed to “Diffeo(4)”, or to Monte

Carlo simulations) mathematical methods presented, *generalized* Maxwell's equations are suggested and then utilized to resolve complex situations, a plethora of new *effects* are described and explained, and rather exotic materials and materials systems are presented. Instead of being confined to existing materials and materials systems, engineers are now able to work with materials scientists to design systems (composites, thin films, etc.) they need even to the nano-level. One author points out that the cause-and-effect orientation of the past is now integrated into a systems approach that has a goal-and-means orientation. Further, in the new realm of nanomaterials, quantum effects also come to the fore.

It is illuminating to list just some of the named *effects* (some of which are *third*- and *fourth-order*, but of increasing engineering significance) listed in this tome, effects that are available for exploitation by the informed and contemporary engineer: Faraday (rotation), Fresnel-Fizeau, Kerr, Matteucci, Mockels, Sagnac, Villari, Voigt (Cotton-Mouton), and Wiedemann. Topics such as natural optical rotation, electro-magnetic- and piezo-toroidics, magnetoelectric, magnetoimpedance, paramagnetoelectric, piezomagnetoelectric, whistler waves, and others are examined. Concepts such as excitons, light-assisted tunneling, photonic crystals, spatial solitons, semiconductor quantum wells, superlattices (metamaterials), and negative phase-velocity materials are considered, particularly as to how they will be effective in new materials with names such as Permalloy, Terfenol-D, and Yablonovite. Faced with such evidence of so many things that were not generally known four decades ago, it is tempting to validate a decade half-life, and to acknowledge that many engineers who thought they were well trained, had only mastered $(1/2)^4 =$ one sixteenth of the information presented in this book!

The broad and exceptionally well-explained contents should be of significant value to three very different populations:

- (i) The engineering scientist whose formal education occurred some decades ago and who wishes to be brought *up-to-date*. . . . The chapters are comprehensive, well written, and informative. Although not condescending, each chapter starts with fundamentals and completely develops the appropriate theory.
- (ii) Those currently active in the very broad arena, who wish a compact yet comprehensive overview of the field as well as of those works that would be complementary to their own. . . . This book should prove to be of real value in expanding the scope of their individual researches.
- (iii) Graduate students. . . . The authors should be commended for expressing their visions as to what remains to *be done*, what is important, and the possible modes of attack. This book should prove to be an excellent source of thesis problems as well as a *map* to achieve the desired solution.

As one who has been in all three of these groups (in reverse order), I found this book to be a treasure trove. . . . I trust you will too.

In an old Pennsylvania Dutch saying, I sign myself as one who is

Old too soon, smart too late.

Preface

“So, what is a complex medium?” Had you asked me this question in 1990, I would not have been able to give you a coherent answer. Although by then I had studied electromagnetic fields in materials with complicated response properties for about seven years, my understanding of electromagnetics lacked the necessary breadth. Furthermore, electromagnetics researchers studying diverse types of response properties were just beginning to interact with each other.

A decade later, the subdiscipline of complex-mediums electromagnetics (CME) has taken shape. At least two series of conferences on CME are held regularly, and many scientific and technical meetings have special sessions devoted to CME. Among other complex mediums, carbon nanotubes, metamaterials, materials in which light bends “differently,” and materials in which light “rotates” are commonly written about in science magazines (such as *Nature*, *Science* and *Materials Today*), as well as in monthly organs of learned societies (such as *OE Magazine*, *Optics and Photonics News* and *IEEE Antennas and Propagation Magazine*).

In 2003, I can give two answers to your question: a short answer, and a long one. The short answer is that a *positive* definition of complex mediums still remains elusive. The consensus among CME researchers is that a complex medium is not a simple medium; and that the response properties of any complex medium must be different from linear, isotropic dielectric. The long answer? Well, read on . . .

Giant strides were made during much of the 20th century in understanding and commercially exploiting the electromagnetic properties of our atmosphere and virtually matter-free space. Yet materials research for the most part remained confined to simplified (preferably dielectric) response properties. The situation began to change during the 1980s. Scientific and technological progress came to be dominated by the conceptualization, characterization, fabrication, and application of many different classes of materials. Although some of these materials are found in nature, laboratory processing is often needed for efficient use. Others are entirely synthetic, created by chemical and physical processes. Certain materials are multiphase composites designed for certain desirable response properties otherwise unavailable. Multifunctional materials as well as functional gradient materials are needed for special purposes. Nanoengineering is often used to make material samples with the same chemical composition but different response characteristics. Thus, novel fabrication techniques and a multifarious understanding of the relationship between the macroscopic properties and the microstructural morphology

of materials led to rapid progress in research on the interaction of the electromagnetic field and matter.

Electromagnetics is a science of the microscopic, though, perhaps reasonably, undergraduate textbooks rarely mention that subtlety. Many graduate textbooks also do not sufficiently emphasize that foundation. Since the 1890s, however, the Lorentz–Heaviside visualization has prevailed over earlier, even Maxwell’s, understanding of electromagnetism. All matter is an ensemble of discrete charges dispersed in free space or vacuum; but an exact treatment of that kind is virtually impossible, even today, when the charge-bearing entities exceed a few million in number. Fortunately, when electromagnetic wavelengths considerably exceed molecular dimensions, matter can be treated as a continuum for a host of technological purposes.

A simple medium—most easily exemplified by a linear, isotropic dielectric material—affects the progress of electromagnetic signals in two ways:

- a delay is created with respect to propagation in vacuum, and
- absorption of electromagnetic energy takes place.

Both effects evince dependencies on frequency, but not on spatial direction. Calculations can be made and measurements can be interpreted on the per unit amplitude/intensity basis. An isotropic dielectric medium is thus equivalent to an isotropic contraction of space with absorption overlaid.

In complex mediums, the progress of electromagnetic signals is additionally affected in one or more of several ways:

- *anisotropy*: the direction-dependent contraction of space and absorption;
- *chirality*: the twisting of space;
- *nonhomogeneity*: the dispersal of energy into different directions by either interfaces between uniform mediums or continuous gradients in material dispersal; and
- *nonlinearity*: the emission of absorbed energy at (generally) some other frequency.

In consequence, CME research has several characteristics different from research on simple mediums.

First, CME formulations are best couched in terms of the fundamental entity in modern electromagnetics: the *electromagnetic field*. It happens to have two parts, named the electric field $\mathbf{E}$ and the magnetic field $\mathbf{B}$, and identified separately for historical reasons as well as convenience. The two parts cannot be separated from the other, except after making some approximation or the other. Take a piece of a material that you think is linear, isotropic and dielectric; and make it move at a constant velocity with respect to you. You will find that it displays bianisotropic properties upon motion. A Lorentz-covariant description is therefore the only proper description of electromagnetic response properties.

Second, causality must be incorporated in CME research. Every material responds after a delay. The instantaneous part of its response properties cannot be different from that of free space; otherwise, the material would possess foreknowledge, a prospect best left for sci-fi authors to exploit. The development of femtosecond-pulse optics and the generation of attosecond pulses suggest that it is better not to cast time aside by the artifice of the Fourier transform. Even in the frequency domain, causality takes the form of dissipation and dispersion, which are the two sides of the same coin.

Third, although matter is nonhomogeneous at microscopic length scales, piecewise homogeneity is commonplace at macroscopic length scales. Statistical techniques provide a bridge between the two length scales. Complicated macroscopic response properties should not be assumed casually. For instance, if a homogeneous piece of a medium with a certain set of response properties cannot be found, the existence of continuously nonhomogeneous analogs of that set at macroscopic length scales is a dubious proposition. The development of homogenization techniques for complex mediums is a major challenge today, despite very recent successes for linear bianisotropic materials.

Fourth, nonlinearity is an essential attribute of wave-material interaction. Nonlinearity introduces dependency on amplitude or strength, and is responsible for the occurrence of multiwavelength processes. It also accounts for the electromagnetic exposure histories of materials. We all know from high-school textbooks that matter modifies electromagnetic waves; but waves also modify matter. Observe how a newspaper yellows after lying in the sun for a few days. Electromagnetic waves emitted by the sun (i.e., sunlight) effect that change.

The complexity of actual materials cannot yet be handled in its entirety. Complexity is like Gulliver, while CME researchers are like the Lilliputians. Although an individual CME researcher takes only one or two meaningful steps towards the taming of complexity, different steps are taken by different CME researchers. CME commands the attentions of scientists from a wide spectrum of disciplines: from physics and optics to electrical and electronic engineering, from chemistry to materials science, to applied mathematics and even biophysics. Thus, CME is presently a multidisciplinary research area spanning basic theoretical and experimental research at universities to the industrial production of a diverse array of electrical, microwave, infrared and optical materials and devices. A recent impetus for multidisciplinary is the unrelenting progress of nanotechnology, which is now beginning to engender mesoscopic approaches in CME.

This book is a collection of essays to explain complex mediums for optical and electromagnetic applications. The genesis of this book lies in a series of conferences organized at the successive Annual Meetings of SPIE from 1999 to 2002. The scope of Conference 3790, *Engineered Nanostructural Thin Films and Materials*, was not fully explained by its title. Subsequently, Conference 4097 was entitled *Complex Mediums*. Further explication being needed, Conference 4467 was named *Complex Mediums II: Beyond Linear Isotropic Dielectrics* and was followed by Conference 4806 *Complex Mediums III: Beyond Linear Isotropic Dielectrics*. All

four were organized by me, very ably assisted by Werner S. Weiglhofer, Russell F. Messier, Ian J. Hodgkinson, Martin W. McCall, and Graeme Dewar. A multitude of CME researchers participated wholeheartedly.

Werner S. Weiglhofer, my co-editor, was involved in all four conferences. He and I felt that the optics community at large should benefit from a relatively broad introduction to complex mediums. Many speakers who had delivered *Key Lectures* and *Critical Review Lectures* at the conferences agreed, as also did Rick Hermann and Sharon Streams of SPIE Press. We therefore invited the presenters to update and expand their initial lectures. Other prominent researchers were invited to contribute essays on CME topics that were deemed important but had not been covered in the four conferences. The essays were edited, reviewed, revised and compiled into this book.

All contributors were requested to write with two aims: first, to educate on phenomenology and terminology; second, to provide a state-of-the-art review of a particular topic. The vast scope of CME exemplified by the actual materials covered in the essays should provide a plethora of opportunities to the novice and the initiated alike. Graduate students in the broad disciplines of electrical engineering, materials science, and physics are likely to find inspiration from one essay or another to pursue CME research; and our fondest hope is that this book would serve the next decade or so as a goldmine for dissertation topics. Experienced researchers desirous of either switching research areas or synthesizing new types of material responses may profit from this book as well. R&D engineers in industry may be able to conceptualize and actualize new types of devices, after reading certain parts of this book.

I must add here that, although Werner and I had agreed to divide editorial responsibilities equally, he was the Managing Editor. To this position, he brought his considerable organizational acumen. He interacted with all contributors and reviewers, as well as with Sharon Streams at SPIE. All contributors were supplied progress reports at suitable intervals; e-mails were promptly answered by him with unfailing courtesy; and so on. When on January 5, 2003, he asked me to initiate the writing of a preface, I replied that the end of March was far away. A week later, he was killed by an avalanche on the slopes of Bispen, a Norwegian mountain that he had ascended 29 times. I had to assume his mantle; I had to write this preface solo. This book is now a memorial to my friend Werner S. Weiglhofer, as you will notice from the inclusion of a section entitled *In Memoriam*.

A linguistic note: You will notice the absence in this book of Latin and Greek plurals of words from those languages commonly used in English. This was a deliberate editorial decision. During some 14 years of collaboration, both Werner and I were appalled at the widespread misuse of plurals—such as *criteria*, *media* and *spectra*—as singulars in scientific literature. Such pluralization is artificial to the native robustness of English. At best, it is an affectation. No wonder so many native and non-native speakers of this language make those mistakes! Taking a leaf from George Bernard Shaw's introductions to his plays that English spelling needs reform, in 1996 we decided in favor of the normal English pluralization of Latin

and Greek singulars. Although uncommon, this practice is not new. Most journals accept it. So do the Royal Society (of London), John Wiley & Sons, and SPIE.

The cooperation that Werner and I received from all contributors and reviewers was nothing short of splendid. Ms. Sharon Streams and others at SPIE have provided unstinted support. Professors David R. Fearn and Edward Spence graciously contributed their memories of Werner; and they also assisted in the transfer of editorial correspondence from Werner's computer to me. In the latter task, they were joined by Mr. David Thom (University of Glasgow) and Professor Joseph P. Cusumano (Pennsylvania State University). I am grateful to everyone involved in this project.

Complex Mediums IV: Beyond Linear Isotropic Dielectrics was convened in early August 2003, by Graeme Dewar and Martin W. McCall. I shall be delighted if a companion volume were published after another two or three editions of this conference. So would Werner, I am sure.

Akhlesh Lakhtakia

August 2003

List of Contributors

David L. Andrews

School of Chemical Sciences
University of East Anglia
Norwich NR4 7TJ United Kingdom

John M. Arnold

Department of Electronics and
Electrical Engineering
University of Glasgow
Glasgow G12 8LT United Kingdom

Toru Asahi

Research Institute for Science
and Engineering
Waseda University
3-4-1 Okubo, Shinjuku-ku
Tokyo 169-8555 Japan

Partha P. Banerjee

Department of Electrical
and Computer Engineering
University of Dayton
Dayton, Ohio 45469-0226 USA

Allan D. Boardman

Joule Physics Laboratory
Institute of Materials Research
University of Salford
Salford M5 4WT United Kingdom

Craig F. Bohren

Department of Meteorology (ret.)
The Pennsylvania State University
University Park, PA 16802 USA

George N. Borzdov

Department of Theoretical Physics
Belarus State University
Fr. Skaryny avenue 4
Minsk, 220050 Belarus

H. John Caulfield

Center for Photonic Materials
and Devices
Fisk University
1000 17th Avenue N
Nashville, TN 37208 USA

Graeme Dewar

Department of Physics
University of North Dakota
Grand Forks, ND 58202 USA

Didier Felbacq

GES UMR-CNRS 5650
Université Montpellier II
CC074 Place Eugène Bataillon
Montpellier Cedex 05
F-34095 France

William J. Firth

Department of Physics
University of Strathclyde
John Anderson Building
107 Rottenrow
Glasgow G4 0NG United Kingdom

Don O. Henderson

Center for Photonic Materials
and Devices
Fisk University
1000 17th Avenue N
Nashville, TN 37208 USA

Dikshitulu K. Kalluri

Electrical and Computer
Engineering Department
University of Massachusetts/Lowell
1 University Avenue
Lowell, MA 01854 USA

Jinzo Kobayashi

Research Institute for Science
and Engineering
Waseda University
3-4-1 Okubo, Shinjuku-ku
Tokyo 169-8555 Japan

Akhlesh Lakhtakia

CATMAS, Department of Engineering
Science and Mechanics
The Pennsylvania State University
University Park, PA 16802 USA

Daniel B. Litvin

Department of Physics
The Eberly College of Science
The Pennsylvania State University
P.O. Box 7009
Reading, PA 19610-6009 USA

Tom G. Mackay

Department of Mathematics and
Statistics
University of Edinburgh
The King's Buildings
Edinburgh EH9 3JZ United Kingdom

Dmitriy P. Makhnovsky

Department of Communication and
Electrical Engineering
University of Plymouth
Drake Circus
Plymouth PL4 8AA United Kingdom

Sergey A. Maksimenko

Institute for Nuclear Problems
Belarus State University
Bobruiskaya 11
Minsk, 220050 Belarus

Martin W. McCall

Department of Physics
The Blackett Laboratory
Imperial College London
Prince Consort Road
London SW7 2BW United Kingdom

Richard P. McNitt

Department Head & Professor Emeritus
Department of Engineering Science
and Mechanics
The Pennsylvania State University
University Park, PA 16802-6812 USA

John M. McSloy

Department of Physics
University of Strathclyde
John Anderson Building
107 Rottenrow
Glasgow G4 0NG United Kingdom

Russell Messier

CATMAS, Department of Engineering
Science and Mechanics
The Pennsylvania State University
University Park, PA 16802 USA

Mikhail A. Noginov

Center for Materials Research
Norfolk State University
700 Park Avenue
Norfolk, VA 23504 USA

Larissa Panina

Department of Communication and
Electrical Engineering
University of Plymouth
Drake Circus
Plymouth PL4 8AA United Kingdom

Evert J. Post

7933 Breen Avenue
Westchester, CA 90045-3357 USA

Veronique Rannou

Département de Recherche en
Électromagnétisme/L.2 S. Supélec
Plateau de Moulon
Gif sur Yvette Cedex
F-91192 France

Stefano Salio

Politecnico di Torino
Electronics Department
Torino, Italy

Andrey K. Sarychev

School of Electrical and
Computer Engineering
Purdue University
West Lafayette, IN 47907 USA

Mathias Schubert

Institute for Experimental Physics II
University of Leipzig
Linnéstrasse 5
Leipzig D-04103 Germany

Vladimir M. Shalaev

School of Electrical and
Computer Engineering
Purdue University
West Lafayette, IN 47907 USA

Gregory Ya. Slepyan

Institute for Nuclear Problems
Belarus State University
Bobruiskaya 11
Minsk, 220050 Belarus

Hans Schmid

Department of Inorganic Analytical
and Applied Chemistry (ret.)
University of Geneva

30 quai Ernest-Ansermet
Geneva 4, CH-1211 Switzerland

Geoffrey B. Smith

Department of Applied Physics
University of Technology Sydney
PO Box 123
Broadway NSW 2007 Australia

Walid Tabbara

Département de Recherche en
Électromagnétisme/L.2 S. Supélec
Plateau de Moulon
Gif sur Yvette Cedex 91192 France

Rodger M. Walser

Department of Electrical and
Computer Engineering
Engineering Science Building 143
University of Texas at Austin
Austin, TX 78712-1084 USA

Werner S. Weiglhofer

Department of Mathematics
University of Glasgow
Glasgow G12 8QW United Kingdom

Ming Xie

Joule Physics Laboratory
Institute of Materials Research
University of Salford
Salford M5 4WT United Kingdom

Frédéric Zolla

Institut Fresnel UMR-CNRS 6133
Faculté des Science de Saint-Jérôme
Case 161-162
Av. Escadrille Normandie Nieman
Marseille Cedex 20 F-13397 France

Separating Field and Constitutive Equations in Electromagnetic Theory

Evert J. Post

Abstract

This essay is an outline of techniques to separate out constitutive functions in order to isolate the purest field law statement that is independent of constitutive specifics. A reorganization of this kind is a near-necessity when dealing with complex electromagnetic mediums. If so desired, it creates the possibility of simultaneously and coherently treating anisotropy, the Fresnel–Fizeau effect, the Sagnac effect, natural optical rotation and Faraday rotation. Aside from these applied aspects, the efforts required to proceed in orderly fashion open up worthwhile perspectives and insights from a purely theoretical angle. The tool of general covariant description is found to have a perceptive potential that goes way beyond the $SR(3)$ group.

1 The beginnings

From the very beginning, Maxwell's inception of the concepts of macroscopic electromagnetic theory was in a form that left little to be desired; so what could be the concern of an essay on the stages of its evolution? The answer to this somewhat rhetorical question can thus be expected to indicate that the conceptual basis of Maxwell's formalism remains largely intact. Yet, special attention is to be devoted to what might be called questions on how we mathematically express the findings of Maxwell. In short, we face matters of mathematical engineering.

With the field equations in conceptually good shape, the task of tackling complicated mediums reminds us of the fact that the properties of a medium are determined by a completely separate set of equations that delineate the physical behavior of that medium. However, more than before in the case of simple mediums, attention needs to focus again on what exactly is the structure of the field equations that can be taken to apply in general. The final objective is then to combine that form with constitutive information solely specifying a particular medium.

As an early example of mixing field and constitutive functions, consider electromagnetic theory relating to wave propagation in matter-free space (i.e., vacuum). The equations then assume the form

$$\operatorname{div} \mathbf{H} = 0, \quad \operatorname{curl} \mathbf{E} = -\frac{1}{c} \frac{\partial \mathbf{H}}{\partial t}, \quad (1)$$

and

$$\operatorname{div} \mathbf{E} = 0, \quad \operatorname{curl} \mathbf{H} = \frac{1}{c} \frac{\partial \mathbf{E}}{\partial t}, \quad (2)$$

where $\mathbf{E}$ and $\mathbf{H}$ are electric and magnetic fields, respectively, whereas the speed of light c is clearly a constitutive parameter. The elimination of either $\mathbf{E}$ or $\mathbf{H}$ reveals both fields satisfying the d'Alembertian wave equation. This result brought optics into the realm of electromagnetic theory and was a dramatic revelation of 19th-century physics [1].

An inspection of this early form of the Maxwell field equations for vacuum does not give much of an inkling about the existence of a set of constitutive equations for the vacuum. The vacuum properties are here represented by the free-space light velocity c , which during the 19th century had been identified as the ratio of electric- and magnetic-based units. In this mixed unit system, $\mathbf{E}$ and $\mathbf{H}$ were taken to have the same physical dimension.

It follows from this treatment of the vacuum situation that Maxwell treated empty space as some sort of nonmedium. The constitutive equations of vacuum were, so to say, built into the field equations. For a material medium exhibiting electric and magnetic polarizabilities, 19th-century physics introduced two new field quantities known as electric displacement $\mathbf{D}$ and magnetic induction $\mathbf{B}$. In a

simple isotropic medium, $\mathbf{B}$ and $\mathbf{D}$ were related to $\mathbf{E}$ and $\mathbf{H}$ through the isotropic constitutive equations

$$\mathbf{D} = \epsilon \mathbf{E}, \quad \mathbf{B} = \mu \mathbf{H}, \quad (3)$$

in which ϵ and μ are merely numbers to describe medium properties; hence, $\mathbf{E}$, $\mathbf{D}$, $\mathbf{H}$ and $\mathbf{B}$ all have the same physical dimension. The field equations then take the following form:

$$\operatorname{div} \mathbf{B} = 0, \quad \operatorname{curl} \mathbf{E} = -\frac{1}{c} \frac{\partial \mathbf{B}}{\partial t}, \quad (4)$$

$$\operatorname{div} \mathbf{D} = 0, \quad \operatorname{curl} \mathbf{H} = \frac{1}{c} \frac{\partial \mathbf{D}}{\partial t}. \quad (5)$$

Equations (3)–(5) led to another landmark conclusion that the speed of propagation u in a material medium would be smaller than c according to $u^2 = c^2/\epsilon\mu$.

For material mediums, (3)–(5) give an effective separation between field equations and constitutive equations. However, the presence of c in the field equations (4) and (5) testifies to the fact that matter-free space still remains in an exceptional position. It was engineering rather than the aesthetics of mathematical physics that would lead to the removal of this remaining defect in the beauty of the mathematical representation of electromagnetic theory.

2 Giorgi's rationalization

The mixed-unit system of physics became highly impractical as electrification began to spread around the world. Around 1900, Giorgi proposed to adapt the electromagnetic units of Maxwell field theory to the practical units used in engineering. To that end, he introduced a set of constitutive equations for vacuum as follows:

$$\mathbf{D} = \epsilon_0 \mathbf{E}, \quad \mathbf{B} = \mu_0 \mathbf{H}. \quad (6)$$

As $c^2 = 1/\epsilon_0\mu_0$, the field equations (4) and (5) accordingly were transformed into

$$\operatorname{div} \mathbf{B} = 0, \quad \operatorname{curl} \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}, \quad (7)$$

and

$$\operatorname{div} \mathbf{D} = 0, \quad \operatorname{curl} \mathbf{H} = \frac{\partial \mathbf{D}}{\partial t}. \quad (8)$$

The physics establishment fought the Giorgi system for a long time, their main argument being that ϵ_0 and μ_0 were mere numerical artifacts undeserving of official status in their serene domain of physics. It would take a quarter century before

the Giorgi system received a modicum of international recognition. It is noted parenthetically that there are recent reports that $\epsilon_0\mu_0 = 1/c^2$ may not be a universal constant [2]. The ratio of μ_0 and ϵ_0 acquired significance as a measure of free-space impedance, which is now increasingly regarded as a universal constant. So, the establishment was only half right with their numerical argument, which means they had no scientific leg to stand on and really deserved a failing mark for aesthetics of scientific description. The upshot is that the constitutive equations of simple isotropic material mediums now assume the form

$$\mathbf{D} = \epsilon_0\epsilon\mathbf{E}, \quad \mathbf{B} = \mu_0\mu\mathbf{H}; \quad (9)$$

yet the field equations are still given as (7) and (8).

After the rationalization of units and the ensuing separation of field and constitutive equations by Giorgi, the first rumblings of relativity could be felt. Establishment physics was at that time still committed to its mixed unit system. So, when Minkowski boldly exposed the world of physics to the challenge of envisioning space and time as parameters of a manifold of four dimensions [3], c was still present in the Maxwell field equations. This was a somewhat disturbing element interfering with all that beautiful mathematical symmetry. While the Giorgi rationalization was already around to remove that defect, it would take a few more decades for full international recognition.

The Giorgi rationalization was either not known in mathematical physics circles or those who knew about it were unwilling to accept a proposition that, to them, seemed motivated by mere engineering considerations. The aftermath of this ill-advised professional chauvinism is still with us today. The reader may open many a book on quantum field theory and will find the convenient substitution $c = 1$ and sometimes, even worse, an imaginary time $x_0 = ict$. Some people just felt that spacetime should have a positive definite metric. These measures were motivated by a false aesthetics. With hindsight, they merely contributed to the drift of theory away from physical reality.

Additionally, the Giorgi rationalization was not popular among the creators of the special and general theories of relativity either. Minkowski and even Einstein were somewhat guilty of condoning the aforementioned convenient substitution for c [3]. It was forced on them by an old system of mixed units. Yet by the same token, as initiators of a new way of thinking, they might have expected others to help smoothen out conflicts ensuing from unsuitable earlier traditions.

Unfortunately, these *ad hoc* adaptations were made to gain a phony idealized representation of a new discipline. They remain a serious factor that has made many physicists uncomfortable with the formalisms of relativity. Without really discounting the whole structure and major consequences thereof, it gives rise to a feeling that only the initiated know what to do and when.

Here we are confronted with organizing an overview of the behavior of complex mediums. By necessity, this includes the discussion of moving mediums and accelerated frames of reference. Such a program cannot be done well without taking full

advantage of the spacetime methods of description as opened up by Minkowski. However, after taking note of the difficulties created by *ad hoc* procedures invited by the mixed system of physical units, we proceed to implement a clean separation between field equations and constitutive equations such as achieved by the Georgi rationalization.

3 Georgi version of Minkowski electrodynamics

Let us now proceed to recast (7) and (8) in the Minkowski spacetime form. Since medium specifics require detail in terms of tensor components, the field equations likewise call for a tensorial form. The first two Maxwell equations (7) then combine into the following single equation, which contains an explicit version of a generalized curl or exterior derivative:

$$\partial_{[\lambda} F_{\nu\kappa]} = 0. \quad (10)$$

The subscripted indexes equal 0, 1, 2, and 3—with 0 as the time label and 1, 2, and 3 as space labels. Equations (8) are also combined into a single equation

$$\partial_\nu \mathfrak{G}^{\lambda\nu} = \mathfrak{C}^\lambda, \quad (11)$$

which involves a generalized divergence but can be equivalently expressed using the exterior derivative as follows:

$$\partial_{[\lambda} \tilde{G}_{\nu\kappa]} = \tilde{C}_{\lambda\nu\kappa}. \quad (12)$$

The index brackets $[\]$ indicate summation over even and odd permutations of the enclosed subscripted/superscripted indexes: even permutations get a plus sign, odd permutations get a negative sign. The components of $\mathfrak{G}^{\lambda\nu}$ and $\tilde{G}_{\nu\kappa}$ are related through the antisymmetric unit tensors; and so are $\mathfrak{C}^\nu$ and $\tilde{C}_{\lambda\nu\kappa}$. The Gothic symbols and tilde markings refer to transformation specifics that are essential to appreciate later the natural metric-free general invariant nature of the Minkowski rendition. More details are provided in Section 4.

At this point some readers may well be discouraged by what seems an excess of unexplained notational engineering. Yet, the truth is that dealing with complex mediums needs disciplined organization in order not be swamped by disorder later on.

Therefore, in the spirit of persistence we now proceed to componentwise identify F and $\mathfrak{G}$ with the usual electromagnetic fields according to the following matrix equivalences:

$$\begin{pmatrix} F_{00} & F_{01} & F_{02} & F_{03} \\ F_{10} & F_{11} & F_{12} & F_{13} \\ F_{20} & F_{21} & F_{22} & F_{23} \\ F_{30} & F_{31} & F_{32} & F_{33} \end{pmatrix} = \begin{pmatrix} 0 & -E_1 & -E_2 & -E_3 \\ E_1 & 0 & -B_3 & B_2 \\ E_2 & B_3 & 0 & -B_1 \\ E_3 & -B_2 & B_1 & 0 \end{pmatrix}, \quad (13)$$

and

$$\begin{pmatrix} \mathfrak{G}^{00} & \mathfrak{G}^{01} & \mathfrak{G}^{02} & \mathfrak{G}^{03} \\ \mathfrak{G}^{10} & \mathfrak{G}^{11} & \mathfrak{G}^{12} & \mathfrak{G}^{13} \\ \mathfrak{G}^{20} & \mathfrak{G}^{21} & \mathfrak{G}^{22} & \mathfrak{G}^{23} \\ \mathfrak{G}^{30} & \mathfrak{G}^{31} & \mathfrak{G}^{32} & \mathfrak{G}^{33} \end{pmatrix} = \begin{pmatrix} 0 & D_1 & D_2 & D_3 \\ -D_1 & 0 & H_3 & -H_2 \\ -D_2 & -H_3 & 0 & H_1 \\ -D_3 & H_2 & -H_1 & 0 \end{pmatrix}. \quad (14)$$

For example, $\partial_{[0}F_{12]} = \partial_0 F_{12} + \partial_2 F_{01} + \partial_1 F_{20} = \partial B_3/\partial t + \partial E_1/\partial y - \partial E_2/\partial x = 0$ is the z component of $\text{curl } \mathbf{E} = -\partial \mathbf{B}/\partial t$, while, similarly, $\partial_{[1}F_{23]} = 0$ is the same as $\text{div } \mathbf{B} = 0$. With patience and persistence, one can make similar identifications for $\tilde{G}_{\nu\kappa}$ as well as $\mathfrak{E}^\nu$ and $\tilde{C}_{\lambda\nu\kappa}$.

These chores are not higher mathematics; instead, they are part of learning what may be seen as a revised language. While learning a new language, it is not wise to start with a *patois* that lacks the sophistication of a well-developed medium of communication. In electromagnetic theory, too many have settled for a *patois*. In the following we adhere, as much as possible, to notational stipulations adopted by Schouten [4] to give due attention to cited distinctions in transformation behavior.

Many readers may have seen these transcriptions and even have worked through them, and then could have asked: “What good is all of this?” Electromagnetic theory has been cast in so many different renditions that one cannot help wondering what is their benefit compared with the usual rendition in terms of vectors [5]? To compare the respective virtues, it is perhaps first necessary to be thoroughly aware of the far-reaching restrictions to which the system of vector analysis is subject.

The system of vector analysis identifies polar and axial vectors, which means that it is restricted to three spatial dimensions without reflections. Its invariance group is $\text{SR}(3)$, which was adequate for early applications in mediums with rotational and inversion symmetries. Yet difficulties arise already with mediums that do not have a center of symmetry: say, solutions of optically active sugars [6-7].

An isotropic, optically active medium supports left- and right-circularly polarized plane waves with distinct propagation velocities [6]-[8]. The latter relate under reflection, yet reflections have no place in vector analysis. Physically, natural optical activity is due to a (dispersive) cross-coupling between magnetic and electric fields, which is a coupling between polar and axial vectors that happen to be identified in vector analysis. The early constitutive relations (3), (6) and (9) do not provide for such an option—not even for complex values of ϵ and μ .

An inspection of the great textbooks of the past (e.g., the 1934 edition of Max Born’s *Optik* [9]) shows that few correctly state the absence of a center of symmetry as a key point in the description of optical activity [10]. Hence, in the past, the alternative has been frequently one of taking recourse to *ad hoc* phenomenological descriptions that could have no counterpart in the microphysical developments of solid state physics.

If an electric-magnetic cross-coupling already hints at a constitutive spacetime description, an electromagnetic description of the Fresnel–Fizeau and the Sagnac

effects [8], [12] further exacerbates the limitations of vector renditions of electromagnetic theory. To avoid a recurring need for *ad hoc* procedures, a full-fledged spacetime description of field and constitutive equations truly becomes a *sine qua non*, no matter what type of objections one may have against the theory of relativity.

Confronted with the specifics of any of the foregoing phenomenon, a lonely researcher well trained in the vector renditions of electromagnetic theory can hardly be expected to be strongly motivated to recast that machinery. The Minkowski renditions available to him suffer a mix of field and constitutive parameters in the field equations, as is clear from the presence of c in (1) and (2). Those mixed units of the past made it impossible to accomplish a clean separation between field equations and constitutive equations. In the end, the researcher may throw up his/her hands in despair and decide not to get involved any further with spacetime projects. From that moment onwards, he/she may prejudice others against such formalisms.

Perhaps, a last-resort effort is still justified in overcoming such negativism. Let it be known that, taking advantage of the Giorgi rationalization of physical units, one finds that the Minkowski rendition does lead to a clean separation between field and constitutive equations. So, let (10)–(14) be testimony to this functional separation between these field equations and the constitutive equations that are to come.

An inspection of the modern textbook literature may show that today the Giorgi rationalization of units has been accepted, yet it does not mean there is a consistent habit of keeping constitutive parameters out of field equations. The widely celebrated *Feynman Lectures* [13] repeatedly state the Maxwell equations in forms wherein the parameters ϵ_0 and μ_0 explicitly occur. This testifies to a lingering belief that ϵ_0 and μ_0 are mere numerical factors, which is a faulty conviction of the past when the establishment was still fighting the Giorgi rationalization.

It is reiterated that ϵ_0 and μ_0 are components of the free-space constitutive tensor. They do not change under the SR(3) invariance group of vector analysis, but they do transform under the much wider group of transformations permitted by the Minkowski rendition—and that is the group needed to account for spacetime motions. In other words, modern textbooks still dwell, at least in this respect, on the remnants of the era of fighting Giorgi rationalization. Habits, once acquired, pass on from generation to generation.

All of this proves there still are no firm convictions universally held as to what are field equations and what are constitutive equations. A principle due to Neumann¹ [11] requires constitutive equations to be numerically invariant under the symmetry group of the medium under consideration. Hence, in order not to prejudicially affect the symmetry of mediums, one likes the field equations to remain invariant in form under at least all the invariance groups of all mediums to which the field equations are to be applied.

¹Karl Gottfried Neumann (1832–1925), not to be confused with John von Neumann (of quantum mechanics fame and a computer pioneer in the 20th century), was a German mathematician whose name is immortalized in Neumann functions and Neumann boundary conditions. References to the Neumann principle are almost completely absent from current textbooks.

The Diffeo(4)-invariant Minkowski rendition of the Maxwell equations [10]–[14] meets that requirement well. The vector rendition of Maxwell equations meets that requirement only for $SR(3)$ and its subgroups, which incidentally excludes isotropic, optically active mediums [6–7]. Since reflections and inversions are not contained in $SR(3)$, we see here how unnecessary restrictive conditions of the past have been inviting *ad hoc* methodology.

The question now is: Where do we go from here? In light of the unusual degree of resistance encountered in implementing revisions of methodology, what are the chances of bringing the Minkowski rendition of electromagnetic theory to the fruition it deserves? Let me mention some of its principal features:

1. The Giorgi version of the Minkowski rendition of electromagnetic theory meets the criterion of separation between field equations and constitutive equations.
2. A further detailed mathematical specification shows that the field equations preserve their form under general differential spacetime transformations totally independent of the metric. This remarkable Diffeo(4) property identified by Kottler [15], Cartan [16] and van Dantzig [17] implies that these macroscopically established equations retain validity in the microscopic domain.
3. The fields thus defined in electromagnetic theory are integrands of either scalar- or pseudoscalar-valued cyclic² integrals that exhibit direct physical relevance, macroscopically as well as microscopically.
4. Since electromagnetic theory first involved the use of topological concepts (e.g., enclosing and linking) as essential features, its mathematical methods were later developed into a general procedure to access the topology of field configurations: a mathematical theory known as *de Rham cohomology* [19]. Topological properties remain invariant under *general differentiable deformations*, and so the term diffeomorphism Diffeo(4) has gained its natural place in electromagnetic theory.
5. The linear constitutive equation of matter-free space is a fourth-rank tensor (i.e., four sub/superscripted indexes are attached to it), which is a concomitant of the metric tensor invoking the free-space impedance. It has the same index symmetries as the Riemann–Christoffel tensor [14].

These are the main highlights of what may be called an *extended Giorgi version of Minkowski electrodynamics*. It recognizes the metric-free Diffeo(4) invariance in addition to the perspectives of de Rham cohomology, which are very fundamental for physical content, especially quantization. Thus, it shows a ramified and very constructive interaction with various domains of physics and mathematics. However, this wider scope manifests a greatly reduced tolerance for a good deal of somewhat wild and *ad hoc* experimentation in description that has been going on under the $SR(3)$ umbrella of vector analysis.

²Cycles are integration domains of zero boundary: a 1-cycle is the topological equivalent of a circle (i.e., a closed loop), a 2-cycle is the topological equivalent of a spherical surface, a 3-cycle could relate to a closed physical universe, and so on, [18].

We are now in a position to bring to the surface a measure of awareness for some of the worst mathematical procedures that have become part of physical tradition during the century-long rule of the system of vector analysis.

4 SR(3)'s suffocating hold on field theories

Vector analysis was developed as a kind of shorthand for physics and engineering purposes, and it has been extremely helpful during the era of the electrical revolution. Föppl in Germany [20] and Gibbs [21] in the USA were among the pioneers who served the compelling needs of mathematical communication in science and engineering. Vector analysis proved to be a valuable compromise for its time. It was never developed with the objective of creating a discipline to serve an intrinsic purpose in mathematics. It was just a shorthand, contingent on the applicability of SR(3).

Crucial in the distinction between (7) and (8) on the one hand, and (10) and (12) on the other, is the tremendous widening of the group of invariance from SR(3) to Diffeo(4)—the latter is the collective of all invertible differentiable spacetime substitutions. Hence, in the Minkowski rendition, the wave equation is of the same *form* for any linear constitutive relation and any coordinate system. All that is needed is the specification of the constitutive relations and the coordinate system; and one is ready thereafter to work on solutions.

Now compare this to the SR(3)-invariant rendition of (7) and (8). The elements of this group are rotations of Cartesian frames; they are linear orthogonal transformations. Transitions to cylindrical, polar or elliptical coordinate systems are not elements of SR(3), yet they are elements of Diffeo(4). Hence, working with (7) and (8), one always needs a compendium of formulas for gradient, curl and divergence for non-Cartesian coordinate systems. If the compendium is unavailable, one has to go through tedious derivations. But when working with the likes of (10) and (12), those derivations are not necessary, because they are Diffeo(4)-invariant. So now we are finding out what was sacrificed in using the compromise known as vector analysis.

The aftermath of the compromise is even worse than it first looks. Working with (7) and (8) under the constraints of SR(3), one finds that all physical vector and tensor fields are defined within the confines of SR(3). Hence, stepping outside the confines of SR(3), one rarely, or rather never, redefines vectors and tensors for this wider realm. The establishment has regarded such a redefinition as too unsettling for reasons that are now buried deep in the literature.

So, what is the alternative of not engaging in the challenge of Diffeo(3) and Diffeo(4) redefinitions of physical fields? The answer is to take recourse to a method that everywhere forces us to keep our little Cartesian boxes around to measure Cartesian-based fields; without them, it is simply impossible to interpret those fields.

In practice, this means local Cartesian frames are chosen, aligned as well as possible with the in-principle arbitrary coordinate directions created by changes of

reference implied by application of Diffeo groups. In full generality, this procedure was developed more than a century ago by the Italian school founded by Ricci-Curbastro [22],³ and is known as the technique of anholonomic references [23]. The motivation for developing such a method may now be hidden in the past, yet insistence in physics circles on descriptions in terms of Cartesian fields may have been a principal factor. Whatever the motivation, after a century the method is alive today under the heading of *curvilinear coordinates* in every textbook [24-25].

It is interesting to note that Lagrange had been able to overcome the Cartesian syndrome in the discipline of mechanics. His generalized coordinates led to a generalized momentum as a covariant vector. Establishment physics has either been resisting a comparable generalization of fields in the realm of electrodynamics, or has denied that anything of that nature actually might exist—all this despite the pioneering work of Kottler, Cartan and van Dantzig in the early 1920s.

Physics is counted as one of the exact sciences; its methodology though does not always testify to that heritage. For the better part of a century—from Sommerfeld, Heisenberg and Schrödinger to the present day—the Hamilton–Jacobi procedure has been hailed as the cradle from which modern quantum mechanics arose. In the process of considering the next step of a quantum electrodynamics, the question presented itself: what would be a natural electrodynamic analog or counterpart of Hamilton–Jacobi mechanics? Since no believable candidate presented itself, the ensuing quantum electrodynamics came to be a discipline founded on Cartesian field concepts [26-27].

However, in witness of the Hamilton–Jacobi procedure being a contingency of the Lagrangian method in mechanics, the question for a Hamilton–Jacobi counterpart ought to be rephrased by asking for an electrodynamic counterpart of Lagrangian mechanics. The latter query had really been well answered by Kottler, Cartan and van Dantzig; yet quantum electrodynamics—as it developed in the 1950s—was unable to incorporate these insights, because the groundwork had already been laid in the 1930s on a Cartesian foundation.

All these Cartesian predicaments are avoidable by using the Diffeo(4)-invariant equations (10) and (12), which means the Cartesian vector and tensor concepts of SR(3) are extended to the realm of Diffeo(4). Such extensions have already been silently incorporated in (10) and (12). Yet mathematical physics has never quite gone through the very instructive effort of assigning the most appropriate vector species to the physical objects that need to be described. The notational differences between F , $\tilde{G}$ and $\mathfrak{G}$ will be explained in detail later.

Schouten's textbook [4] gives a striking example how the fields $\mathbf{E}$, $\mathbf{D}$, $\mathbf{H}$ and $\mathbf{B}$ are really four very distinct vector species under Diffeo(3). It was a distinction for which Faraday had already developed his very own method of geometric awareness in terms of tubes of force, potential surface pairs, and inner and outer orientations. All these little gems homing in on the central truth of intrinsic field renditions had

³Gregorio Ricci-Curbastro (1853–1925) was an Italian mathematician more commonly known as simply Ricci.

been indiscriminately bypassed and left to wither by the perhaps well-intended compromise of vector analysis.

Presentations of the theories of relativity as well as their transcription into differential forms have greatly suffered in content and impact through this undue restriction to SR(3)-reduced vector and tensor species. The method of raising and lowering indexes with the help of the metric tensor, as also used by Einstein himself, has been largely responsible for a *de facto* continuation of SR(3)-type reductions in the Riemannian world of General Relativity.

In fact, mathematical inadequacies of similar nature can be expected to permeate all of modern physics. It would seem that, instead of following the lure of higher dimensions, it would be better first to tidy up in three and four dimensions. The potential for cleanup and adding subtlety underlies any effort of establishing an intrinsic transformation behavior of physical fields under Diffeo(4). A basic procedure guided by physical dimensions is discussed in chapter II of my first monograph [8], and it leads quite naturally to the choice of transformation behavior as implied in the notation of (10) and (12).

As a deeper source for these dimensional recipes, an article by Truesdell and Noll [28] is recommended. Consider hereby that the basic laws of electromagnetism are global in origin, as expressed by integrals that have the same dimensions as the quanta of charge $[e]$, flux $[h/e]$ and action $[h]$. This very fact invites use of revised dimensional basis $[h, e, \ell, t]$ instead of the usual basis $[m, e, \ell, t]$, where m stands for mass, ℓ for length, and t for time.

In fact, a lot of basic physics is associated with the aforementioned integrals. Of these, the Aharonov–Bohm integral is today a favorite tool in quantum interferometry as a counter of flux quanta [18]. To be an accurate counter, the integration loop of the Aharonov–Bohm integral theoretically has to reside in a field-free region $\mathbf{E} = \mathbf{B} = \mathbf{0}$. If those conditions are met, the Aharonov–Bohm integral becomes a residue- or period integral [18–19]. However, the Aharonov–Bohm integral also shows relevance in orbital situations. Then, $\mathbf{E} \neq \mathbf{0}$ and $\mathbf{B} \neq \mathbf{0}$ and hence the period integral conditions are not met. Yet, if the particle has an $\mathbf{E} = \mathbf{B} = \mathbf{0}$ interior, one has an asymptotic period integral condition, see Ref. 18 for further explorations of this argument.

The orbital relevance of the Aharonov–Bohm integral indicates a microphysical applicability. This extension of validity from the macroscopic to the microscopic domains reveals a mathematical feature that is not talked about much. The Aharonov–Bohm integral is a flux-quantum counter in both domains, thus indicating that the integral is required to be metric-independent. In fact, only the metric is a reference for the macro-micro distinction. Once alerted to this feature of metric independence, an explicit proof of that property comes naturally.

The foregoing bird's eye view of closely interrelated matters was deemed necessary to evoke fuller awareness of significant ramifications of (10) and (12). It is worth emphasizing that the formalism delineated here carries well beyond the topic of complex electromagnetic mediums. Universal attention is justified, because it is the natural preparation for transitions from prequantum to quantum conditions.

5 Mathematical specifics

5.1 Transformation of tensors

Henceforth, we shall largely use the notation of Schouten [4]. Tensors of weighted transformation are denoted by Gothic or sometimes Greek symbols. Tensors weighted by sign change, with or without magnitude change, are marked by a tilde. These are all essential distinctions to be incorporated in a Diffeo description that is also accounting for orientation matters.

Suppose that q^λ and $q^{\lambda'}$ are two general spacetime coordinate systems, distinguished by unprimed versus primed indexes. Two Jacobian matrixes are defined as follows:

$$A_{\lambda}^{\lambda'} = \frac{\partial q^{\lambda'}}{\partial q^{\lambda}}, \quad A_{\lambda'}^{\lambda} = \frac{\partial q^{\lambda}}{\partial q^{\lambda'}}. \quad (15)$$

The determinant $\Delta = |A_{\lambda}^{\lambda'}|$, while the product $A_{\nu}^{\lambda'} A_{\lambda'}^{\kappa} = \delta_{\nu}^{\kappa}$ wherein δ_{ν}^{κ} is the Kronecker unit tensor.

Differentiability implies that the Jacobian matrixes are constants under alternating (exterior) differentiation; thus,

$$\partial_{[\nu'} A_{\lambda']}^{\lambda} = \partial_{[\nu} A_{\lambda]}^{\lambda'} = 0. \quad (16)$$

Furthermore, the chain rule holds, i.e.,

$$A_{\lambda'}^{\lambda} \partial_{\lambda} = \partial_{\lambda'}, \quad A_{\lambda}^{\lambda'} \partial_{\lambda'} = \partial_{\lambda}. \quad (17)$$

The transformation of some covariant fields, decorated by a tilde, is affected by the factor $\Delta/|\Delta|$, which changes sign for transformations with negative determinant. The latter is also a constant under differentiation and integration.⁴

Since the properties of flux and charge are governed by scalar-valued and pseudoscalar-valued integrals, the transformation of the electromagnetic fields have to be specified as follows:

$$F_{\lambda'\nu'} = A_{\lambda'}^{\lambda} A_{\nu'}^{\nu} F_{\lambda\nu}, \quad (18)$$

$$\tilde{G}_{\lambda'\nu'} = \frac{\Delta}{|\Delta|} A_{\lambda'}^{\lambda} A_{\nu'}^{\nu} \tilde{G}_{\lambda\nu}. \quad (19)$$

The Diffeo(4)-invariant version of the Minkowski form of the Maxwell equations in terms of these field renditions can be verified by using (16) and (17); thus,

⁴This appears just a triviality but determines the existence of *pair* and *impair* forms (these terms are defined later).

$\partial_{[\lambda} F_{\nu\kappa]} = 0$, as found previously in (10), and also $\partial_{[\lambda} \tilde{G}_{\nu\kappa]} = \tilde{C}_{\lambda\nu\kappa}$ given before as (12). The differential equation (10) implies that

$$F_{\nu\kappa} = \partial_{[\nu} A_{\kappa]}, \quad (20)$$

in which A is the 4-vector potential (distinct from $A_{\lambda'}^{\lambda}$).

The linear Diffeo(4) constitutive equation

$$\tilde{G}_{\lambda\nu} = \tilde{\chi}_{\lambda\nu}{}^{\sigma\kappa} F_{\sigma\kappa}, \quad (21)$$

now replaces (6) and (9), and can accommodate any linear medium properties, howsoever uncommon.⁵ In it, the linear constitutive tensor $\tilde{\chi}_{\lambda\nu}{}^{\sigma\kappa}$ has the same index symmetries as the Riemann–Christoffel tensor [14], [29] thus, by simple combinatorics, permitting 20 parameters. Substitution of (21) into (12), while using (20), yields the wave equation

$$\partial_{[\mu} \tilde{\chi}_{\lambda\nu]}{}^{\sigma\kappa} \partial_{\sigma} A_{\kappa} = \tilde{C}_{\mu\lambda\nu}. \quad (22)$$

The more traditional form of the Minkowski rendition of the Maxwell equations features the source-incorporating equation (12) in terms of a divergence of a contravariant tensor density. A transition between the two perfectly equivalent versions is obtained with the help of the two totally antisymmetric Levi–Civita unit tensors [4], [29]. The contravariant unit tensor density $\mathfrak{E}$ is needed to obtain the contravariant equivalents of $\tilde{G}$ and the 4-vector current density as follows:

$$\mathfrak{E}^{\lambda} = \tilde{\mathfrak{E}}^{\lambda\nu\sigma\kappa} \tilde{C}_{\nu\sigma\kappa}, \quad (23)$$

$$\mathfrak{G}^{\lambda\nu} = \tilde{\mathfrak{E}}^{\lambda\nu\sigma\kappa} \tilde{G}_{\sigma\kappa}. \quad (24)$$

Since the unit tensor commutes with differentiation, the operation of ∂_{ν} on (24)—with the help of (23)—yields the more familiar (11).

Note that $\mathfrak{G}$ transforms as a density of weight -1 and so does $\mathfrak{E}$. It means that the transformation of (12) is weighted by the reciprocal of the absolute value of the determinant. The numerical invariance of the unit tensor is easily verified by its transformation formula, weighted -1 but with the determinant itself and has a tilde marker as per

$$\tilde{\mathfrak{E}}^{\lambda' \nu' \sigma' \kappa'} = \Delta^{-1} A_{\lambda}^{\lambda'} A_{\nu}^{\nu'} A_{\sigma}^{\sigma'} A_{\kappa}^{\kappa'} \tilde{\mathfrak{E}}^{\lambda\nu\sigma\kappa}. \quad (25)$$

Contraction with the antisymmetric unit tensor is a determinant-forming operation and simply leads to Δ times the numerical value one cares to assign to the unit

⁵Time-delayed response (which is characteristic of material mediums) means the algebraic constitutive relation (21) should be replaced by a convolution integral. A temporal Fourier transform thereafter creates a complex constitutive tensor whose real and imaginary parts must be related by the Kramers–Kronig relations. The essential argumentation in this chapter is not affected.

tensor. In order not to overload this exposition with undue specifics, it may be convenient to leave that numerical value open, so as to be adjusted according to need.

Chances are the reader may be exasperated by all this detail. If it helps in keeping the faith, one should remember that vector analysis has obliterated reflections and inversions for more than a century. The distinctions are mandatory if the limitations due to SR(3) are to be repaired.

The wave equation (22) for the new description is

$$\partial_\nu \chi^{\lambda\nu\sigma\kappa} \partial_\sigma A_\kappa = \mathfrak{E}^\lambda, \quad (26)$$

where the isomer form⁶

$$\chi^{\lambda\nu\sigma\kappa} = \tilde{\chi}_{\mu\rho}^{\lambda\nu} \tilde{\mathfrak{E}}^{\mu\rho\sigma\kappa} \quad (27)$$

of the constitutive tensor relates to the Levi–Civita tensor. The isomer form of the constitutive tensor is weighted with the determinant’s absolute value as per

$$\chi^{\lambda'\nu'\sigma'\kappa'} = |\Delta|^{-1} A_{\lambda'}^{\lambda'} A_{\nu'}^{\nu'} A_{\sigma'}^{\sigma'} A_{\kappa'}^{\kappa'} \chi^{\lambda\nu\sigma\kappa}, \quad (28)$$

whereas the nonisomer form transforms with the sign of the Jacobian determinant as follows:

$$\tilde{\chi}_{\lambda'\nu'}^{\sigma'\kappa'} = \frac{\Delta}{|\Delta|} A_{\lambda'}^{\lambda'} A_{\nu'}^{\nu'} A_{\sigma'}^{\sigma'} A_{\kappa'}^{\kappa'} \tilde{\chi}_{\lambda\nu}^{\sigma\kappa}. \quad (29)$$

For consistent processing of improper transformations (i.e., reflections and inversions), it is essential to take these distinctions into account. If it is any consolation at all, consider it as a consequence of a century of neglect; even in early 20th century, Voigt [30] saw the impossibility of codifying crystal physics on the basis of vector analysis.

Now at the culmination of this bird’s eye view of Diffeo tensors, it is important to point out these methods culminate in the Cartan method of differential forms. Unlike the SR-group-restricted method of vector analysis in three dimensions, the method of forms is Diffeo-invariant for arbitrary dimensions, provided the presumably physically inspired de Rham distinctions between *pair* and *impair* forms is upheld. The method of forms is of special interest for integrals, because it came into existence while extending the local differential views of geometry to a global level for the purposes of topology.

The integration elements in spacetime are (by definition) contravariant antisymmetric tensors—defined by independent parameter lines⁷ q , r , s and u —spanning

⁶All four indexes are superscripted in the isomer form. The regular or the nonisomer form has two subscripted and two superscripted indexes.

⁷These are intersections of integration domains and coordinate surfaces.

the following subdomains of integration:

$$\begin{aligned}
 \text{line element:} & \quad dq^\lambda \\
 \text{surface element:} & \quad dq^{\lambda\nu} = dq^{[\lambda} dr^{\nu]} \\
 \text{3-“surface”element:} & \quad dq^{\lambda\nu\kappa} = dq^{[\lambda} dr^{\nu} ds^{\kappa]} \\
 \text{4-volume element:} & \quad dq^{\lambda\nu\kappa\sigma} = dq^{[\lambda} dr^{\nu} ds^{\kappa} du^{\sigma]}
 \end{aligned} \tag{30}$$

Instead of indexes in square brackets for the determinant–inspired product formation, the contemporary literature on differential forms uses the $\wedge$ notation, e.g.,

$$dq^{\lambda\nu\kappa} = dq^{[\lambda} dr^{\nu} ds^{\kappa]} = dq^\lambda \wedge dr^\nu \wedge ds^\kappa. \tag{31}$$

A differential form is defined as the scalar contraction of the here-encountered antisymmetric electromagnetic covariant p -fields with a contravariant p -element of integration. Electromagnetic theory has two absolute scalar-valued (i.e., pair) forms: a one-form

$$A = A_\lambda dq^\lambda \tag{32}$$

and a two-form

$$F = F_{\lambda\nu} dq^{\lambda\nu}. \tag{33}$$

Both have the physical dimension of flux $[h/e]$. Electromagnetic theory also has two pseudoscalar-valued (i.e., impair) forms, viz.,

$$\tilde{G} = \tilde{G}_{\lambda\nu} dq^{\lambda\nu}, \quad \tilde{C} = \tilde{C}_{\lambda\nu\kappa} dq^{\lambda\nu\kappa}. \tag{34}$$

Both $\tilde{G}$ and $\tilde{C}$ have the physical dimension of charge $[e]$.

Starting now with the dual of antisymmetric contravariant densities of weight -1 , there is another way of constructing differential forms by exterior product formation of a tensor of rank p and an $(n - p)$ -dimensional volume element; thus,

$$\tilde{C} = \mathfrak{C}^{[\lambda} dq^{\nu\sigma\kappa]}. \tag{35}$$

Since the alternation over four indexes in (35) amounts to the formation of a determinant, the alternating summation can be accomplished by contraction with the covariant alternating unit tensor of Levi–Civita (which is of weight $+1$) as follows:

$$\tilde{C} = \tilde{\mathfrak{E}}_{\lambda\nu\sigma\kappa} \mathfrak{C}^\lambda dq^{\nu\sigma\kappa}. \tag{36}$$

Equations (35) and (36) are identical because

$$\tilde{\mathfrak{E}}_{\lambda\nu\sigma\kappa} C^\sigma = \tilde{C}_{\lambda\nu\kappa}, \tag{37}$$

the last relation being the inversion of (23) which is consistent with the transformation of the covariant Levi–Civita tensor as per

$$\tilde{\mathfrak{E}}_{\lambda' \nu' \sigma' \kappa'} = \Delta A_{\lambda'}^{\lambda} A_{\nu'}^{\nu} A_{\sigma'}^{\sigma} A_{\kappa'}^{\kappa} \tilde{\mathfrak{E}}_{\lambda \nu \sigma \kappa}. \quad (38)$$

Without explicitly calling on the presence of covariant and contravariant aspects, the Cartan formalism retains all its detail without sacrificing any potential. Vector analysis, by contrast, requires the Cartesian restriction to do that, and cannot possibly deal with orientation changes. De Rham’s incorporation of impair and pair forms, finally, covers the need for dealing with orientation changes in physics and secures a one-to-one relationship with the way Schouten [4] treats tensors. While vector analysis’ SR(3) invariance has simply constant transformation elements, the Diffeo world of Cartan, Schouten and de Rham has nonconstant transformation elements, yet they behave as constants under exterior differentiation.

5.2 Differential forms and de Rham cohomology

Let us digress to a little *intermezzo* about forms and de Rham cohomology, although it is not directly connected to complex mediums, as it gives some unusual perspectives on the mathematical structure of electromagnetic theory.

In the language of differential forms, the laws of electrodynamics can be summarized in terms of properties of integrals of forms. The cyclic integral of the spacetime, pair, 2-form F vanishes for all cycles c_2 , i.e.,

$$\oint\!\!\!\oint_{c_2} F = 0. \quad (39)$$

The cyclic integral of the spacetime, impair, 3-form $\tilde{C}$ vanishes for all cycles c_3 as well:

$$\oint\!\!\!\oint\!\!\!\oint_{c_3} \tilde{C} = 0. \quad (40)$$

Differential forms obeying such properties are said to be *exact*.

According to a theorem by de Rham, exact forms are derivable by exterior differentiation from forms of lower order $F = dA$ and $\tilde{C} = d\tilde{G}$. The forms A and $\tilde{G}$ have cyclic integrals that do not vanish for all cycles:

$$\oint_{c_1} A \neq 0, \quad \oint\!\!\!\oint_{c_2} \tilde{G} \neq 0. \quad (41)$$

The integral of A —known as the Aharonov–Bohm integral—counts flux quanta, provided c_1 resides where $F = 0$. The other integral in (41) is the spacetime version of the Gauss integral to count the quanta of net charge, provided c_2 resides where $\tilde{C} = 0$. Equations (41) are Diffeo(4) quantum laws compatible with Einstein’s principle of general covariance.

Thus, electromagnetic theory inspired de Rham cohomology [19-18]. Yet, thus far, the all-pervasive Heisenberg uncertainty principle and the reluctance to extrapolate metric-free statements to the microscopic domain have blocked the application of de Rham cohomology to probe the topology of particle structures.

5.3 Constitutive specifics

Where traditional physics in free space (or vacuum) is seen as having one or perhaps two parameters that may really be taken as constants, I invite the reader to instead consider a constitutive tensor that can have no fewer than 20 independent components. In the previous section its mathematical characteristics were delineated, but what do we know physically about that tensor?

Present knowledge seems to indicate that, in an inertial frame, the constitutive tensor of vacuum in a Cartesian basis can be written as a 6×6 matrix as follows in Table 1. The numerical invariance of the configuration shown in Table 1 defines the conformal group [14]. The Lorentz group is the invariance group of the spacetime metric tensor and is a subgroup of the conformal group.

It is instructive to view these matters through the telescope of history. After Maxwell’s formulation of the macroscopic field equations, Lorentz—in his electron theory—postulated their validity in the microscopic domain and established their invariance under transformation now known as the Lorentz group. A few years later, Bateman [31] and Cunningham [32] established the conformal invariance of (1) and (2). Although rarely mentioned, the somewhat abstract discovery by Bateman and Cunningham supported the microscopic validity of the Maxwell equations postulated by Lorentz.

Let us recall the argument about (1) and (2) being a mixed field- and free-space constitutive statement. Clearly, then, the conformal invariance [14] emerges from the constitutive tensor in Table 1, because the field equations obey the wider $\text{Diffeo}(4)$ invariance.

Since the invariance of the constitutive tensor in Table 1 defines the conformal group as an optimum collective, one may ask: What kind of physically interesting transformation, outside the conformal group, is going to affect the form of this constitutive tensor? Using the transformation rule (28) and $\chi^{\lambda\nu\sigma\kappa}$ from Table 1,

Table 1 The 6×6 constitutive matrix of vacuum. It is equivalent to the fourth-rank constitutive tensor.

		E			B		
$\chi^{\lambda\nu\sigma\kappa}$		01	02	03	23	31	12
−D	01	$-\epsilon_0$	0	0	0	0	0
	02	0	$-\epsilon_0$	0	0	0	0
	03	0	0	$-\epsilon_0$	0	0	0
H	23	0	0	0	$1/\mu_0$	0	0
	31	0	0	0	0	$1/\mu_0$	0
	12	0	0	0	0	0	$1/\mu_0$

one finds [12] that a Galilei rotation at constant angular rate Ω and simultaneous transition to cylindrical coordinates $(r, \phi = \Omega t, z)$ yields Table 2. Substitution of the constitutive tensor from Table 2 into (26) yields a time-asymmetric wave equation describing electromagnetic wave propagation as seen from a cylindrical frame rotating with angular velocity Ω . Solving for circular solutions around the axis of rotation one finds the familiar frequency shift between left- and right-circulating waves of the Sagnac effect. More detailed discussions are available elsewhere [12].

Some readers will undoubtedly remark that they know simpler ways of arriving at the conclusions concerning the Sagnac effect. The point is well taken. Nobody likes to work through a transformation of a fourth-rank tensor, unless it is for proving some basic point. Unlike those simpler ways that cleverly circumvent electromagnetic fundamentals, the objective here is to show how the Sagnac effect can be consistently incorporated in electromagnetic theory. In the process of doing so, we have learned how to separate field equations and constitutive equations, and gained additional insight in the physical implications of the conformal group.

An inspection of Table 2 reveals that a cross-coupling exists between electric and magnetic fields in a rotating frame. Those who doubt the physical existence of such an effect are referred to experiments performed by Kennard [33] and Pegram [34]. The so-called Kennard–Pegram– effect is hereby identified as a cause of the Sagnac effect [12].

The question of whether or not the free-space constitutive tensor indeed shares with the Riemann–Christoffel tensor the condition

$$\chi^{[\lambda\nu\sigma\kappa]} = 0, \quad (42)$$

has been a point of fairly recent discussion, in reference to the so-called nonreciprocal biisotropic materials [35–36]. Whereas Sihvola takes the position that $\chi^{[\lambda\nu\sigma\kappa]}$ can be nonzero, Lakhtakia and Weiglhofer hold it to be identically zero [37–38].

Since the square brackets in (42) denote a determinant-forming operation, contraction with the covariant Levi–Civita unit tensor indeed accomplishes this determinant formation. Hence,

$$\tilde{\mathfrak{E}}_{\lambda\nu\sigma\kappa} \chi^{\lambda\nu\sigma\kappa} = \tilde{\chi}_{\sigma\kappa}^{\sigma\kappa} \quad (43)$$

Table 2 The 6×6 constitutive matrix of vacuum as seen from a frame rotating with constant angular velocity Ω about the distinguished axis of a cylindrical coordinate system.

$\chi^{\lambda\nu\sigma\kappa}$		E			B		
		01	02	03	23	31	12
–D	01	$-r\epsilon_0$	0	0	0	0	$\Omega r\epsilon_0$
	02	0	$-\epsilon_0/r$	0	0	0	0
	03	0	0	$-r\epsilon_0$	$-\Omega r\epsilon_0$	0	0
H	23	0	0	$-\Omega r\epsilon_0$	$1/r\mu_0$	0	0
	31	0	0	0	0	r/μ_0	0
	12	$\Omega r\epsilon_0$	0	0	0	0	$1/r\mu_0$

is a pseudoscalar. According to the Neumann principle, pseudoscalars cannot exist in a medium that has inversion symmetry. Except for conceivably extremely small gravity-induced asymmetry, matter-free space obeys inversion symmetry even on the surface of our planet. Hence in support of the Lakhtakia–Weiglhofer proposition, the pseudoscalar $\tilde{\epsilon}_{\lambda\nu\sigma\kappa}\chi^{\lambda\nu\sigma\kappa}$ must be taken to vanish for all mediums that have inversion symmetry [11].

Yet objects without inversion symmetry can exist in free space. Such objects normally occur in mirrored pairs (enantiomorphism), of which left- and right-handed quartz are examples in the macroscopic domain. Since inversion and mirror-reflection operations are independent of the metric field (which is solely responsible for the macro/micro distinction), there is no reason why matters of inversion symmetry are to be restricted to the macroscopic domain. Hence, electron and positron constitute an enantiomorphic pair in the microscopic domain.⁸

It now follows that the Sihvola option is only feasible in the absence of spatial inversion symmetry. The Sagnac effect manifests itself in an absence of spatial inversion symmetry and the off-diagonal 3×3 submatrixes appearing in Table 2 indeed have nonzero components. Those terms were generated, though, by transformation to rotating frames. A check shows that the pseudoscalar (43) remains zero. Once the right side of (42) is zero, its invariance secures the foregoing conclusion as a permanent feature.

Since a coordinate transformation can have no role in creating a nonzero pseudoscalar, only a matter-related feature can generate a magnetic-electric cross-coupling. Solutions of optically active sugars give examples of the phenomenon of natural optical activity [6-7]. The off-diagonal 3×3 submatrixes in the constitutive tensor of an optically active solution have—according to Condon [10]—nonzero traces, which are imaginary, accounting for the very pronounced dispersion of this effect. Nondissipative mediums, however, call for Hermitian properties of the 6×6 constitutive matrix as in Table 3. It means the pseudoscalar (43) is still zero. A conceivable relevance of Sihvola’s proposition of a nonvanishing $\chi^{[\lambda\nu\sigma\kappa]}$ could probably be illuminated by future developments.

Table 3 The 6×6 constitutive matrix of an isotropic medium exhibiting natural optical activity.

		E			B		
$\chi^{\lambda\nu\sigma\kappa}$		01	02	03	23	31	12
−D	01	$-\epsilon_0\epsilon$	0	0	$i\gamma$	0	0
	02	0	$-\epsilon_0\epsilon$	0	0	$i\gamma$	0
	03	0	0	$-\epsilon_0\epsilon$	0	0	$i\gamma$
H	23	$-i\gamma$	0	0	$1/\mu_0\mu$	0	0
	31	0	$-i\gamma$	0	0	$1/\mu_0\mu$	0
	12	0	0	$-i\gamma$	0	0	$1/\mu_0\mu$

⁸The electric charge emerges here as a pseudoscalar, an implication not generally adhered to in high-energy (particle) physics. Instead, the proposition of charge conjugation is meant to make up for this deficiency ensuing from the earlier SR(3)-invariant rendition of electromagnetism.

Table 4 The 6×6 constitutive matrix of a medium simultaneously exhibiting natural optical activity and the Faraday effect.

$\chi^{\lambda\nu\sigma\kappa}$		E			B		
		01	02	03	23	31	12
−D	01	$-\epsilon_0\epsilon$	0	0	$i\gamma$	0	0
	02	0	$-\epsilon_0\epsilon$	$i\epsilon_{23}$	0	$i\gamma$	0
	03	0	$-i\epsilon_{23}$	$-\epsilon_0\epsilon$	0	0	$i\gamma$
H	23	$-i\gamma$	0	0	$1/\mu_0\mu$	0	0
	31	0	$-i\gamma$	0	0	$1/\mu_0\mu$	0
	12	0	0	$-i\gamma$	0	0	$1/\mu_0\mu$

An experimental and theoretical test substantiating a measure of relevance of the described constitutive procedures has been made by Schultz *et al.* [39]. For the magnetically induced Faraday effect, Hermitian $i\epsilon_{23}$ terms in the 6×6 constitutive matrix are taken to be responsible. The constitutive tensor of Table 4 thus accounts for the Faraday effect as well as natural optical activity. The Faraday effect injects time-inversion asymmetry T and the natural optical rotation injects the spatial inversion asymmetry P . Substitution of the constitutive tensor from Table 4 into (26) gives a wave equation yielding a fourfold splitting of propagation modes.⁹ They have been shown to exist in a ringlaser setting.

6 Conclusion

If this essay on mathematical descriptions of complex medium properties has ended up in a plea for replacing SR(3)- by Diffeo(4)-invariant descriptions, let the reader rest assured that this proposition was made with the best of intentions. Many of us have seen such diversity in presentations of electromagnetic theory, that—out of despair—one just about has to fall back on good old vector analysis, if one wants to communicate at all.

The problem with that diversity of vector, tensor and spinor¹⁰ representations of electromagnetic theory is that few, if any, stipulate an underlying group of invariance. Even when the invariance group is stated, the reasons for choosing it remain in the dark. As a result, the contemporary situation hinges on Cartesian-based field quantities, which by necessity have a constitutive connotation, because the metric has a constitutive function in free space. Hence, the separation of field equations and constitutive equations requires a metric-free Diffeo(4)-invariant form of the Maxwell equations. Kottler [15], Cartan [16] and van Dantzig [17] have shown that such a form indeed exists. Any compromise that does not meet these natural

⁹Professor (emeritus) Hugh Walker (University of Houston) has observed that the four modes verify the four irreducible representations of Klein’s *Vierer* group [40] generated by time reversal T and space inversion $P : \{I, P, T, PT\}$.

¹⁰The two-valuedness of spinors is a consequence of the orientability of spacetime and its physical subspace. Hence spinors emerged when stepping out of the confines of SR(3) became unavoidable; see Ref. 18.

conditions is bound to end up in *ad hoc* concessions that compound a situation that is difficult enough anyway.

At the end of Section 4, it was mentioned that measures proposed here carry well beyond the realm of complex mediums. The inherently metric-independent Diffeo qualities of the methods of differential forms can hardly be expected to operate at their best potential when used in conjunction with Cartesian-based fields. That is a major reason why physical theorizing has for many years ignored de Rham's distinction between closed and exact forms. In contemporary physical theory, it is quite normal to consider magnetic monopoles along with the experimentally well-established flux quanta. In rejecting global mathematical consistency, one forfeits major perspectives on the relation between period integration and truly Diffeo(4)-invariant quantization statements, because quantization is intrinsically global.

The moral of this story is that many established physics procedures have failed to follow up on the principle of general covariance's potential to develop a measure of perceptual perfection in conjunction with the Neumann principle. Extending physical description, by admitting the whole gamut of groups from Diffeo(4) down to $SR(3)$ and discrete crystal groups to have their say, is a procedure that permits a more discerning assessment of physical law. Unbeknown to many is that spinors relate to aspects of orientability and object enantiomorphism [18]; they belong in the Diffeo hierarchy. It is just so that Diffeo(4) can see where $SR(3)$ has seriously impaired vision.

Notwithstanding very successful interactions between physics and mathematics in the 20th century, contemporary trends to specialize have nevertheless forced the two disciplines unduly apart. Yet, the laws of nature may be expected to be independent of how humans care to define their realms of expertise and what names they give it. When asking a question of Mother Nature, undue specialization can easily turn into prejudice.

Acknowledgments

In his lectures and writings, the late J. A. Schouten, Professor of Mathematics at the University of Amsterdam and the Delft Institute of Technology, used to argue how the geometric field images of Faraday testified to a measure of mathematical conscience. Unfortunately, many who later took leading roles in setting up electromagnetic theory missed that same measure of conscience. If the plea presented here for a sharper mathematical vision in physical theory were to arouse a change, much would be owed to Schouten's teachings. Van Dantzig [17] was an early recipient of his insights, and I had the same privilege decades later. I hope this effort not only meets those earlier endeavors, but also adds useful new perspectives.

I also thank both editors for their gracious help in overcoming software hurdles, which now seem to be inevitable companions of cybernetic progress. Yet, it also facilitated the injection of adaptations, which may enhance wider applicability.

References

- [1] A. Sommerfeld, *Lectures on theoretical physics. Vol. 3: Electrodynamics*, Academic Press, New York, NY, USA, 1952.
- [2] P.C.W. Davies, T.M. Davis and C.H. Lineweaver, Black holes constrain varying constants, *Nature* **418** (2002), 602–603.
- [3] H.A. Lorentz, A. Einstein, H. Minkowski and H. Weyl, *The principle of relativity*, Dover Press, New York, NY, USA, 1997. [Reprinted from Methuen, London, UK, 1923.]
- [4] J.A. Schouten, *Tensor analysis for physicists*, Clarendon Press, Oxford, UK, 1951.
- [5] J.D. Jackson, *Classical electrodynamics*, Wiley, New York, NY, USA, 1999.
- [6] E. Charney, *The molecular basis of optical activity*, Krieger, Malabar, FL, USA, 1985.
- [7] A. Lakhtakia (ed), *Selected papers on natural optical activity*, SPIE Press, Bellingham, WA, USA, 1990.
- [8] E.J. Post, *Formal structure of electromagnetics*, Dover Press, New York, NY, USA, 1952. [Reprinted from North-Holland, Amsterdam, The Netherlands, 1962.]
- [9] M. Born, *Optik*, Springer, Berlin, Germany, 1934.
- [10] E.U. Condon, Theories of optical rotatory power, *Rev Mod Phys* **9** (1937), 432–457. [Condon corrected the description of natural optical activity provided by Max Born [9].]
- [11] J.F. Nye, *Physical properties of crystals*, Clarendon Press, Oxford, UK, 1985.
- [12] E.J. Post, Sagnac effect, *Rev Mod Phys* **39** (1967), 475–493.
- [13] R.P. Feynman, R.B. Leighton and M.L. Sands, *The Feynman lectures on physics*, Addison-Wesley, Redwood City, CA, USA, 1965.
- [14] E.J. Post, The constitutive map and some of its ramifications, *Ann Phys* **71** (1972), 497–518.
- [15] F. Kottler, Maxwell'sche Gleichung und Metrik, *Sitzungsber Akad Wiss Wien Math-natur Kl. IIa* **131** (1922), 119–146.
- [16] E. Cartan, Les variétés a connexion affine, *Ann l'Ecole Norm Sup* **41** (1924), 1–25.
- [17] D. van Dantzig, The fundamental equations of electromagnetism, independent of metrical geometry, *Proc Camb Phil Soc* **30** (1934), 421–427.
- [18] E.J. Post, *Quantum reprogramming*, Kluwer, Dordrecht, The Netherlands, 1995.
- [19] G. de Rham, *Variétés différentiables*, Hermann, Paris, France, 1955.
- [20] A. Föppl, *Vorlesungen über technische Mechanik, Bd. 1-6*, Leibniz, Munich, Germany, 1921–1949. [Originally published during the late 19th century, some volumes were co-authored by others.]

- [21] J.W. Gibbs, *The scientific papers of J. Willard Gibbs*, Dover Press, New York, NY, USA, 1961.
- [22] D.J. Struik, From Riemann to Ricci: the origins of the tensor calculus, *Analysis, geometry and groups: a Riemann legacy volume* (H.M. Srivastava, ed), Hadronic Press, Palm Harbor, FL, USA, 1993.
- [23] H. Kleinert, *Path integrals in quantum mechanics, statistics, and polymer physics*, World Scientific, Singapore, 1990.
- [24] P. Lorrain and D. Corson, *Electromagnetic fields and waves*, Freeman, New York, NY, USA, 1970.
- [25] J. van Bladel, *Electromagnetic fields*, Hemisphere, Washington, DC, USA, 1985.
- [26] I. Bialynicki-Birula and Z. Bialynicki-Birula, *Quantum electrodynamics*, Pergamon, Oxford, UK, 1975.
- [27] S.N. Gupta, *Quantum electrodynamics*, Gordon & Breach, New York, NY, USA, 1977.
- [28] C.A. Truesdell and W. Noll, The non-linear field theories of mechanics, *Handbuch der Physik III/3* (S. Flügge, ed), Springer, Berlin, Germany, 1965.
- [29] M.N.L. Narasimhan, *Principles of continuum mechanics*, Wiley, New York, NY, USA, 1993.
- [30] W. Voigt, *Lehrbuch der Kristallphysik*, Teubner, Leipzig, Germany, 1910. [See the Preface.]
- [31] H. Bateman, The transformation of the electrodynamical equations, *Proc Lond Math Soc* **8** (1910), 223–264.
- [32] E. Cunningham, The principle of relativity in electrodynamics and an extension thereof, *Proc Lond Math Soc* **8** (1910), 77–98.
- [33] E.H. Kennard, On unipolar induction: another experiment and its significance for the existence of the aether, *Phil Mag* **33** (1917), 179–190.
- [34] G.B. Pegram, Unipolar induction and electron theory, *Phys Rev* **10** (1917), 591–600.
- [35] A. Lakhtakia and W.S. Weiglhofer, Are linear, nonreciprocal bi-isotropic media forbidden? *IEEE Trans Microw Theory Tech* **42** (1994), 1715–1716.
- [36] A.H. Sihvola, Are nonreciprocal bi-isotropic media forbidden indeed? *IEEE Trans Microw Theory Tech* **43** (1995), 2160–2162.
- [37] A. Lakhtakia and W.S. Weiglhofer, Constraint on linear, spatiotemporally nonlocal, spatiotemporally nonhomogeneous constitutive relations, *Int J Infrared Millim Waves* **17** (1996), 1867–1878.
- [38] W.S. Weiglhofer and A. Lakhtakia, The Post constraint revisited, *Arch Elektron Übertrag* **52** (1998), 276–279.
- [39] A.K. Schultz, R.M. Kiehn, E.J. Post and J.B. Roberds, Lifting of the four-fold EM degeneracy and PT asymmetry, *Phys Lett A* **74** (1979), 384–386.
- [40] M. Hamermesh, *Group theory and its application to physical problems*, Addison-Wesley, Reading, MA, USA, 1962.



Evert Jan Post was born in 1914 in Rotterdam, The Netherlands. Initially he envisioned a career in the merchant navy, but his impaired color vision forced him to reconsider. After a period of indecision, he opted for engineering physics at the Technical University of Delft. In 1946, he started working on crystal physics at the laboratory of the Netherlands Telecommunication Service in The Hague. During that period, he also became an active member of the circle of the mathematician J. A. Schouten at the University of Amsterdam. His doctoral thesis, on the principle of general covariance and the electromagnetic medium, is largely covered in the last four chapters of a 1962 monograph [8], which was reprinted in 1997 by the Dover Press. In 1957, he went to work at Bell Labs in New Jersey, USA. In 1960, he moved to the U.S. Air Force Cambridge Research Labs in Boston, participating in activities on laser navigation systems employing the Sagnac effect. He also spent a year at CalTech as a research associate. From 1970 to 1980, he intermittently taught physics at the University of Houston, Texas, and the University of Salford in the UK, and became acquainted with the de Rham theory. Extracurricular activities ensuing therefrom became focused on the quantum Hall effect and quantization from a de Rham perspective, which resulted in a 1995 monograph [18]. He continues to work on incorporating epistemic perspectives in the principle of covariance.

Constitutive Characterization of Simple and Complex Mediums

Werner S. Weiglhofer

Abstract

The solution of electromagnetic field problems requires two basic ingredients: the Maxwell equations and constitutive relations. The purpose of the latter is to provide a description of the electromagnetic properties of the medium or material in which the electromagnetic process (be it propagation, radiation or scattering) takes place. Here the focus is on some of the fundamental aspects of the constitutive characterization of both simple and complex mediums, approached through the dual viewpoints of field analysis in the space-time and the space-frequency domains. Subsequently, the topic is illuminated with a succession of case studies relating to mediums with an ever-increasing structural complexity, ranging from the classical vacuum to bianisotropic and nonlinear mediums. Along the way, important mathematical and physical requirements in the formulation of constitutive relations are encountered. A further focus of this chapter is the emergence of special types of constitutive relations from phenomenological models of materials and via the application of homogenization formalisms.

1 Introduction: the curtain rises

Evoking a symbolic image, one can relate theories of physical processes to elaborate plays with intricate plots. In electromagnetics and optics, these plots involve the radiation, propagation and scattering of electromagnetic waves, in a play where the rules are founded on the mathematical apparatus developed by James Clerk Maxwell in the 19th century. Having run for over a century, what could possibly be novel in such a play? Of course, a change of plot gives rise to a new play. Yet, there is an aspect, sometimes neglected, sometimes overlooked, that is supplied by the stage itself. In electromagnetics, that stage is the *medium* in which the electromagnetic processes take place.

Throughout the 20th century, the emphasis of electromagnetics research has gradually shifted towards the applied, and one can reasonably expect that trend to continue. An applied perspective is provided by the multitude of effects that arise due to the interaction of electromagnetic radiation with different types of mediums. A very close link between electromagnetics and the materials sciences is therefore established, and it is for this reason that the terms *medium* and *material* are used synonymously in this chapter.

What, then, is a *medium*? The purist's point of view may be that there is no such thing; all that is required are charged particles in vacuum, interacting with each other through well-defined and well-established physical laws. While that view can hardly be challenged from a fundamental perspective, it is of little practical use to the person holding a chunk of bulk material and wishing to characterize its electromagnetic and optical properties. Determining the motions of 10^{26} particles is not an intrinsically easy task.

The *vacuum* (or *free space*) itself is like Janus, a deity from the Roman mythology, with quantum and classical faces. In the days before quantum mechanics became to be understood as well as it is now, the vacuum (more accurately, the *classical* vacuum) was a somewhat boring place which by its definition had no content at all. Textbooks on quantum field theory now tell us that the vacuum is a place of tremendous complexity where virtual particles are created and annihilated due to constant energetic fluctuations. In that sense, the vacuum has perhaps become the most complex of all mediums. Nevertheless, unless one is specifically interested in quantum electrodynamics—and we are not in this chapter—it is the classical vacuum that provides the steady and comforting backdrop of our work.

On that basis, one can argue that the classical vacuum should not be called a medium after all, being completely devoid of any material interactions. Therefore, the classical vacuum is more appropriately designated as a *reference* medium, and so one arrives from the initial question: What is a medium? (in the context of electromagnetics) at the further question: When is a medium *simple* and when is it *complex*?

Electromagnetics research relating to different mediums/materials has of course existed for centuries, dating back to the earliest experiments of the interaction of light with glass, to crystallography and onwards to magnetized plasmas,

to name but a few examples of familiar milestones. Yet, it has only been during the final decade of the 20th century that even the major electromagnetics conferences would have special sessions dedicated to complex mediums or that specialist conferences on the electromagnetics and optics of complex mediums would spring up. The series of conferences that is the intellectual background of the book that contains the present chapter is the most prominent example [1]-[4], while more specialized events are also worth mentioning [5]-[7]; and an increasing number of books are being published on the topic of complex mediums [8]-[14].

Where it relates to the electromagnetic properties of a certain material, the term *complex* has remained only very loosely defined and can encompass vastly different research topics. A sensible working hypothesis takes the view that a linear, homogeneous, isotropic, dielectric-magnetic medium is a *simple* medium, while the relaxation of any of these attributes, be it linearity, homogeneity or (achiral) isotropy leads to a *complex* medium. Consequently, complex mediums range from the standard dielectrics of nonlinear optics to various exotic examples of anisotropic and bianisotropic mediums.

Whatever medium is being considered, in order to understand electromagnetic processes therein, we must be in possession of a characterization of its electromagnetic properties. This is the realm of *constitutive relations* that provides a basis for a classification of mediums by certain criteria such as distinctions on the basis of linearity versus nonlinearity, isotropy versus anisotropy/bianisotropy, homogeneity versus nonhomogeneity, spatial and/or temporal dispersion, and others. Not only will a thorough study of constitutive relations acquaint us with the stage for the play of electromagnetics and optics, it will equally contribute towards a better understanding of the genesis of electromagnetic complexity.

The purpose of this chapter, however, is less to provide a classification of materials or a definition of complexity by using constitutive relations. Rather, various fundamental aspects associated with the delineation of constitutive relations are highlighted, the genesis of constitutive relations in certain physical models of bulk matter are investigated and their mathematical structure as it is constrained by physical and mathematical conditions are studied. Needless to say, this is a vast topic and the author makes no claim to providing a comprehensive presentation in such a limited space. The selection of the covered topics will thus undoubtedly reflect a personal, but hopefully interesting and illuminating, viewpoint.

The layout of the chapter is as follows. Section 2 delineates two complementary approaches to electromagnetic field analysis in the space-time and space-frequency domains, respectively. In Section 3, fundamental aspects of constitutive relations are outlined. In Section 4, an exposition of constitutive relations for simple mediums is given, while Section 5 explores the vast realm of complex mediums, including the important distinction between linear and nonlinear mediums. Section 6 sheds some light onto symmetries and constraints, while Section 7 returns to some fundamental issues as to how constitutive dyadics can be determined by using electromagnetic homogenization formalisms. The chapter is rounded off with concluding remarks in Section 8.

The mathematical approach adopted here is based on the standard vector calculus of 3-vectors and 3×3 dyadics (second-rank Cartesian tensors). Of course, a variety of alternative descriptions have been developed. Examples are the covariant tensor formalism [15-16], the utilization of exterior differential forms [17] or the technique of 6-vectors and 6×6 dyadics [18-19].

The following notational conventions are used: vectors are in boldface, while dyadics/tensors of second rank are in normal face and underlined twice. Contraction of indexes is symbolized by a dot; that is, $\mathbf{a} \cdot \mathbf{b}$ stands for $\sum_n a_n b_n$ and $\mathbf{c} = \underline{\underline{A}} \cdot \mathbf{a}$ is a vector with components $c_n = \sum_p A_{np} a_p$. A dyadic $\underline{\underline{A}}$ is akin to a 3×3 matrix with elements A_{np} , and can be thought of as a linear combination of two or more dyads. A dyad, denoted by $\mathbf{a}\mathbf{b}$, can also be represented as a 3×3 matrix $a_n b_p$ [20]. The triplet of Cartesian unit vectors is denoted by $(\mathbf{u}_x, \mathbf{u}_y, \mathbf{u}_z)$, and the position vector by $\mathbf{r} = x\mathbf{u}_x + y\mathbf{u}_y + z\mathbf{u}_z$. The unit dyadic is $\underline{\underline{I}} = \mathbf{u}_x\mathbf{u}_x + \mathbf{u}_y\mathbf{u}_y + \mathbf{u}_z\mathbf{u}_z$, and $\underline{\underline{0}}$ is the null dyadic. The superscripts $^{-1}$ and ^{T} indicate inversion and transposition of a dyadic, respectively. Time-dependent quantities are distinguished by a tilde above their respective symbols.

2 Basics: the Maxwell equations

2.1 Space and time

Classical, i.e., nonquantum, electromagnetic theory is founded on the Maxwell equations. We shall first consider them in their standard, macroscopic representation with field vectors depending on the spatial variable $\mathbf{r}$ and time t . Thus,

$$\nabla \times \tilde{\mathbf{E}}(\mathbf{r}, t) + \frac{\partial \tilde{\mathbf{B}}(\mathbf{r}, t)}{\partial t} = -\tilde{\mathbf{J}}_m(\mathbf{r}, t), \quad (1)$$

$$\nabla \cdot \tilde{\mathbf{B}}(\mathbf{r}, t) = \tilde{\rho}_m(\mathbf{r}, t), \quad (2)$$

$$-\frac{\partial \tilde{\mathbf{D}}(\mathbf{r}, t)}{\partial t} + \nabla \times \tilde{\mathbf{H}}(\mathbf{r}, t) = \tilde{\mathbf{J}}_e(\mathbf{r}, t), \quad (3)$$

$$\nabla \cdot \tilde{\mathbf{D}}(\mathbf{r}, t) = \tilde{\rho}_e(\mathbf{r}, t). \quad (4)$$

The fields and sources that appear in these first-order partial differential equations are defined in Table 1.

To begin with, it is assumed in the present context that all sources on the right sides of (1)–(4) are prescribed in a consistent manner. Magnetic current and magnetic charge densities, $\mathbf{J}_m$ and ρ_m , respectively, are fictitious quantities; however, their introduction into the Maxwell equations serves two purposes: (i) the equations acquire a higher degree of symmetry (and show an associated invariance with respect to certain types of duality transformations [21]) and (ii) in some applications such usage leads to a simplified analysis; see the literature for more details [22-23].

Table 1 Fields and sources in the Maxwell equations in the space-time domain (1)–(4). The abbreviations for the units are V (volt), A (ampere = coulomb/s), m (meter) and s (second). See the text for a comment about the terminology regarding $\tilde{\mathbf{B}}$ and $\tilde{\mathbf{H}}$.

Symbol	Name	Unit
$\tilde{\mathbf{E}}(\mathbf{r}, t)$	electric field	V/m
$\tilde{\mathbf{B}}(\mathbf{r}, t)$	magnetic field	V s/m ²
$\tilde{\mathbf{D}}(\mathbf{r}, t)$	dielectric displacement	A s/m ²
$\tilde{\mathbf{H}}(\mathbf{r}, t)$	magnetic induction	A/m
$\tilde{\mathbf{J}}_e(\mathbf{r}, t)$	electric current density	A/m ²
$\tilde{\mathbf{J}}_m(\mathbf{r}, t)$	magnetic current density	V/m ²
$\tilde{\rho}_e(\mathbf{r}, t)$	electric charge density	A s/m ³
$\tilde{\rho}_m(\mathbf{r}, t)$	magnetic charge density	V s/m ³

Regarding the interpretation of the various field vectors that appear in (1)–(4), a clear distinction is provided in the space-time domain. The macroscopic field vectors $\tilde{\mathbf{E}}(\mathbf{r}, t)$ and $\tilde{\mathbf{B}}(\mathbf{r}, t)$ arise from spatially averaged microscopic fields. Indeed, $\tilde{\mathbf{E}}(\mathbf{r}, t)$ and $\tilde{\mathbf{B}}(\mathbf{r}, t)$ appear in the fundamental electromagnetic quantity, the Lorentz force, and may therefore be called *primitive* fields; while $\tilde{\mathbf{D}}(\mathbf{r}, t)$ and $\tilde{\mathbf{H}}(\mathbf{r}, t)$ are macroscopic fields induced by the presence of material interactions and appropriately termed *induction* fields [15], [22], [24]. The reader is warned, however, that such usage requires some care. Historically, magnetic terminology is confused; it is very common in the literature to call $\tilde{\mathbf{H}}(\mathbf{r}, t)$ the magnetic field and $\tilde{\mathbf{B}}(\mathbf{r}, t)$ the magnetic induction.

As a result of the distinction between primitive fields and induction fields that is adopted in this chapter, one of the special features of the representation of the Maxwell equations at the level of the differential equations is that primitive fields and induction fields are not coupled. Equations (1) and (2) are the realm of the former, (3) and (4) of the latter. This distinction is further investigated in Section 5.2.

2.2 Space and frequency

The fields introduced in the previous subsection are *real-valued* vector fields depending on position $\mathbf{r}$ and time t . A significant amount, if not the vast majority, of electromagnetics research dealing with simple and complex mediums alike is performed for frequency-dependent fields, however. In the standard procedure, one can apply a temporal Fourier transform

$$\tilde{\mathcal{Z}}(\mathbf{r}, t) = \int_{-\infty}^{+\infty} \mathcal{Z}(\mathbf{r}, \omega) \exp(-i\omega t) d\omega,$$
$$\mathcal{Z} = \mathbf{D}, \mathbf{E}, \mathbf{B}, \mathbf{H}, \mathbf{J}_e, \mathbf{J}_m, \rho_e, \rho_m,$$

(5)

to transform (1)–(4) into

$$\nabla \times \mathbf{E}(\mathbf{r}, \omega) - i\omega \mathbf{B}(\mathbf{r}, \omega) = -\mathbf{J}_m(\mathbf{r}, \omega), \quad (6)$$

$$\nabla \cdot \mathbf{B}(\mathbf{r}, \omega) = \rho_m(\mathbf{r}, \omega), \quad (7)$$

$$i\omega \mathbf{D}(\mathbf{r}, \omega) + \nabla \times \mathbf{H}(\mathbf{r}, \omega) = \mathbf{J}_e(\mathbf{r}, \omega), \quad (8)$$

$$\nabla \cdot \mathbf{D}(\mathbf{r}, \omega) = \rho_e(\mathbf{r}, \omega). \quad (9)$$

The two systems of vector partial differential equations (1)–(4) and (6)–(9) are common starting points for electromagnetic field analysis. As such they are not equivalent, however. In the first instance, not every function has a Fourier transform. Furthermore, and perhaps more importantly, the reverse Fourier synthesis is not unique because the inverse Fourier transform is not unique (the reader is referred to classical texts on Fourier and other transforms for more detailed background [25–26]). Quite apart from these issues, one must also appreciate the fundamental conceptual difference between the two systems: (1)–(4) contain real-valued (and thus measurable) field vectors depending on $\mathbf{r}$ and t , while in (6)–(9) the involved field quantities are *complex-valued field phasors* depending on $\mathbf{r}$ and angular frequency ω . As such, the latter are simply mathematical constructs. Important consequences of this distinction will become apparent later, especially in Section 4.2.

The main significance of the formulation of the Maxwell equations in the way it was done in this section is—to remain within the imagery of Section 1—that it contains both the players and the stage. Consequently, the equations remain unchanged regardless of the layout of the stage, i.e., regardless of the specific type of medium that is being studied.

3 Setting the stage: constitutive relations

Charge is the fundamental quantity in electromagnetics because it creates the fields while currents are generated by the motion of charges. Macroscopically, charges and currents, or rather their respective densities defined previously, satisfy the continuity conditions

$$\nabla \cdot \tilde{\mathbf{J}}_p(\mathbf{r}, t) + \frac{\partial \tilde{\rho}_p(\mathbf{r}, t)}{\partial t} = 0, \quad p = e, m, \quad (10)$$

which follow easily from (1)–(4). It will be assumed that all sources comply with the continuity conditions (10). Then, from the viewpoint of the mathematical analysis, the divergence equations (2) and (4) do not need to be further considered here as they are implied by the curl equations (1) and (3)—despite the fact that it is the divergence equations that contain the fundamental sources. As a consequence, the following system of partial differential equations is considered for fields depending

on space and time:

$$\nabla \times \tilde{\mathbf{E}}(\mathbf{r}, t) + \frac{\partial \tilde{\mathbf{B}}(\mathbf{r}, t)}{\partial t} = -\tilde{\mathbf{J}}_m(\mathbf{r}, t), \quad (1)$$

$$-\frac{\partial \tilde{\mathbf{D}}(\mathbf{r}, t)}{\partial t} + \nabla \times \tilde{\mathbf{H}}(\mathbf{r}, t) = \tilde{\mathbf{J}}_e(\mathbf{r}, t). \quad (3)$$

Equivalently, for field phasors depending on space and frequency, the continuity conditions become

$$\nabla \cdot \mathbf{J}_p(\mathbf{r}, \omega) - i\omega\rho_p(\mathbf{r}, \omega) = 0, \quad p = e, m. \quad (11)$$

Thus, one finds

$$\nabla \times \mathbf{E}(\mathbf{r}, \omega) - i\omega\mathbf{B}(\mathbf{r}, \omega) = -\mathbf{J}_m(\mathbf{r}, \omega), \quad (6)$$

$$i\omega\mathbf{D}(\mathbf{r}, \omega) + \nabla \times \mathbf{H}(\mathbf{r}, \omega) = \mathbf{J}_e(\mathbf{r}, \omega). \quad (8)$$

These two systems were repeated here such that one can see very clearly that (1) and (3) on the one hand, and (6) and (8) on the other, do not provide a sufficient number of differential equations to derive a solution for the fields or phasors; we recall that the source terms on the respective right sides are prescribed. In both cases there are twelve unknown components: three each for the field vectors $\tilde{\mathbf{D}}(\mathbf{r}, t)$, $\tilde{\mathbf{E}}(\mathbf{r}, t)$, $\tilde{\mathbf{H}}(\mathbf{r}, t)$ and $\tilde{\mathbf{B}}(\mathbf{r}, t)$ in the time-dependent case; and three each for the field phasors $\mathbf{D}(\mathbf{r}, \omega)$, $\mathbf{E}(\mathbf{r}, \omega)$, $\mathbf{H}(\mathbf{r}, \omega)$ and $\mathbf{B}(\mathbf{r}, \omega)$ in the frequency-dependent case, while the two systems only provide six partial differential equations each. As a consequence, six further relations (two vector equations) must be added in each case for a self-consistent description of the electromagnetic field. Initially, we specify these relations in a very general form as the vector equations

$$\mathbf{Y}_p^t \{ \tilde{\mathbf{E}}(\mathbf{r}, t), \tilde{\mathbf{D}}(\mathbf{r}, t), \tilde{\mathbf{H}}(\mathbf{r}, t), \tilde{\mathbf{B}}(\mathbf{r}, t) \} = \mathbf{0}, \quad p = e, m, \quad (12)$$

or

$$\mathbf{Y}_p^\omega \{ \mathbf{E}(\mathbf{r}, \omega), \mathbf{D}(\mathbf{r}, \omega), \mathbf{H}(\mathbf{r}, \omega), \mathbf{B}(\mathbf{r}, \omega) \} = \mathbf{0}, \quad p = e, m, \quad (13)$$

respectively. Therein, $\mathbf{Y}_{e,m}^t$ and $\mathbf{Y}_{e,m}^\omega$ are vector operators denoting some general functional dependence between the involved fields and phasors. The equations (12) are the *constitutive relations* for space-time dependent fields, while (13) are the constitutive relations for space-frequency dependent field phasors. What specific functional dependence is present in (12) or (13) depends on three issues.

1. In which kind of medium does the electromagnetic process take place: What is the *global* structure?

2. What constraints do physical conditions (such as symmetry requirements in crystals, for example, or reciprocity conditions) impose on the constitutive parameters: What are the *intrinsic* properties?
3. Are there structural conditions such as mathematical uniqueness requirements that further constrain the constitutive properties?

Providing the answers to these questions will be the aim of the following sections; for a conceptually different approach to constitutive relations the reader is referred to the chapter of Post [16].

To summarize this juncture, a self-consistent problem is therefore described by the Maxwell equations plus the constitutive relations that provide specific information about the material that is under consideration. To avoid misinterpretations of the usage of *self-consistent* in this context, it must be added that the delineation of a complete electromagnetic problem further requires the specification of boundary and/or initial conditions.

4 Exploring the stage: simple mediums

4.1 The classical vacuum as reference medium

The simplest implementation of constitutive relations applies to the classical vacuum as per

$$\tilde{\mathbf{D}}(\mathbf{r}, t) = \epsilon_0 \tilde{\mathbf{E}}(\mathbf{r}, t), \quad (14)$$

$$\tilde{\mathbf{H}}(\mathbf{r}, t) = \frac{1}{\mu_0} \tilde{\mathbf{B}}(\mathbf{r}, t). \quad (15)$$

As ϵ_0 and μ_0 are *true* constants,¹ the corresponding constitutive relations for the frequency-dependent field phasors, which follow from the inverse Fourier transform, are identical in appearance:

$$\mathbf{D}(\mathbf{r}, \omega) = \epsilon_0 \mathbf{E}(\mathbf{r}, \omega), \quad (16)$$

$$\mathbf{H}(\mathbf{r}, \omega) = \frac{1}{\mu_0} \mathbf{B}(\mathbf{r}, \omega). \quad (17)$$

It is important to realize that the vacuum is the only “medium” for which this identity in form holds. Vacuum is thus a reference medium for classical electromagnetism.

The constitutive relations of vacuum also define a *minimum* mathematical structure for constitutive relations in general. When delineating general functional de-

¹They are representing permittivity and permeability of vacuum, respectively, and are related via $\epsilon_0 \mu_0 = 1/c_0^2$, where c_0 is the speed of light in vacuum.

dependencies as in (12) or (13), one must aim to exclude pathological or contradictory, i.e., mathematically inconsistent, formulations (see [27] for some illuminating insights on this topic). One can require that the constitutive relations of any material must be reducible to those of the classical vacuum when all material interactions are removed.

4.2 Homogeneous isotropic dielectric-magnetic mediums

The linear, homogeneous, isotropic, dielectric-magnetic medium is the most widely employed material of electromagnetics research. Its frequency-dependent constitutive relations are given by

$$\mathbf{D}(\mathbf{r}, \omega) = \epsilon_0 \epsilon(\omega) \mathbf{E}(\mathbf{r}, \omega), \quad (18)$$

$$\mathbf{H}(\mathbf{r}, \omega) = \frac{1}{\mu_0 \mu(\omega)} \mathbf{B}(\mathbf{r}, \omega). \quad (19)$$

The true constants ϵ_0 and μ_0 are supplemented by two (dimensionless) constitutive parameters: relative permittivity $\epsilon(\omega)$ and relative permeability $\mu(\omega)$, both dependent on the circular frequency ω .

For this medium, one encounters for the first time the fundamental conceptual difference between constitutive relations in the frequency domain and in the time domain. For convenience, we first introduce an electric susceptibility χ_e and a magnetic susceptibility function χ_m as per

$$\epsilon(\omega) = 1 + \chi_e(\omega), \quad \mu^{-1}(\omega) = 1 - \chi_m(\omega). \quad (20)$$

We note parenthetically that χ_m was introduced for the benefit of the mathematical representation [28] and that its definition here does not correspond to the standard definition of a magnetic susceptibility [29]. Next, we define the electric and magnetic susceptibility kernels $\tilde{\chi}_e(t)$ and $\tilde{\chi}_m(t)$ as the Fourier transforms of the susceptibilities according to

$$\tilde{\chi}_p(t) = \int_{-\infty}^{+\infty} \chi_p(\omega) \exp(-i\omega t) d\omega, \quad p = e, m. \quad (21)$$

It follows that, on using the Fourier transform (5), (18)–(21) lead to

$$\tilde{\mathbf{D}}(\mathbf{r}, t) = \epsilon_0 \tilde{\mathbf{E}}(\mathbf{r}, t) + \int_{-\infty}^{+\infty} \epsilon_0 \tilde{\chi}_e(\tau) \tilde{\mathbf{E}}(\mathbf{r}, t - \tau) d\tau, \quad (22)$$

$$\tilde{\mathbf{H}}(\mathbf{r}, t) = \frac{1}{\mu_0} \tilde{\mathbf{B}}(\mathbf{r}, t) - \int_{-\infty}^{+\infty} \frac{1}{\mu_0} \tilde{\chi}_m(\tau) \tilde{\mathbf{B}}(\mathbf{r}, t - \tau) d\tau. \quad (23)$$

The representation in (22) and (23) provides a very instructive interpretation in separating the material response of the medium into a contribution from the

reference medium (the vacuum) and the actual material interaction (expressed in terms of convolution integrals). Indeed, it is motivated by the standard approach in describing the material properties of matter in the form

$$\tilde{\mathbf{D}}(\mathbf{r}, t) = \epsilon_0 \tilde{\mathbf{E}}(\mathbf{r}, t) + \tilde{\mathbf{P}}(\mathbf{r}, t), \quad (24)$$

$$\tilde{\mathbf{H}}(\mathbf{r}, t) = \frac{1}{\mu_0} \tilde{\mathbf{B}}(\mathbf{r}, t) - \frac{1}{\mu_0} \tilde{\mathbf{M}}(\mathbf{r}, t), \quad (25)$$

in terms of an electric polarization $\tilde{\mathbf{P}}(\mathbf{r}, t)$ and a magnetic polarization $\tilde{\mathbf{M}}(\mathbf{r}, t)$. There is a clear identification of the relevant terms in (24) and (25) with those in (22) and (23), respectively.

To interpret the constitutive relations (22) and (23) correctly [28], one must first realize they are nonlocal with respect to time. Indeed, if one imposes the requirements that

- (i) no material medium can respond before it is stimulated and
- (ii) no material can have an instantaneous response

(in other words, the medium must be *causal*), one is led to the conditions

$$\tilde{\chi}_e(t) \equiv 0, \quad \tilde{\chi}_m(t) \equiv 0, \quad t \leq 0. \quad (26)$$

The constitutive relations (22) and (23), together with (26) provide the proper material description of a causal, linear, homogeneous, isotropic, dielectric-magnetic medium. We note that (26) is also in accordance with the Kramers–Kronig relations [22]

$$\text{Re}[\gamma(\omega)] = 1 + \frac{1}{\pi} \mathcal{P} \int_{-\infty}^{+\infty} \frac{\text{Im}[\gamma(\omega')]}{\omega' - \omega} d\omega', \quad (27)$$

$$\text{Im}[\gamma(\omega)] = -\frac{1}{\pi} \mathcal{P} \int_{-\infty}^{+\infty} \frac{\text{Re}[\gamma(\omega')] - 1}{\omega' - \omega} d\omega', \quad (28)$$

where $\gamma = \epsilon$ or μ , while Re and Im are the real and imaginary parts of a complex-valued quantity, respectively, and $\mathcal{P}$ indicates a principal value operation [30].

In fact, the use of constitutive relations that technically apply only for phasors (which are frequency-dependent) to conduct a time-dependent field analysis remains an issue of considerable confusion in the research literature to the present day [31]–[35]. Such continuing practice is surprising in view of the observation that appropriate treatments through convolution integrals and causality conditions exist both in older [15], [22] and much more recent publications [28], [36]–[38].

5 A plethora of complex mediums

5.1 Beyond isotropy

Beyond the world of isotropy lies the vast realm of anisotropic and bianisotropic mediums. It is approached here through an example involving a plasma of electrons in thermal motion. For simplicity, a homogeneous, collisionless, compressible fluid model is employed [39]. The physical system comprises electrons in a classical vacuum; and the state of the system is described with $\tilde{\mathbf{E}}(\mathbf{r}, t)$ and $\tilde{\mathbf{B}}(\mathbf{r}, t)$, an average electron velocity $\tilde{\mathbf{v}}(\mathbf{r}, t)$ and an average electron pressure $\tilde{p}(\mathbf{r}, t)$. The Maxwell equations for vacuum are thus augmented by an electric current density due to the motion of the electrons as per

$$\nabla \times \tilde{\mathbf{E}}(\mathbf{r}, t) + \mu_0 \frac{\partial \tilde{\mathbf{H}}(\mathbf{r}, t)}{\partial t} = \mathbf{0}, \quad (29)$$

$$-\epsilon_0 \frac{\partial \tilde{\mathbf{E}}(\mathbf{r}, t)}{\partial t} + \nabla \times \tilde{\mathbf{H}}(\mathbf{r}, t) + n_0 q \tilde{\mathbf{v}}(\mathbf{r}, t) = \mathbf{0}. \quad (30)$$

These equations are supplemented by the Euler equations for a charged inviscid fluid, given by

$$\frac{1}{\gamma p_0} \frac{\partial \tilde{p}(\mathbf{r}, t)}{\partial t} + \nabla \cdot \tilde{\mathbf{v}}(\mathbf{r}, t) = 0, \quad (31)$$

$$n_0 q \tilde{\mathbf{E}}(\mathbf{r}, t) + \nabla \tilde{p}(\mathbf{r}, t) + n_0 m \frac{\partial \tilde{\mathbf{v}}(\mathbf{r}, t)}{\partial t} + n_0 q B_0 \tilde{\mathbf{v}}(\mathbf{r}, t) \times \mathbf{u}_0 = \mathbf{0}. \quad (32)$$

Therein, $-n_0 q$ and $n_0 m$ are the background electron charge and mass densities, p_0 is the background electron pressure and γ is the specific heat ratio for electrons. The model further contains a quasi-static magnetic field $\mathbf{B}_0 = B_0 \mathbf{u}_0$ of strength B_0 and fixed direction in space ($\mathbf{u}_0$ is a unit vector) that influences the motion of the electrons. The Lorentz force density $-n_0 q (\tilde{\mathbf{E}} + \tilde{\mathbf{v}} \times \mathbf{B}_0)$ is clearly recognizable in (32). No externally prescribed source terms have been added to the system (29)–(32).

It is apparent that, by writing (29)–(32), we have adopted the viewpoint that a medium consists essentially of electric charges interacting in vacuum. But, it is also clear that these equations are not microscopic anymore as the definition of an average electron velocity and an average electron pressure already implies an averaging procedure. In the next stage, however, the aim is to obtain an effective electromagnetic formalism by elimination of the nonelectromagnetic variables $\tilde{\mathbf{v}}(\mathbf{r}, t)$ and $\tilde{p}(\mathbf{r}, t)$. Straightforward manipulations permit recasting of (29)–(32) in the form (1) and (3) (setting $\tilde{\mathbf{J}}_e = \tilde{\mathbf{J}}_m = \mathbf{0}$ there), provided these two equations are supplemented by the constitutive relations

$$\tilde{\mathbf{D}}(\mathbf{r}, t) = \epsilon_0 \underline{\underline{\epsilon}} \star \tilde{\mathbf{E}}(\mathbf{r}, t), \quad (33)$$

$$\tilde{\mathbf{H}}(\mathbf{r}, t) = \frac{1}{\mu_0} \tilde{\mathbf{B}}(\mathbf{r}, t). \quad (34)$$

The effective electromagnetic medium has the same magnetic behavior as the vacuum, see (15), but the elimination of $\tilde{p}(\mathbf{r}, t)$ and $\tilde{\mathbf{v}}(\mathbf{r}, t)$ has manifested itself in a complicated constitutive relation linking $\tilde{\mathbf{D}}(\mathbf{r}, t)$ and $\tilde{\mathbf{E}}(\mathbf{r}, t)$. In (33), $\tilde{\underline{\underline{\epsilon}}} \star$ is a dyadic operator in space and time defined by

$$\tilde{\underline{\underline{\epsilon}}} \star = \underline{\underline{I}} + \omega_p^2 \left[\frac{\partial^2}{\partial t^2} \underline{\underline{I}} - a^2 \nabla \nabla - \omega_c (\mathbf{u}_0 \times \underline{\underline{I}}) \frac{\partial}{\partial t} \right]^{-1}. \quad (35)$$

Therein, ω_p is the *plasma frequency* defined by $\omega_p^2 = n_0 q^2 / m \epsilon_0$, and $\omega_c = q B_0 / m$ is the *gyrofrequency*, while $a^2 = \gamma p_0 / n_0 m$ is the *spatial dispersion parameter*.

The operator representation in (35) is only a formal one, but it exhibits two new features not discussed before. First, $\tilde{\mathbf{D}}(\mathbf{r}, t)$ and $\tilde{\mathbf{E}}(\mathbf{r}, t)$ are not parallel anymore; in other words, *anisotropy* has entered the stage. Second, in addition to *temporal dispersion* (by virtue of the presence of $\partial/\partial t$) there is also *spatial dispersion* due to the appearance of derivative operators ∇ in $\tilde{\underline{\underline{\epsilon}}} \star$. For didactic purposes these concepts shall be investigated separately here.

5.1.1 Anisotropy without spatial dispersion

Spatial dispersion in the plasma model disappears when the spatial dispersion parameter a is set equal to 0. Physically, this means that the plasma pressure $\tilde{p}(\mathbf{r}, t)$ is uniform in all space and does not vary with time. Consequently, (35) simplifies to

$$\tilde{\underline{\underline{\epsilon}}}_{an} \star = \underline{\underline{I}} + \omega_p^2 \left[\frac{\partial^2}{\partial t^2} \underline{\underline{I}} - \omega_c (\mathbf{u}_0 \times \underline{\underline{I}}) \frac{\partial}{\partial t} \right]^{-1}, \quad (36)$$

which can be manipulated into

$$\begin{aligned} \tilde{\underline{\underline{\epsilon}}}_{an} \star &= \left(1 + \frac{\omega_p^2}{\omega_c^2 + \partial_{tt}} \right) (\underline{\underline{I}} - \mathbf{u}_0 \mathbf{u}_0) + \frac{\omega_c \omega_p^2}{(\omega_c^2 + \partial_{tt}) \partial_t} \mathbf{u}_0 \times \underline{\underline{I}} \\ &\quad + \left(1 + \frac{\omega_p^2}{\partial_{tt}} \right) \mathbf{u}_0 \mathbf{u}_0. \end{aligned} \quad (37)$$

The last expression was obtained by using simple dyadic manipulations [20]; furthermore, the abbreviations $\partial_t \equiv \partial/\partial t$, $\partial_{tt} \equiv \partial^2/\partial t^2$ were used. A medium thus described is often called a *cold magnetized plasma*; a detailed analysis of the properties of electromagnetic waves propagating in it is available in the chapter by Kalluri [40], while related issues of linear magneto-optics are addressed by Boardman and Xie [41].

The operator representations in (36) and (37) are only formal ones. Yet, they clearly exhibit the special feature of anisotropy whereby $\tilde{\mathbf{D}}(\mathbf{r}, t)$ and $\tilde{\mathbf{E}}(\mathbf{r}, t)$ are not parallel anymore. Temporal dispersion is present by virtue of the presence of (inverse) time derivatives in the operator. Of course, one needs to clarify how to

calculate inverses of composite operators containing temporal derivatives. One can assign the meaning

$$\left(\frac{\partial}{\partial t}\right)^{-1} \cdot \equiv \int_{-\infty}^t \cdot d\tau, \quad (38)$$

and thus

$$\frac{1}{c + \partial_{tt}} f(t) = g(t) \Leftrightarrow (c + \partial_{tt})g(t) = f(t), \quad (39)$$

for some functions f and g and a constant c . This interpretation is equivalent to the use of convolution integrals as in (22) and (23).

It is in the frequency domain that a clearer interpretation of (33), (36) and (37) emerges. By using the inverse Fourier transform, the frequency-dependent constitutive relations

$$\mathbf{D}(\mathbf{r}, \omega) = \epsilon_0 \underline{\underline{\epsilon}}(\omega) \cdot \mathbf{E}(\mathbf{r}, \omega), \quad (40)$$

$$\mathbf{H}(\mathbf{r}, \omega) = \frac{1}{\mu_0} \mathbf{B}(\mathbf{r}, \omega), \quad (41)$$

result, wherein

$$\underline{\underline{\epsilon}}(\omega) = \epsilon_1(\omega) (\underline{\underline{I}} - \mathbf{u}_0 \mathbf{u}_0) + i\epsilon_2(\omega) \mathbf{u}_0 \times \underline{\underline{I}} + \epsilon_3(\omega) \mathbf{u}_0 \mathbf{u}_0. \quad (42)$$

The constitutive parameters are

$$\epsilon_1(\omega) = 1 + \frac{\omega_p^2}{\omega^2 - \omega_c^2}, \quad \epsilon_2(\omega) = -\frac{\omega_c \omega_p^2}{\omega(\omega^2 - \omega_c^2)}, \quad \epsilon_3(\omega) = 1 - \frac{\omega_p^2}{\omega^2}. \quad (43)$$

The transition from the operator $\tilde{\underline{\underline{\epsilon}}}_{an} \star$ to the dyadic $\underline{\underline{\epsilon}}(\omega)$ can be formally facilitated by the simple substitution $\partial/\partial t \rightarrow -i\omega$. These constitutive relations describe a *gyrotropic* medium of electric type. A similar dyadic structure for the magnetic properties describes a gyrotropic medium of magnetic type as exemplified by magnetized ferrites [20]. Macroscopically, an electrically gyrotropic medium is characterized by an anisotropy due to the presence of the distinguished direction $\mathbf{u}_0$ and three constitutive parameters ϵ_1 , ϵ_2 and ϵ_3 arranged in the specific structure (42). Beyond that, however, the physical model that was employed to study the electron plasma has also provided in (43) the exact dependence of those constitutive parameters on the circular frequency ω and on the fundamental plasma parameters, the plasma frequency ω_p and the gyrofrequency ω_c . Indeed, the earlier restriction of a collisionless plasma can now be relaxed by the formal substitution $\omega \rightarrow \omega - i\nu$, where ν is a suitably chosen *damping* parameter.

We have thus seen that nonlocal behavior in time is equivalent to local behavior in frequency: the frequency-dependent constitutive relations permit a local connection between $\mathbf{D}(\mathbf{r}, \omega)$ and $\mathbf{E}(\mathbf{r}, \omega)$, which is considerably easier to handle than the dependence of $\tilde{\mathbf{D}}(\mathbf{r}, t)$ on $\tilde{\mathbf{E}}(\mathbf{r}, t)$ via convolution integrals.

5.1.2 Spatial dispersion: the isotropic plasma

Let us now investigate the effect of spatial dispersion. For that purpose, we eliminate the effects of anisotropy by simply switching off the quasi-static magnetic field $\mathbf{B}_0$ in the plasma equations. As a consequence, (35) reduces to

$$\begin{aligned}\tilde{\underline{\underline{\epsilon}}}_{sd} \star &= \underline{\underline{I}} + \omega_p^2 \left(\frac{\partial^2}{\partial t^2} \underline{\underline{I}} - a^2 \nabla \nabla \right)^{-1} \\ &= \left(1 + \frac{\omega_p^2}{\partial_{tt}} \right) \underline{\underline{I}} + \frac{\omega_p^2 a^2}{\partial_{tt} (\partial_{tt} - a^2 \nabla^2)} \nabla \nabla,\end{aligned}\quad (44)$$

where ∇^2 is the Laplace operator. A transform into the space-frequency domain, as in the previous subsection, leaves (41) unchanged while we obtain

$$\mathbf{D}(\mathbf{r}, \omega) = \epsilon_0 \tilde{\underline{\underline{\epsilon}}}_{sd\omega} \star \mathbf{E}(\mathbf{r}, \omega), \quad (45)$$

where

$$\tilde{\underline{\underline{\epsilon}}}_{sd\omega} \star = \left(1 - \frac{\omega_p^2}{\omega^2} \right) \underline{\underline{I}} + \frac{\omega_p^2 a^2}{\omega^2 (\omega^2 + a^2 \nabla^2)} \nabla \nabla \quad (46)$$

is still an operator. Its specific form shows that the medium exhibits spatial nonlocality.

A constitutive relation that is fully local in its variables can only be obtained if one takes the further step from the space-frequency to the wavevector-frequency domain. This is accomplished by a three-dimensional Fourier transform with respect to the spatial coordinates as per

$$\mathcal{Z}(\mathbf{r}, \omega) = \int_{\mathbf{k}} \hat{\mathcal{Z}}(\mathbf{k}, \omega) \exp(i\mathbf{k} \cdot \mathbf{r}) d^3\mathbf{k} \quad (47)$$

for the involved fields, where the integration domain for the wavevector $\mathbf{k}$ is the complete unbounded $\mathbf{k}$ -space. Consequently, (45) becomes

$$\hat{\mathbf{D}}(\mathbf{k}, \omega) = \epsilon_0 \hat{\underline{\underline{\epsilon}}}(\mathbf{k}, \omega) \cdot \hat{\mathbf{E}}(\mathbf{k}, \omega), \quad (48)$$

where

$$\hat{\underline{\underline{\epsilon}}}(\mathbf{k}, \omega) = \left(1 - \frac{\omega_p^2}{\omega^2} \right) \underline{\underline{I}} - \frac{\omega_p^2 a^2}{\omega^2 (\omega^2 - a^2 \mathbf{k} \cdot \mathbf{k})} \mathbf{k} \mathbf{k}. \quad (49)$$

Table 2 Schematic diagram connecting the domains of space, time, frequency and wavevector. Horizontal equivalences are facilitated by the temporal Fourier transform (5) and its inverse, vertical ones by the spatial Fourier transform (47) and its inverse, and diagonal ones (not indicated by arrows) by the combined Fourier transform (50) and its inverse.

space-time $\mathbf{r}, t$	$\Leftrightarrow$	space-frequency $\mathbf{r}, \omega$
$\Updownarrow$		$\Updownarrow$
wavevector-time $\mathbf{k}, t$	$\Leftrightarrow$	wavevector-frequency $\mathbf{k}, \omega$

Phasors and dyadics in the wavevector-frequency domain are denoted by a caret ^ above the respective symbol.

More details on the approach to constitutive relations directly through the route of the wavevector-frequency domain are available to the interested reader in the literature [20], [42]. For the present purpose it is sufficient to state that the temporal Fourier transform (5) can be combined with its spatial equivalent (47) into

$$\tilde{\mathcal{Z}}(\mathbf{r}, t) = \int_{\mathbf{k}} \int_{\omega=-\infty}^{+\infty} \hat{\mathcal{Z}}(\mathbf{k}, \omega) \exp(i\mathbf{k} \cdot \mathbf{r} - i\omega t) d^3\mathbf{k} d\omega.$$

(50)

This expression facilitates a direct four-dimensional Fourier transform linking the space-time domain to the wavevector-frequency domain. Table 2 shows a schematic diagram about the connections between the various domains.

While the effects of anisotropy and spatial dispersion were isolated in the last two sections for didactic purposes, there is no reason against tackling the full problem, i.e., inverting the operator $\underline{\underline{\epsilon}} \star$ in (35), without making any simplifications. Readily available formulas for dyadic manipulations [20] give the mathematical means for such an enterprise which is left to the enthusiastic reader. With both anisotropy and spatial dispersion present, the medium is called a *warm* magnetized plasma [18].

5.2 Constitutive relations: generalities

We return to the general constitutive relations (12) and assume that they can be solved for $\tilde{\mathbf{D}}(\mathbf{r}, t)$ and $\tilde{\mathbf{H}}(\mathbf{r}, t)$. Thus,

$$\tilde{\mathbf{D}}(\mathbf{r}, t) = \tilde{\mathcal{L}}_D \{ \tilde{\mathbf{E}}(\mathbf{r}, t), \tilde{\mathbf{B}}(\mathbf{r}, t) \},$$

(51)

$$\tilde{\mathbf{H}}(\mathbf{r}, t) = \tilde{\mathcal{L}}_H \{ \tilde{\mathbf{E}}(\mathbf{r}, t), \tilde{\mathbf{B}}(\mathbf{r}, t) \},$$

(52)

for time-dependent fields, wherein $\tilde{\mathcal{L}}_D$ and $\tilde{\mathcal{L}}_H$ symbolize certain functional dependencies that may be of algebraic or differential/integral operator form. While these are still very general representations, they reflect the fact that for all practical purposes, constitutive relations are being substituted into the Maxwell equations, leading to an elimination of two of the unknown field vectors. Implicit within such a procedure is also a decision of which two of the four field vectors are to be eliminated. The foregoing choice reflects the point made in Section 2.1 about the distinction between the primitive fields and induction fields. In (51) and (52), the fields induced by the material interactions, $\tilde{\mathbf{D}}(\mathbf{r}, t)$ and $\tilde{\mathbf{H}}(\mathbf{r}, t)$, are explicitly expressed in terms of the primitive fields $\tilde{\mathbf{E}}(\mathbf{r}, t)$ and $\tilde{\mathbf{B}}(\mathbf{r}, t)$ [15], [22], [24]. As a consequence, substitution of (51) and (52) into (1) and (3) leads to field analysis exclusively in terms of $\tilde{\mathbf{E}}(\mathbf{r}, t)$ and $\tilde{\mathbf{B}}(\mathbf{r}, t)$; once these are obtained, $\tilde{\mathbf{D}}(\mathbf{r}, t)$ and $\tilde{\mathbf{H}}(\mathbf{r}, t)$ can be calculated, if so desired, from (51) and (52).

Pursuing the same strategy for the frequency-dependent field phasors, we restructure (13) into

$$\mathbf{D}(\mathbf{r}, \omega) = \mathcal{L}_D\{\mathbf{E}(\mathbf{r}, \omega), \mathbf{B}(\mathbf{r}, \omega)\}, \quad (53)$$

$$\mathbf{H}(\mathbf{r}, \omega) = \mathcal{L}_H\{\mathbf{E}(\mathbf{r}, \omega), \mathbf{B}(\mathbf{r}, \omega)\}, \quad (54)$$

where, again, $\mathcal{L}_D$ and $\mathcal{L}_H$ symbolize some general functional dependencies. These entail a field analysis in terms of $\mathbf{E}(\mathbf{r}, \omega)$ and $\mathbf{B}(\mathbf{r}, \omega)$.

However, being devoid of direct physical observability, field phasors are complex-valued quantities that can be used interchangeably. The bulk of all frequency-domain electromagnetics research is carried out in terms of $\mathbf{E}(\mathbf{r}, \omega)$ and $\mathbf{H}(\mathbf{r}, \omega)$, based on writing (53) and (54) in the alternative form

$$\mathbf{D}(\mathbf{r}, \omega) = \mathcal{L}_{D'}\{\mathbf{E}(\mathbf{r}, \omega), \mathbf{H}(\mathbf{r}, \omega)\}, \quad (55)$$

$$\mathbf{B}(\mathbf{r}, \omega) = \mathcal{L}_{B'}\{\mathbf{E}(\mathbf{r}, \omega), \mathbf{H}(\mathbf{r}, \omega)\}. \quad (56)$$

For linear mediums, as considered later in Section 5.3, a one-to-one correspondence between the $\mathbf{E}, \mathbf{H}$ (often called the *Tellegen* representation) and the $\mathbf{E}, \mathbf{B}$ formalism (often called the *Boys–Post* representation) can easily be facilitated [43–44]. Inclination towards the use of $\mathbf{E}(\mathbf{r}, \omega)$ and $\mathbf{H}(\mathbf{r}, \omega)$ in frequency-domain analysis is based on the conveniences

- (i) that substitution of (55) and (56) does not affect the terms involving spatial differentiation in (6) and (8);
- (ii) that boundary conditions on material interfaces are generally formulated in terms of $\mathbf{E}(\mathbf{r}, \omega)$ and $\mathbf{H}(\mathbf{r}, \omega)$; and
- (iii) that $\mathbf{E}(\mathbf{r}, \omega)$ and $\mathbf{H}(\mathbf{r}, \omega)$ define the time-averaged Poynting vector—an issue that is of some importance in the engineering literature.

The case studies of Sections 4.2 and 5.1 undoubtedly convey the impression that electromagnetic field analysis is more complicated when pursued in the

time domain than in the frequency domain. In addition to the necessity to take into proper account material responses that may be anisotropic or nonhomogeneous, time-domain formulations require complicated convolution integrals in their constitutive relations such that causality requirements are fulfilled. Therefore, a frequency-domain approach to a specific problem is often more feasible keeping in mind, as we saw in Section 5.1.2, that further transfer into the wavevector domain may provide additional advantages when spatial nonlocality occurs. Thus, we shall henceforth and for the remainder of this chapter keep our feet firmly rooted in the frequency domain. Final results of any specific calculation of fields *may*, if so desired, be transferred into the time domain by the appropriate inverse temporal Fourier transform.²

5.3 Linear mediums: bianisotropy

We now assume that the operators $\mathcal{L}_D$, $\mathcal{L}_H$, $\mathcal{L}_{D'}$ and $\mathcal{L}_{B'}$ appearing in (53)–(56) are *linear* operators to reflect linear response properties of the medium. Within the context of linearity, the most general type of medium is called *bianisotropic*. A linear, bianisotropic medium can thus be characterized by frequency-dependent constitutive relations given by

$$\mathbf{D}(\mathbf{r}, \omega) = \underline{\underline{\epsilon}}(\mathbf{r}, \omega) \cdot \mathbf{E}(\mathbf{r}, \omega) + \underline{\underline{\alpha}}(\mathbf{r}, \omega) \cdot \mathbf{B}(\mathbf{r}, \omega), \quad (57)$$

$$\mathbf{H}(\mathbf{r}, \omega) = \underline{\underline{\beta}}(\mathbf{r}, \omega) \cdot \mathbf{E}(\mathbf{r}, \omega) + \underline{\underline{\mu}}^{-1}(\mathbf{r}, \omega) \cdot \mathbf{B}(\mathbf{r}, \omega). \quad (58)$$

Therein, in the Boys–Post representation, $\underline{\underline{\epsilon}}(\mathbf{r}, \omega)$ is the permittivity dyadic, $\underline{\underline{\mu}}(\mathbf{r}, \omega)$ is the permeability dyadic, while $\underline{\underline{\alpha}}(\mathbf{r}, \omega)$ and $\underline{\underline{\beta}}(\mathbf{r}, \omega)$ are the two magnetoelectric dyadics. Spatial dispersion is not included here but can be easily implemented into (57) and (58) by interpreting the medium dyadics as operators in the same way as in Section 5.1.

To briefly illuminate the genesis of the term *bianisotropic*, it is noted that a medium with frequency-dependent constitutive relations of the form

$$\mathbf{D}(\mathbf{r}, \omega) = \underline{\underline{\epsilon}}(\mathbf{r}, \omega) \cdot \mathbf{E}(\mathbf{r}, \omega), \quad (59)$$

$$\mathbf{H}(\mathbf{r}, \omega) = \underline{\underline{\mu}}^{-1}(\mathbf{r}, \omega) \cdot \mathbf{B}(\mathbf{r}, \omega), \quad (60)$$

is traditionally called *anisotropic*. The “bi” in *bianisotropic* then indicates that each induction field depends not only on one primitive field, as in (59) and (60), but on *both* primitive fields, as in (57) and (58).

²The reader should recall the caveat expressed in the discussion following (9) in Section 2.2 about the nonuniqueness of the inverse Fourier transform and the amenability of certain functions to Fourier transforms.

Returning to the general bianisotropic case, we may alternatively use the Tellegen representation (hence the subscript T on the constitutive dyadics)

$$\mathbf{D}(\mathbf{r}, \omega) = \underline{\underline{\epsilon}}_T(\mathbf{r}, \omega) \cdot \mathbf{E}(\mathbf{r}, \omega) + \underline{\underline{\xi}}_T(\mathbf{r}, \omega) \cdot \mathbf{H}(\mathbf{r}, \omega), \quad (61)$$

$$\mathbf{B}(\mathbf{r}, \omega) = \underline{\underline{\zeta}}_T(\mathbf{r}, \omega) \cdot \mathbf{E}(\mathbf{r}, \omega) + \underline{\underline{\mu}}_T(\mathbf{r}, \omega) \cdot \mathbf{H}(\mathbf{r}, \omega). \quad (62)$$

The terminology for the constitutive dyadics $\underline{\underline{\epsilon}}_T$ (permittivity), $\underline{\underline{\mu}}_T$ (permeability), $\underline{\underline{\xi}}_T$ and $\underline{\underline{\zeta}}_T$ (magnetoelectric) remains the same. The two sets of constitutive relations (57) and (58) as well as (61) and (62) are linked through the identifications [43]

$$\underline{\underline{\epsilon}} = \underline{\underline{\epsilon}}_T - \underline{\underline{\xi}}_T \cdot \underline{\underline{\mu}}_T^{-1} \cdot \underline{\underline{\zeta}}_T, \quad (63)$$

$$\underline{\underline{\alpha}} = \underline{\underline{\xi}}_T \cdot \underline{\underline{\mu}}_T^{-1}, \quad (64)$$

$$\underline{\underline{\beta}} = -\underline{\underline{\mu}}_T^{-1} \cdot \underline{\underline{\zeta}}_T, \quad (65)$$

$$\underline{\underline{\mu}} = \underline{\underline{\mu}}_T; \quad (66)$$

or, vice versa,

$$\underline{\underline{\epsilon}}_T = \underline{\underline{\epsilon}} - \underline{\underline{\alpha}} \cdot \underline{\underline{\mu}} \cdot \underline{\underline{\beta}}, \quad (67)$$

$$\underline{\underline{\xi}}_T = \underline{\underline{\alpha}} \cdot \underline{\underline{\mu}}, \quad (68)$$

$$\underline{\underline{\zeta}}_T = -\underline{\underline{\mu}} \cdot \underline{\underline{\beta}}, \quad (69)$$

$$\underline{\underline{\mu}}_T = \underline{\underline{\mu}}. \quad (70)$$

Henceforth, choosing either the Boys–Post or the Tellegen representation when dealing with frequency-dependent fields will simply be a matter of convenience.

Regardless which representation is chosen, the four constitutive dyadics contain the full information of the electromagnetic response of the bianisotropic medium. They provide a vast parameter space with a multitude of specializations relating to different types of mediums. Specific structures in the constitutive dyadics/tensors, or generally, operators, give rise to specific electromagnetic effects. And so, for many decades a vast variety of different types of linear, bianisotropic mediums, homogeneous and nonhomogeneous, have been studied.

There is consequently an extensive playground between the maximum of $4 \times 9 = 36$ complex-valued constitutive parameters and the greatest simplification to only two quantities. The latter arise when the reduction to a linear, homogeneous, isotropic, dielectric-magnetic medium is considered. There, $\underline{\underline{\epsilon}} = \epsilon_0 \underline{\underline{I}}$, $\underline{\underline{\alpha}} = \underline{\underline{\beta}} \equiv \underline{\underline{0}}$ and $\underline{\underline{\mu}} = \mu_0 \underline{\underline{I}}$ provide a description in terms of just the two complex-valued scalars ϵ and μ . In view of the considerable attention bestowed on such mediums

in Section 4.2, let us briefly remain within the realm of isotropy before moving on to the terminology characterizing different types of anisotropy and bianisotropy.

The frequency-dependent constitutive relations

$$\mathbf{D}(\mathbf{r}, \omega) = \epsilon_0 \epsilon(\omega) \mathbf{E}(\mathbf{r}, \omega) + \alpha(\omega) \mathbf{B}(\mathbf{r}, \omega), \quad (71)$$

$$\mathbf{H}(\mathbf{r}, \omega) = \alpha(\omega) \mathbf{E}(\mathbf{r}, \omega) + \frac{1}{\mu_0 \mu(\omega)} \mathbf{B}(\mathbf{r}, \omega) \quad (72)$$

define a linear, isotropic *chiral* medium. Its constitutive characterization contains a third parameter α , the so-called *chirality* parameter. Crucially, while isotropy is retained, the main characteristic of this medium is a lack of mirror symmetry. Since the discovery of *optical activity* by Arago and Biot nearly 200 years ago, much research has been focused on such mediums; the interested reader can find out more about the many interesting electromagnetic and optical phenomena in isotropic chiral mediums in the chapter of Bohren [45] or in comprehensive monograph [9], encyclopedia [46] or review [47] publications. By virtue of the working hypothesis for a definition of the term (electromagnetic) complexity put forward in Section 1, an isotropic chiral material is not *simple* anymore but indeed complex. Yet, one of the great attractions to both theoreticians and practitioners in the field lies in the observation that the mathematical analysis of electromagnetic field problems in linear, isotropic chiral mediums can be reduced to (superposition of) the results readily available for linear, homogeneous, isotropic, dielectric-magnetic mediums. To quote Bohren [45]: “*Every single time-harmonic problem in electromagnetic theory that has ever been solved for linear, isotropic, achiral materials can also be solved for linear, isotropic, chiral materials.*”

Many of the mediums that inhabit the vast parameter space of bianisotropy and which are characterized by certain properties have well-known names that are based on the following dyadic structures:

$$\text{isotropic:} \quad \underline{\underline{a}} = a_1 \underline{\underline{I}}, \quad (73)$$

$$\text{uniaxial:} \quad \underline{\underline{a}} = a_1 \underline{\underline{I}} + a_3 \mathbf{u}_n \mathbf{u}_n, \quad (74)$$

$$\text{gyrotropic:} \quad \underline{\underline{a}} = a_1 \underline{\underline{I}} + a_2 \mathbf{u}_n \times \underline{\underline{I}} + a_3 \mathbf{u}_n \mathbf{u}_n, \quad (75)$$

$$\text{biaxial:} \quad \underline{\underline{a}} = a_1 \underline{\underline{I}} + \frac{a_3}{2} (\mathbf{u}_n \mathbf{u}_m + \mathbf{u}_m \mathbf{u}_n). \quad (76)$$

In these formulas, $\underline{\underline{a}}$ is a representative for $\underline{\underline{\epsilon}}_T$, $\underline{\underline{\xi}}_T$, $\underline{\underline{\zeta}}_T$ or $\underline{\underline{\mu}}_T$; a_1 , a_2 and a_3 are scalar parameters; while $\mathbf{u}_m$ and $\mathbf{u}_n$ are unit vectors.

Let us briefly exemplify nomenclature based on these dyadic structures. In the anisotropic regime, where the magnetoelectric dyadics $\underline{\underline{\xi}}_T$ and $\underline{\underline{\zeta}}_T$ vanish identically, a uniaxial $\underline{\underline{\epsilon}}_T$ and an isotropic $\underline{\underline{\mu}}_T$ defines a *uniaxial dielectric* medium; a gyrotropic $\underline{\underline{\epsilon}}_T$ and a gyrotropic $\underline{\underline{\mu}}_T$ denotes a *gyrotropic dielectric-magnetic* medium. In the bianisotropic regime, a *biaxial bianisotropic* medium is characterized by four

constitutive dyadics that are all biaxial. The reader is warned, however, that, for bianisotropic mediums, terminology can be dependent on the choice of representation as the identifications between the Boys–Post and the Tellegen representation given earlier readily show; see Ref. 48 for an example of such an occurrence. For various attempts towards consistent classification schemes the reader is referred to the literature [13], [43–44].

5.4 Beyond homogeneity

When constitutive parameters do not vary in space, they describe a *homogeneous* medium. While it is easy enough to postulate a *nonhomogeneous* medium by defining the constitutive parameters as arbitrary functions of position $\mathbf{r}$, actual field solutions cannot be expected to emerge without specifying mathematical functions that detail these dependencies. Not surprisingly, the mathematical analysis often becomes significantly more difficult when nonhomogeneous mediums are considered.

A specific type of nonhomogeneity that manifests itself as periodicity is considered henceforth. Well-studied examples of this type of nonhomogeneity are cholesteric liquid crystals (CLCs) with their frequency-domain constitutive relations given by [49]

$$\mathbf{D}(\mathbf{r}, \omega) = \epsilon_0 \underline{\underline{\epsilon}}(\mathbf{r}, \omega) \cdot \mathbf{E}(\mathbf{r}, \omega), \quad (77)$$

$$\mathbf{H}(\mathbf{r}, \omega) = \frac{1}{\mu_0} \mathbf{B}(\mathbf{r}, \omega), \quad (78)$$

where now

$$\underline{\underline{\epsilon}}(\mathbf{r}, \omega) = \epsilon_a(\omega) \underline{\underline{I}} + [\epsilon_b(\omega) - \epsilon_a(\omega)] \mathbf{u}(z) \mathbf{u}(z), \quad (79)$$

with

$$\mathbf{u}(z) = \mathbf{u}_x \cos \frac{\pi z}{\Omega} + \mathbf{u}_y \sin \frac{\pi z}{\Omega}. \quad (80)$$

This describes a helicoidal structure with periodicity along the z axis, with $\epsilon_a(\omega)$ and $\epsilon_b(\omega)$ as scalar parameters and $\mathbf{u}(z)$ as the local optical axis or *director*. The medium has a periodicity of Ω in the z -direction as per $\underline{\underline{\epsilon}}(z + \Omega, \omega) = \underline{\underline{\epsilon}}(z, \omega)$ and is a twisted, uniaxial, dielectric medium.

A speculative generalization of such a helicoidal structure led to the postulation of *helicoidal bianisotropic mediums* [50]–[52], specified by

$$\mathbf{D}(\mathbf{r}, \omega) = \epsilon_0 [\underline{\underline{\alpha}}_{11}(z, \omega) \cdot \mathbf{E}(\mathbf{r}, \omega) + \underline{\underline{\alpha}}_{12}(z, \omega) \cdot \mathbf{H}(\mathbf{r}, \omega)], \quad (81)$$

$$\mathbf{B}(\mathbf{r}, \omega) = \mu_0 [\underline{\underline{\alpha}}_{21}(z, \omega) \cdot \mathbf{E}(\mathbf{r}, \omega) + \underline{\underline{\alpha}}_{22}(z, \omega) \cdot \mathbf{H}(\mathbf{r}, \omega)], \quad (82)$$

where the nonhomogeneous constitutive dyadics $\underline{\underline{\alpha}}_{\lambda\nu}(z, \omega)$ have the representations

$$\underline{\underline{\alpha}}_{\lambda\nu}(z, \omega) = \underline{\underline{S}}(z) \cdot \underline{\underline{\beta}}_{\lambda\nu}(\omega) \cdot \underline{\underline{S}}^T(z), \quad \lambda = 1, 2, \quad \nu = 1, 2. \quad (83)$$

The unitary dyadic $\underline{\underline{S}}(z)$, given by

$$\underline{\underline{S}}(z) = \cos \frac{\pi z}{\Omega} (\mathbf{u}_x \mathbf{u}_x + \mathbf{u}_y \mathbf{u}_y) + \sin \frac{\pi z}{\Omega} (\mathbf{u}_y \mathbf{u}_x - \mathbf{u}_x \mathbf{u}_y) + \mathbf{u}_z \mathbf{u}_z, \quad (84)$$

is responsible for the rotational nonhomogeneity of the helicoidal bianisotropic medium.

Since the original conceptualization of helicoidal bianisotropic mediums, research has progressed rapidly, both theoretically and experimentally, towards sculptured thin films (STFs), which provide many interesting applications in optical and other technologies [53]. The constitutive characterization of STFs, in their various specifications, can exceed that of the helicoidal bianisotropic mediums even further in mathematical complexity; the unitary dyadic $\underline{\underline{S}}(z)$ may be a product of three different rotation dyadics because of the need to introduce as many as three rotation angles to describe the morphology of these highly complex mediums. More details can be found in the chapter by Lakhtakia and Messier elsewhere in this book [54]. The interested reader is also directed to a comprehensive review of STFs [55] and other review presentations [56-57]. The many papers in the *Complex Mediums* conferences [1]-[4] and a predecessor conference [58] dealing with various aspects of STFs further exhibit the great significance of these nanoengineered materials.

5.5 Nonlinear mediums

We choose here for the frequency-domain description of the nonlinear constitutive response of a medium a route via the introduction of a polarization phasor $\mathbf{P}(\mathbf{r}, \omega)$ and a magnetization phasor $\mathbf{M}(\mathbf{r}, \omega)$ in the frequency-domain counterparts of (24) and (25) as

$$\mathbf{D}(\mathbf{r}, \omega) = \epsilon_0 \mathbf{E}(\mathbf{r}, \omega) + \mathbf{P}(\mathbf{r}, \omega), \quad (85)$$

$$\mathbf{H}(\mathbf{r}, \omega) = \frac{1}{\mu_0} \mathbf{B}(\mathbf{r}, \omega) - \frac{1}{\mu_0} \mathbf{M}(\mathbf{r}, \omega). \quad (86)$$

At this stage it is convenient to introduce two 6-vectors (distinguished from 3-vectors by a single underlining) as per

$$\underline{\mathbf{E}}(\mathbf{r}, \omega) = \begin{pmatrix} \mathbf{E}(\mathbf{r}, \omega) \\ \mathbf{H}(\mathbf{r}, \omega) \end{pmatrix}, \quad (87)$$

$$\underline{\mathbf{Q}}(\mathbf{r}, \omega) = \begin{pmatrix} \mathbf{P}(\mathbf{r}, \omega) \\ \mathbf{M}(\mathbf{r}, \omega) \end{pmatrix}. \quad (88)$$

We may then split $\underline{\mathbf{Q}}(\mathbf{r}, \omega)$ into a linear and a nonlinear contribution:

$$\underline{\mathbf{Q}}(\mathbf{r}, \omega) = \underline{\mathbf{Q}}^L(\mathbf{r}, \omega) + \underline{\mathbf{Q}}^{NL}(\mathbf{r}, \omega), \quad (89)$$

where the linear part $\underline{\mathbf{Q}}^L(\mathbf{r}, \omega)$ contains the vacuum contribution and the linear constitutive properties.

The description of nonlinear medium properties is facilitated by first introducing a set of $M > 1$ angular frequencies as per [59]

$$\mathcal{W} = \{\omega_1, \omega_2, \omega_3, \dots, \omega_M\}, \quad (90)$$

with no requirement that all members of $\mathcal{W}$ shall be distinct. One further supposes that there exists an ensemble of M fields $\underline{\mathbf{F}}(\mathbf{r}, \omega_m)$, $1 \leq m \leq M$. Then, the simultaneous action of this ensemble of fields in the medium gives rise to the nonlinear part of $\underline{\mathbf{Q}}(\mathbf{r}, \omega)$ at frequency $\omega = \omega_{NL}$. The j -th element of this 6-vector is given by

$$\begin{aligned} Q_j^{NL}(\mathbf{r}, \omega_{NL}) = & \sum_{j_1=1}^6 \sum_{j_2=1}^6 \cdots \sum_{j_m=1}^6 \cdots \sum_{j_M=1}^6 \left\{ \chi_{j j_1 j_2 \dots j_m \dots j_M}^{NL}(\omega_{NL}; \mathcal{W}) \right. \\ & \left. \times \prod_{n=1}^M [F_{j_n}(\mathbf{r}, \omega_n)] \right\}, \quad 1 \leq j \leq 6. \end{aligned} \quad (91)$$

The frequency ω_{NL} is related to all members of $\mathcal{W}$ as

$$\omega_{NL} = \sum_{m=1}^M a_m \omega_m; \quad a_m = \pm 1, \quad 1 \leq m \leq M, \quad (92)$$

and it is implicit here that $F_{j_n}(\mathbf{r}, \omega_n)$ must be replaced by its complex conjugate on the right side of (91) and its derivative expressions when $a_n = -1$.

All nonlinear material properties are contained within the nonlinear susceptibility tensor

$$\chi_{j j_1 j_2 \dots j_m \dots j_M}^{NL}(\omega_{NL}; \mathcal{W}), \quad (93)$$

which is a 6-tensor of rank $M + 1$. The constitutive relation (91) is a mathematical description of the fact that fields oscillating at $\omega_m \in \mathcal{W}$, ($1 \leq m \leq M$), that are launched into the nonlinear medium, create fields oscillating at ω_{NL} . A vast number of known nonlinear dielectric effects [60-61] are described by this equation. In addition, (91) can represent magnetic as well as magnetoelectric effects of the nonlinear variety [62]-[64].

Many interesting papers of nonlinear optics and electromagnetics have appeared in the *Complex Mediums* conferences to date, and the reader is referred

to the relevant proceedings [1]–[4]. Particular mention is made of the results arising from nonlinear magneto–optics and the quantum mechanical work relating to semiconductors that can be found elsewhere in this book in the chapters of Boardman and Xie [41] and Arnold [65], respectively.

6 Regulating the stage: symmetries and constraints

6.1 General remarks

Theoretical exploration of the vast parameter space, spanned by the four constitutive dyadics, can very quickly become very difficult or at best tedious. Consider first the problem of plane wave propagation in a linear, homogeneous, bianisotropic medium. One substitutes (61) and (62) into the source-free version of the Maxwell (curl) equations (6) and (8), and assumes that the field phasors have a spatial dependence of the form $\exp(ik\mathbf{a} \cdot \mathbf{r})$, with k being the wavenumber and $\mathbf{a}$ being the direction of propagation. This procedure leads to a standard algebraic eigenvalue problem for the amplitudes of the phasors of $\mathbf{E}(\mathbf{r}, \omega)$ and $\mathbf{H}(\mathbf{r}, \omega)$. From a mathematical perspective, such an eigenvalue problem is trivial. The eigenvalues $k_p = k_p(\mathbf{a})$, $p = 1, 2, 3, 4$, arise from a quartic equation. Writing them and the corresponding eigenvectors explicitly when $\underline{\underline{\epsilon}}_T$, $\underline{\underline{\xi}}_T$, $\underline{\underline{\zeta}}_T$ and $\underline{\underline{\mu}}_T$ retain their full, unconstrained structure, is possible but appears to be of little benefit. The solution contains the 36 complex-valued parameters in such a complicated algebraic form that the influence of one or even a small group of constitutive parameters can hardly be understood. The stage has become so crowded that the plot is indiscernible! Yet, such an understanding is highly desirable as it relates to the recognition of novel electromagnetic and optical effects. This applies first to a theoretical aspect where one wishes to obtain a clear view as to what a certain constitutive parameter does. Following on from that, it leads to the applied aspect of working with materials (either natural, if available, or specifically manufactured) that allow the exploitation of a novel property towards new technology.

The situation is different when one investigates radiation problems driven by prescribed sources. The key quantity in this enterprise are the infinite-medium dyadic Green functions. Obtaining these mathematical objects in explicit form is only possible for very specialized choices for the elements of $\underline{\underline{\epsilon}}_T$, $\underline{\underline{\xi}}_T$, $\underline{\underline{\zeta}}_T$ and $\underline{\underline{\mu}}_T$ which, to phrase it in a somewhat simplified manner, must not contain more than a small number of nonvanishing constitutive parameters. For what can and cannot be achieved, the interested reader is referred to a review on infinite–medium dyadic Green functions [66].

A reduction of the number of constitutive parameters is thus desirable. Specializations and/or simplifications in general medium descriptions should therefore be founded on two important principles:

- (i) they must be based on sound physical and mathematical considerations and
- (ii) they must contain, at least in the form of a principal mechanism, a scheme how such a medium, if it does not exist in natural form, can be manufactured.

In the following, the second and third issues highlighted in Section 3 (after constitutive relations were introduced for the first time) are addressed. In the more specific context of linear, bianisotropic mediums considered now, this requires an understanding of how symmetries and constraints influence the structure of the four constitutive dyadics and thus how the constitutive parameter space is reduced.

6.2 Reciprocity

A very common requirement in theoretical studies consists in the imposition that a specific medium should be Lorentz-reciprocal. Expressed mathematically, in the Tellegen representation, this means [67-68]

$$\underline{\underline{\epsilon}}_T = \underline{\underline{\epsilon}}_T^T, \quad \underline{\underline{\xi}}_T = -\underline{\underline{\xi}}_T^T, \quad \underline{\underline{\mu}}_T = \underline{\underline{\mu}}_T^T. \quad (94)$$

The conditions (94) reduce the constitutive parameter space from 36 to 21 complex-valued parameters. While many mediums, for example, uniaxial or biaxial dielectric-magnetic mediums, are reciprocal, others are not. The magnetized plasma of Section 5.1 and more general types of gyrotropic mediums belong to the most prominent representatives of nonreciprocal mediums.

6.3 Losslessness

Continuing with this theme, there may be further conditions based on appropriate approximations; for example, absence of dissipation at a certain frequency requires [23]

$$\underline{\underline{\epsilon}}_T = \underline{\underline{\epsilon}}_T^\dagger, \quad \underline{\underline{\xi}}_T = \underline{\underline{\xi}}_T^\dagger, \quad \underline{\underline{\mu}}_T = \underline{\underline{\mu}}_T^\dagger. \quad (95)$$

In (95), the superscript † denotes transposition and complex conjugation. Considered by themselves, the conditions of losslessness also reduce the parameter space from 36 to 21 complex-valued constitutive parameters.

Superimposed, a reciprocal medium that is (approximately) free of dissipation is characterized by at most 21 constitutive parameters, 12 of which are real-valued and the remaining 9 are purely imaginary.

6.4 A structural constraint

Is it possible that the Maxwell equations (1)–(4) may have hidden within them any structural and uniqueness constraints that can provide conditions for the structure

of a medium? While original work on this question has lain dormant in the literature for 40 years [15], it was only recognized within the last decade that at least one such constraint does indeed exist. It turns out that any linear bianisotropic medium must satisfy the Post constraint [15], [24], [69]–[71]

$$\text{Trace}[\underline{\underline{\alpha}}(\mathbf{r}, \omega) - \underline{\underline{\beta}}(\mathbf{r}, \omega)] \equiv 0, \quad (96)$$

for constitutive relations expressed in the Boys–Post form, or

$$\text{Trace}[\underline{\underline{\mu}}^{-1}_T(\mathbf{r}, \omega) \cdot (\underline{\underline{\xi}}_T(\mathbf{r}, \omega) + \underline{\underline{\zeta}}_T(\mathbf{r}, \omega))] \equiv 0, \quad (97)$$

in the Tellegen form. The Post constraint is an algebraic condition, relevant—as can easily be seen by inspection—whenever magnetoelectric coupling is present. All purely anisotropic (as opposed to *bianisotropic*) mediums fulfill (96) identically, as can be seen by substituting $\underline{\underline{\alpha}}(\mathbf{r}, \omega) = \underline{\underline{\beta}}(\mathbf{r}, \omega) \equiv \underline{\underline{0}}$ into (96) or, equivalently, $\underline{\underline{\xi}}_T(\mathbf{r}, \omega) = \underline{\underline{\zeta}}_T(\mathbf{r}, \omega) \equiv \underline{\underline{0}}$ into (97), respectively. The Post constraint thus reduces the constitutive parameter space of the general bianisotropic medium by one complex-valued quantity; it *always* does, regardless if other constraints such as reciprocity or losslessness also exist.

It is clear that the Post constraint is not a reciprocity condition; compare (97) with (94). However, the Post constraint has its most important consequence for linear *biisotropic* materials [10]. Such materials are mathematically described by four complex-valued constitutive parameters according to

$$\underline{\underline{\epsilon}} = \epsilon \underline{\underline{I}}, \quad \underline{\underline{\alpha}} = \alpha \underline{\underline{I}}, \quad \underline{\underline{\beta}} = \beta \underline{\underline{I}}, \quad \underline{\underline{\mu}} = \mu \underline{\underline{I}}. \quad (98)$$

Substitution of (98) into (96) leads immediately to the requirement that $\alpha = \beta$. This means that the Post constraint for a biisotropic medium, which is nonreciprocal as such, coincides with the condition that also renders it reciprocal. Yet a reciprocal biisotropic medium is an *isotropic chiral* medium [9], and its constitutive relations were given in (71) and (72). It emerges therefore, that the Post constraint negates the recognizable existence of *nonreciprocal* biisotropic mediums: the isotropic chiral medium is the most general isotropic medium that contains magnetoelectric coupling. For more details on this issue the reader is referred to a comprehensive review [71] that contains a literature survey plus detailed evaluation of experimental and theoretical background on this subject.

6.5 Symmetries: biaxial bianisotropic mediums

A widely used approach to the reduction of the large constitutive parameter space is the exploitation of symmetries. Symmetries in constitutive relations can occur because they appear intrinsically in the studied materials, and crystals provide well-known examples; see Litvin’s chapter on crystallography [72]. We exemplify here

the incorporation of symmetries into constitutive dyadics via the concept of *biaxial bianisotropy*.

Biaxial bianisotropy can be defined by a simple symmetry requirement in which we demand that a linear, homogeneous, bianisotropic medium as a whole is characterized by two distinct (crystallographic) axes. We then obtain the constitutive relations as [73]

$$\mathbf{D}(\mathbf{r}, \omega) = \epsilon_0 \underline{\underline{\epsilon}}_{bi}(\omega) \cdot \mathbf{E}(\mathbf{r}, \omega) + (1/\eta_0) \underline{\underline{\alpha}}_{bi}(\omega) \cdot \mathbf{B}(\mathbf{r}, \omega), \quad (99)$$

$$\mathbf{H}(\mathbf{r}, \omega) = (1/\eta_0) \underline{\underline{\beta}}_{bi}(\omega) \cdot \mathbf{E}(\mathbf{r}, \omega) + (1/\mu_0) \underline{\underline{\mu}}_{bi}^{-1}(\omega) \cdot \mathbf{B}(\mathbf{r}, \omega), \quad (100)$$

($\eta_0 = \sqrt{\mu_0/\epsilon_0}$). The dimensionless constitutive dyadics are given by [73]

$$\underline{\underline{\epsilon}}_{bi}(\omega) = \epsilon_a(\omega) \underline{\underline{I}} + \epsilon_b(\omega)(\mathbf{u}_m \mathbf{u}_n + \mathbf{u}_n \mathbf{u}_m), \quad (101)$$

$$\underline{\underline{\alpha}}_{bi}(\omega) = \alpha_a(\omega) \underline{\underline{I}} + \alpha_b(\omega)(\mathbf{u}_m \mathbf{u}_n + \mathbf{u}_n \mathbf{u}_m), \quad (102)$$

$$\underline{\underline{\beta}}_{bi}(\omega) = \beta_a(\omega) \underline{\underline{I}} + \beta_b(\omega)(\mathbf{u}_m \mathbf{u}_n + \mathbf{u}_n \mathbf{u}_m), \quad (103)$$

$$\underline{\underline{\mu}}_{bi}^{-1}(\omega) = \sigma_a(\omega) \underline{\underline{I}} + \sigma_b(\omega)(\mathbf{u}_m \mathbf{u}_n + \mathbf{u}_n \mathbf{u}_m). \quad (104)$$

In the foregoing expressions, $\mathbf{u}_m$ and $\mathbf{u}_n$ are the two distinguished axes of the biaxial structure. The biaxial bianisotropic medium is a nonreciprocal medium. It is characterized by the eight complex-valued and frequency-dependent constitutive parameters ϵ_a , ϵ_b , α_a , α_b , β_a , β_b , σ_a , σ_b (the last two being essentially inverse permeability scalars), plus a real-valued angle ϕ , whereby $\cos \phi = \mathbf{u}_m \cdot \mathbf{u}_n$.

We take the opportunity to illustrate the concepts of Sections 6.2 and 6.4. The Post constraint (96) simplifies for the biaxial bianisotropic medium to

$$3(\alpha_a - \beta_a) + 2(\alpha_b - \beta_b) \cos \phi = 0, \quad (105)$$

thus reducing the number of unconstrained medium parameters from $8 \times 2 + 1 = 17$ to 15 real-valued quantities. Reciprocity, on the other hand, requires

$$\alpha_a = \beta_a, \quad \alpha_b = \beta_b. \quad (106)$$

One can therefore recognize that every biaxial bianisotropic medium that is reciprocal also satisfies the Post constraint, while the converse is not true. The reader should appreciate however, that the status of these two conditions is not equivalent. Equation (105) is dictated by the structure of the Maxwell equations themselves in that only a biaxial bianisotropic medium with constitutive parameters in accordance with (105) can recognizably exist. The reciprocity condition (106), on the other hand, may or may not be satisfied.

We mention in passing that the theoretical characterization of biaxial bianisotropic mediums [73] provided the basis for detailed numerical homogenization studies [74]-[76]; see also a comprehensive review of those results [77].

7 Preparing the stage: homogenization

Previously, in Section 5.1, an instructive example was presented that had as its basis a simple physical system, that of an electron plasma in thermal motion. It was shown how the elimination of the nonelectromagnetic field variables led to an effective electromagnetic formalism; in other words, “simple” ingredients generated a complex medium. The purpose of this final section is to highlight a different, but exceedingly valuable source for the delineation of constitutive relations. These are the homogenization formalisms.

Homogenization formalisms aim to provide quantitative estimates for the constitutive parameters of a homogenized composite medium, based on a knowledge of the constitutive parameters and various geometrical/topological quantities pertaining to two or more constituent mediums. Again, complexity arises from simplicity, but through a different mechanism from the plasma case outlined previously.

Homogenization formalisms operate within well-defined limits of applicability. Best known among the various mathematical formalisms are those associated with the names of Maxwell Garnett and Bruggeman [78] and variations thereof [79]. Among the vast amount of literature on homogenization reference is made to introductory works [80] as well as detailed reviews [81-82]. Very recent work has also provided significant progress arising from the application and extension of another homogenization formalism, the strong-property-fluctuation theory (SPFT), both in the linear [83]-[85] and nonlinear [86-87] regime. A survey of recent electromagnetic homogenization research can be found in the chapter by Mackay elsewhere in this book [88].

7.1 Faraday chiral mediums

The homogenization procedure is illustrated by laying out the recipe to homogenize a chiroferrite composite. Take two component mediums, the first one a magnetically biased ferrite characterized by [20]

$$\mathbf{D}(\mathbf{r}, \omega) = \epsilon_0 \epsilon^f(\omega) \mathbf{E}(\mathbf{r}, \omega), \quad (107)$$

$$\mathbf{B}(\mathbf{r}, \omega) = \mu_0 [\mu^f(\omega) (\underline{\underline{I}} - \mathbf{u}\mathbf{u}) + i\mu_g^f(\omega) \mathbf{u} \times \underline{\underline{I}} + \mu_u^f(\omega) \mathbf{u}\mathbf{u}] \cdot \mathbf{H}(\mathbf{r}, \omega), \quad (108)$$

where ϵ^f is the relative permittivity scalar, while μ^f , μ_g^f and μ_u^f are relative permeability scalars. One may compare these constitutive relations of a gyrotropic medium of the magnetic type with (40)–(42) for a gyrotropic medium of the electric type (the cold magnetized plasma of Section 5.1). The second component is an isotropic chiral medium [9]:

$$\mathbf{D}(\mathbf{r}, \omega) = \epsilon_0 \epsilon^c(\omega) \mathbf{E}(\mathbf{r}, \omega) + i\sqrt{\epsilon_0 \mu_0} \xi^c(\omega) \mathbf{H}(\mathbf{r}, \omega), \quad (109)$$

$$\mathbf{B}(\mathbf{r}, \omega) = -i\sqrt{\epsilon_0 \mu_0} \xi^c(\omega) \mathbf{E}(\mathbf{r}, \omega) + \mu_0 \mu^c(\omega) \mathbf{H}(\mathbf{r}, \omega), \quad (110)$$

wherein ϵ^c , μ^c and ξ^c are the relative permittivity, relative permeability and the chirality parameter, respectively.³

The ferrite medium and the isotropic chiral medium are then mixed in a certain volumetric proportion and the Bruggeman formalism [78], for example, can be successfully employed to extract the properties of the homogenized composite medium (HCM). Upon usage of the constitutive relations (107)–(110), this procedure yields

$$\mathbf{D}(\mathbf{r}, \omega) = \epsilon_0 \underline{\underline{\epsilon}}^{HCM}(\omega) \cdot \mathbf{E}(\mathbf{r}, \omega) + i\sqrt{\epsilon_0 \mu_0} \underline{\underline{\xi}}^{HCM}(\omega) \cdot \mathbf{H}(\mathbf{r}, \omega), \quad (111)$$

$$\mathbf{B}(\mathbf{r}, \omega) = i\sqrt{\epsilon_0 \mu_0} \underline{\underline{\zeta}}^{HCM}(\omega) \cdot \mathbf{E}(\mathbf{r}, \omega) + \mu_0 \underline{\underline{\mu}}^{HCM}(\omega) \cdot \mathbf{H}(\mathbf{r}, \omega). \quad (112)$$

Each of the four dimensionless constitutive dyadics $\underline{\underline{\epsilon}}^{HCM}$, $\underline{\underline{\mu}}^{HCM}$, $\underline{\underline{\xi}}^{HCM}$ and $\underline{\underline{\zeta}}^{HCM}$ is of the form

$$\underline{\underline{\alpha}}^{HCM}(\omega) = \alpha^{HCM}(\omega)(\underline{\underline{I}} - \mathbf{u}\mathbf{u}) + i\alpha_g^{HCM}(\omega)\mathbf{u} \times \underline{\underline{I}} + \alpha_u^{HCM}(\omega)\mathbf{u}\mathbf{u}, \quad (113)$$

which can be recognized as of gyrotropic type again.

A medium described by (111)–(113) is called a *Faraday chiral medium*; in this instance, a *chiroferrite*. The detailed numerical homogenization results [89] fully confirmed the correct characterization of chiroferrites on purely theoretical grounds [48]. But, of course, these studies provide much more detailed information than just about the basic dyadic structure. They also show, for example, for all choices of the constitutive parameters and all volumetric proportions of the component mediums, that the relation $\underline{\underline{\xi}}^{HCM} = -\underline{\underline{\zeta}}^{HCM}$ holds to high numerical accuracy. That relation is sufficient to fulfill the Post constraint (96). At the same time, the homogenized composite medium is nonreciprocal, which is of course not surprising as one of its component mediums is a ferrite that is itself nonreciprocal.

By following a similar homogenization procedure, whereby the ferrite is replaced by a cold magnetized plasma (as described by (40)–(42)) leads to a Faraday chiral medium, again described by (111)–(113), this time termed *chiroplasma*; see Ref. 90.

It can be seen in a straightforward manner, by comparing the constitutive relations of the ferrite and the isotropic chiral medium with those of the homogenized composite medium, that the latter medium has acquired a higher degree of medium complexity. Needless to say, however, that all constitutive parameters appearing in (111) and (112) are fully determined by the constitutive parameters of the two constituent mediums (supplemented, in this simple example, by only one other parameter, the ratio by which the constituents are mixed to form the composite).

³The constitutive relations of an isotropic chiral medium, given in Tellegen form here, were previously provided in (71) and (72) in Boys–Post form.

8 Concluding remarks

Constitutive relations lie at the very basis of theoretical, numerical and experimental investigations in the electromagnetics of complex mediums. They supplement the macroscopic Maxwell equations (in addition to boundary and/or initial conditions) to provide a self-consistent specification of electromagnetic field problems. Greater complexity in the medium that needs to be characterized is reflected by greater complexity in the constitutive relations which link the primitive fields and the induction fields (or their respective associated phasors). Constitutive relations can be linear or nonlinear, homogeneous or nonhomogeneous, isotropic or anisotropic/bianisotropic, local or nonlocal in space and/or time; their mathematical expression can be in differential, integral or algebraic form. As for their genesis, they may be based on purely phenomenological argumentation, be derived from some fundamental underlying physical principles, or result as the outcome of different types of homogenization formalisms.

In this chapter, a basic introduction into constitutive relations was provided and the reader was presented with a flavor of the difficulties and challenges that need to be addressed through a series of case studies involving different simple and complex mediums. Complex mediums play a very important role in electromagnetics and optics research and the technological opportunities presented by novel materials are vast. The play on the stage of electromagnetics has always been fascinating and will doubtlessly remain so in the future.

References

- [1] A. Lakhtakia, W.S. Weiglhofer and R.F. Messier (eds), *Complex mediums*, Proc SPIE **4097**, Bellingham, WA, USA, 2000.
- [2] A. Lakhtakia, W.S. Weiglhofer and I.J. Hodgkinson (eds), *Complex mediums II: Beyond linear isotropic dielectrics*, Proc SPIE **4467**, Bellingham, WA, USA, 2001.
- [3] A. Lakhtakia, G. Dewar and M.W. McCall (eds), *Complex mediums III: Beyond linear isotropic dielectrics*, Proc SPIE **4806**, Bellingham, WA, USA, 2002.
- [4] M.W. McCall and G. Dewar (eds), *Complex mediums IV: Beyond linear isotropic dielectrics*, Proc SPIE **5218**, Bellingham, WA, USA, 2003 [to be published].
- [5] W.S. Weiglhofer (ed), *Proceedings of Bianisotropics'97, International conference and workshop on electromagnetics of complex media*, Glasgow, UK, 1997.
- [6] A.F. Jacob and J. Reinert (eds), *Proceedings of Bianisotropics'98, 7th international conference on complex media*, Braunschweig, Germany, 1998.
- [7] A.M. Barbosa and A.L. Topa (eds), *Proceedings of Bianisotropics 2000, 8th international conference on electromagnetics of complex media*, Lisbon, Portugal, 2000.
- [8] A. Lakhtakia, V.K. Varadan and V.V. Varadan, *Time-harmonic electromagnetic fields in chiral media*, Springer, Heidelberg, Germany, 1989.

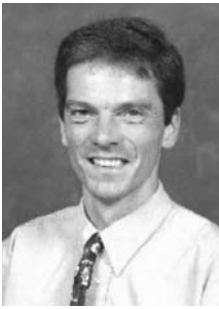
- [9] A. Lakhtakia, *Beltrami fields in chiral media*, World Scientific, Singapore, 1994.
- [10] I.V. Lindell, A.H. Sihvola, S.A. Tretyakov and A.J. Viitanen, *Electromagnetic waves in chiral and bi-isotropic media*, Artech House, London, UK, 1994.
- [11] A. Priou, A. Sihvola, S. Tretyakov and A. Vinogradov, *Advances in complex electromagnetic materials*, Kluwer, Dordrecht, The Netherlands, 1997.
- [12] O.N. Singh and A. Lakhtakia (eds), *Electromagnetic fields in unconventional materials and structures*, Wiley, New York, NY, USA, 2000.
- [13] A. Serdyukov, I. Semchenko, S. Tretyakov and A. Sihvola, *Electromagnetics of bi-anisotropic materials*, Gordon and Breach, Amsterdam, The Netherlands, 2001.
- [14] S. Zouhdi, A. Sihvola and M. Arsalane (eds), *Advances in electromagnetics of complex media and metamaterials*, Kluwer, Dordrecht, The Netherlands, 2003.
- [15] E.J. Post, *Formal structure of electromagnetics*, Dover Press, New York, USA, 1997 (reprinted from North-Holland, Amsterdam, The Netherlands, 1962).
- [16] E.J. Post, Separating field and constitutive equations in electromagnetic theory, *Introduction to complex mediums for optics and electromagnetics* (W.S. Weiglhofer and A. Lakhtakia, eds), SPIE Press, Bellingham, WA, USA, 2003.
- [17] W.E. Thirring, *A course in mathematical physics 2: Classical field theory*, Springer, New York, NY, USA, 1979.
- [18] C. Altman and K. Suchy, *Reciprocity, spatial mapping and time reversal in electromagnetics*, Kluwer, Dordrecht, The Netherlands, 1991.
- [19] I.V. Lindell, A.H. Sihvola and K. Suchy, Six-vector formalism in electromagnetics of bianisotropic media, *J Electromag Waves Appl* **9** (1995), 887–903.
- [20] H.C. Chen, *Theory of electromagnetic waves*, McGraw-Hill, New York, NY, USA, 1983.
- [21] A. Lakhtakia, Covariances and invariances of the Maxwell postulates, *Advanced electromagnetism: Foundations, theory and applications* (T.W. Barrett and D.M. Grimes, eds), World Scientific, Singapore, 1995, 390–410.
- [22] J.D. Jackson, *Classical electrodynamics*, 3rd ed, Wiley, New York, NY, USA, 1999.
- [23] J.A. Kong, *Electromagnetic wave theory*, 2nd ed, EMW Publishing, Cambridge, MA, USA, 2000.
- [24] A. Lakhtakia and W.S. Weiglhofer, Lorentz covariance, Occam's razor, and a constraint on linear constitutive relations, *Phys Lett A* **213** (1996), 107–111; erratum: **222** (1996), 459.
- [25] E.C. Titchmarsh, *Introduction to the theory of Fourier integrals*, 2nd ed, Clarendon Press, Oxford, UK, 1948.
- [26] R.V. Churchill, *Operational mathematics*, 3rd ed, McGraw-Hill, New York, NY, USA, 1972.

- [27] A. Lakhtakia, An electromagnetic trinity from “negative permittivity” and “negative permeability,” *Int J Infrared Millim Waves* **22** (2001), 1731–1734; correctly reprinted: **23** (2002), 813–818.
- [28] W.S. Weiglhofer and A. Lakhtakia, On causality requirements for material media, *Arch Elektron Übertrag* **50** (1996), 389–391.
- [29] J. Van Bladel, *Electromagnetic fields*, Hemisphere, Washington, D.C., USA, 1985.
- [30] M.R. Spiegel, *Advanced calculus*, 2nd ed, McGraw-Hill, New York, NY, USA, 2002.
- [31] P. Hillion, Electromagnetism in a moving chiral medium, *Phys Rev E* **48** (1993), 3060–3065.
- [32] S. Ragusa, Electromagnetic first-order conservation laws in a chiral medium, *J Phys A: Math Gen* **27** (1994), 2887–2890.
- [33] K.M. Flood and D.L. Jaggard, Effective charge densities and current densities in isotropic chiral media, *J Opt Soc Am A* **12** (1995), 177–183.
- [34] P. Hillion, Electromagnetic waves in linear media, *J Phys A: Math Gen* **28** (1995), 2647–2659.
- [35] W.S. Weiglhofer and A. Lakhtakia, Comment on ‘Rigorous solution for transient propagation of electromagnetic waves through a medium: causality plus diffraction in time,’ *Opt Lett* **26** (2001), 1218–1219.
- [36] W.C. Chew, *Waves and fields in inhomogeneous media*, Van Nostrand-Reinhold, New York, NY, USA, 1990.
- [37] A. Karlsson and G. Kristensson, Constitutive relations, dissipation, and reciprocity for the Maxwell equations in the time domain, *J Electromag Waves Appl* **6** (1992), 1517–1535.
- [38] W.S. Weiglhofer and A. Lakhtakia, Causality and natural optical activity (chirality), *J Opt Soc Am A* **13** (1996), 385–386.
- [39] L.B. Felsen and N. Marcuvitz, *Radiation and scattering of waves*, IEEE Press, Piscataway, NJ, USA, 1994 (reprinted from Prentice-Hall, Englewood Cliffs, NJ, USA, 1973).
- [40] D.K. Kalluri, Frequency shifts induced by a time-varying magnetoplasma medium, *Introduction to complex mediums for optics and electromagnetics* (W.S. Weiglhofer and A. Lakhtakia, eds), SPIE Press, Bellingham, WA, USA, 2003.
- [41] A.D. Boardman and M. Xie, Magneto-optics: A critical review, *Introduction to complex mediums for optics and electromagnetics* (W.S. Weiglhofer and A. Lakhtakia, eds), SPIE Press, Bellingham, WA, USA, 2003.
- [42] D. Censor and T. Melamed, A note on Volterra differential constitutive operators and locality considerations in electromagnetic theory, *Complex mediums III: Beyond linear isotropic dielectrics* (A. Lakhtakia, G. Dewar and M.W. McCall, eds), Proc SPIE **4806**, Bellingham, WA, USA, 2002, 81–91.
- [43] W.S. Weiglhofer, A perspective on bianisotropy and *Bianisotropics ’97*, *Int J Appl Electromagn Mech* **9** (1998), 93–101.

- [44] W.S. Weiglhofer, Electromagnetic theory of complex materials, *Engineered nanostructural films and materials* (A. Lakhtakia and R.F. Messier, eds), Proc SPIE **3790**, Bellingham, WA, USA, 1999, 66–76.
- [45] C.F. Bohren, Isotropic chiral materials, *Introduction to complex mediums for optics and electromagnetics* (W.S. Weiglhofer and A. Lakhtakia, eds), SPIE Press, Bellingham, WA, USA, 2003.
- [46] A. Lakhtakia, Chirality, *Wiley Encyclopedia of Electrical and Electronics Engineering, Vol 3*, Wiley, New York, NY, USA, 1999, 307–317.
- [47] A. Lakhtakia, A mini-review on isotropic chiral mediums, *Electromagnetic fields in unconventional materials and structures* (O.N. Singh and A. Lakhtakia, eds), Wiley, New York, NY, USA, 2000, 125–149.
- [48] W.S. Weiglhofer and A. Lakhtakia, The correct constitutive relations of chiroplasmas and chiroferrites, *Microw Opt Technol Lett* **17** (1998), 405–408.
- [49] S. Chandrasekhar, *Liquid crystals*, Cambridge University Press, Cambridge, UK, 1977.
- [50] A. Lakhtakia and W.S. Weiglhofer, Axial propagation in general helicoidal bianisotropic media, *Microw Opt Technol Lett* **14** (1993), 804–806.
- [51] A. Lakhtakia and W.S. Weiglhofer, On light propagation in helicoidal bianisotropic mediums, *Proc Roy Soc Lond A* **448** (1995), 419–437; erratum: **454** (1998), 3275.
- [52] A. Lakhtakia and W.S. Weiglhofer, Further results on light propagation in helicoidal bianisotropic mediums: Oblique propagation, *Proc Roy Soc Lond A* **453** (1997), 93–105; erratum: **454** (1998), 3275.
- [53] A. Lakhtakia, Sculptured thin films: accomplishments and emerging uses, *Mater Sci Engg C* **19** (2002), 424–434.
- [54] A. Lakhtakia and R. Messier, The past, the present, and the future of sculptured thin films, *Introduction to complex mediums for optics and electromagnetics* (W.S. Weiglhofer and A. Lakhtakia, eds), SPIE Press, Bellingham, WA, USA, 2003.
- [55] V.C. Venugopal and A. Lakhtakia, Sculptured thin films: Conception, optical properties, and applications, *Electromagnetic fields in unconventional materials and structures* (O.N. Singh and A. Lakhtakia, eds), Wiley, New York, NY, USA, 2000, 151–216.
- [56] A. Lakhtakia, Batman presents: Applications of sculptured thin films with chiropterous flavor, *Proceedings of Bianisotropics '98, 7th international conference on complex media* (A.F. Jacob and J. Reinert, eds), Braunschweig, Germany, 1998, 297–300.
- [57] A. Lakhtakia, Filamentary, my dear Watson! *Proceedings of Bianisotropics 2000, 8th international conference on electromagnetics of complex media* (A.M. Barbosa and A.L. Topa, eds), Lisbon, Portugal, 2000, 169–174.
- [58] A. Lakhtakia and R.F. Messier (eds), *Engineered nanostructural films and materials*, Proc SPIE **3790**, Bellingham, WA, USA, 1999.
- [59] A. Lakhtakia and W.S. Weiglhofer, Maxwell Garnett formalism for weakly nonlinear, bianisotropic, dilute, particulate composite media, *Int J Electron* **87** (2000), 1401–1408.

- [60] R.W. Boyd, *Nonlinear optics*, Academic Press, San Diego, CA, USA, 1992.
- [61] T. Kobayashi, Introduction to nonlinear optical materials, *Nonlinear Optics* **1** (1991), 91–117.
- [62] S. Kielich and A. Piekara, Frequency- and spatially variable electric and magnetic polarizations induced in nonlinear media by electromagnetic fields, *Acta Phys Pol* **29** (1966), 875–898.
- [63] P.W. Atkins and L.D. Barron, Quantum field theory of optical birefringence phenomena I. Linear and nonlinear optical rotation, *Proc Roy Soc Lond A* **304** (1968), 303–317.
- [64] W.S. Weiglhofer and A. Lakhtakia, Mediation of nonlinear polarization by the magnetic field in a composite medium with a chiral component, *Microw Opt Technol Lett* **13** (1996), 285–287.
- [65] J.M. Arnold, Nonlinear optics using semiconductor quantum wells, *Introduction to complex mediums for optics and electromagnetics* (W.S. Weiglhofer and A. Lakhtakia, eds), SPIE Press, Bellingham, WA, USA, 2003.
- [66] W.S. Weiglhofer, Frequency-dependent dyadic Green functions for bianisotropic media, *Advanced electromagnetism: Foundations, theory and applications* (T.W. Barrett and D.M. Grimes, eds), World Scientific, Singapore, 1995, 376–389.
- [67] C.M. Krowne, Electromagnetic theorems for complex anisotropic media, *IEEE Trans Antennas Propagat* **32** (1984), 1224–1230.
- [68] J.A. Kong, Theorems of bianisotropic media, *Proc IEEE* **60** (1972), 1036–1046.
- [69] A. Lakhtakia and W.S. Weiglhofer, Constraint on linear, homogeneous constitutive relations, *Phys Rev E* **50** (1994), 5017–5019.
- [70] A. Lakhtakia and W.S. Weiglhofer, On a constraint on the electromagnetic constitutive relations of nonhomogeneous linear media, *IMA J Appl Math* **54** (1995), 301–306.
- [71] W.S. Weiglhofer and A. Lakhtakia, The Post constraint revisited, *Arch Elektron Übertrag* **52** (1998), 276–279.
- [72] D.B. Litvin, Point group symmetries, *Introduction to complex mediums for optics and electromagnetics* (W.S. Weiglhofer and A. Lakhtakia, eds), SPIE Press, Bellingham, WA, USA, 2003.
- [73] W.S. Weiglhofer and A. Lakhtakia, On electromagnetic waves in biaxial bianisotropic media, *Electromagnetics* **19** (1999), 351–362.
- [74] T.G. Mackay and W.S. Weiglhofer, Homogenization of biaxial composite materials: Nondissipative dielectric properties, *Electromagnetics* **21** (2001), 15–26.
- [75] T.G. Mackay and W.S. Weiglhofer, Homogenization of biaxial composite materials: Dissipative anisotropic properties, *J Opt A: Pure Appl Opt* **2** (2000), 426–432.
- [76] T.G. Mackay and W.S. Weiglhofer, Homogenization of biaxial composite materials: Bianisotropic properties, *J Opt A: Pure Appl Opt* **3** (2001), 45–52.
- [77] T.G. Mackay and W.S. Weiglhofer, A review of homogenization studies for biaxial bianisotropic materials, *Advances in electromagnetics of complex me-*

- dia and metamaterials* (S. Zouhdi, A. Sihvola and M. Arsalane, eds), Kluwer, Dordrecht, The Netherlands, 2003, 211–228.
- [78] W.S. Weiglhofer, A. Lakhtakia and B. Michel, Maxwell Garnett and Bruggeman formalisms for a particulate composite with bianisotropic host medium, *Microw Opt Technol Lett* **15** (1997), 263–266; erratum: **22** (1999), 221.
 - [79] B. Michel, A. Lakhtakia, W.S. Weiglhofer and T.G. Mackay, Incremental and differential Maxwell Garnett formalisms for bi-anisotropic composites, *Compos Sci Technol* **61** (2001), 13–18.
 - [80] A. Lakhtakia (ed), *Selected papers on linear optical composite materials*, SPIE Press, Bellingham, WA, USA, 1996.
 - [81] B. Michel, Recent developments in the homogenization of linear bianisotropic composite materials, *Electromagnetic fields in unconventional materials and structures* (O.N. Singh and A. Lakhtakia, eds), Wiley, New York, NY, USA, 2000, 39–82.
 - [82] W.S. Weiglhofer, Homogenization of particulate materials, *Complex mediums* (A. Lakhtakia, W.S. Weiglhofer and R. Messier, eds), Proc SPIE **4097**, Bellingham, WA, USA, 2000, 146–154.
 - [83] T.G. Mackay, A. Lakhtakia and W.S. Weiglhofer, Strong-property-fluctuation theory for homogenization of bianisotropic composites: Formulation, *Phys Rev E* **62** (2000), 6052–6064; erratum: **63** (2001), 049901.
 - [84] T.G. Mackay, A. Lakhtakia and W.S. Weiglhofer, Third-order implementation and convergence of the strong-property-fluctuation theory in electromagnetic homogenization, *Phys Rev E* **64** (2000), 066616.
 - [85] T.G. Mackay, A. Lakhtakia and W.S. Weiglhofer, The strong-property-fluctuation theory applied to the homogenisation of linear bianisotropic composites, *Complex mediums II: Beyond linear isotropic dielectrics* (A. Lakhtakia, W.S. Weiglhofer and I.J. Hodgkinson, eds), Proc SPIE **4467**, Bellingham, WA, USA, 2001, 243–255.
 - [86] A. Lakhtakia, Application of strong permittivity fluctuation theory for isotropic, cubically nonlinear, composite mediums, *Opt Commun* **192** (2001), 145–151.
 - [87] T.G. Mackay, A. Lakhtakia and W.S. Weiglhofer, Homogenisation of isotropic, cubically nonlinear, composite mediums by the strong-permittivity-fluctuation theory: third-order considerations, *Opt Commun* **204** (2002), 219–228.
 - [88] T.G. Mackay, Homogenization of linear and nonlinear complex composite materials, *Introduction to complex mediums for optics and electromagnetics* (W.S. Weiglhofer and A. Lakhtakia, eds), SPIE Press, Bellingham, WA, USA, 2003.
 - [89] W.S. Weiglhofer, A. Lakhtakia and B. Michel, On the constitutive parameters of a chiroferrite composite medium, *Microw Opt Technol Lett* **18** (1998), 342–345.
 - [90] W.S. Weiglhofer and T.G. Mackay, Numerical studies on the constitutive parameters of a chiroplasma composite medium, *Arch Elektron Übertrag* **54** (2000), 259–264.



Werner S. Weiglhofer was born in Bruck an der Mur, Austria. He obtained the degrees of Dipl.-Ing. and Dr. Techn. from the Technische Universität Graz, Austria. He joined the University of Glasgow in 1988 as a Research Assistant in the Department of Mathematics, where he became a Professor of Applied Mathematics. He authored/coauthored 135 refereed journal publications and was the author of an undergraduate mathematics textbook. He died in a tragic mountaineering accident on January 12, 2003, in Norway.

Isotropic Chiral Materials

Craig F. Bohren

Abstract

Isotropic chiral materials are rotationally symmetric but not mirror symmetric. Their salient optical property is that only a circularly polarized plane wave can propagate (in any direction) in them without change in polarization. In general, the major axis of an elliptically polarized plane wave changes direction—called optical rotation—and its ratio of minor to major axes increases. Such materials have different wavenumbers for circularly polarized waves of opposite handedness. The equations for arbitrary time-harmonic electromagnetic fields can be cast in a form such that every problem that has ever been solved for linear, isotropic achiral materials (such as reflection and transmission by layers, and scattering by particles) can also be solved for linear, isotropic chiral materials.

1 Introduction

An isotropic chiral medium is rotationally symmetric but not superposable on its mirror image. A simple example is a great jumble of ordinary (right-handed) wood screws. If you rotate this jumble through any angle, you do not sense that anything has changed. Although any given screw changes its orientation, another screw (at least in a statistical sense) takes up the original orientation. But if the jumble is observed in a mirror, all right-handed screws become left-handed.

From the point of view of continuum electromagnetic theory, it is irrelevant whether the screws in a chiral medium are the kind you can fish out of a bin in a hardware store or are of molecular dimensions. Continuum electromagnetic theory (what is almost universally taught, usually without the essential qualifier *continuum*) is based on the approximation—or call it a myth, to be more blunt—that matter is absolutely homogeneous. But matter is composed of atoms, and hence is heterogeneous at least on the atomic if not on larger scales. And yet continuum electromagnetic theory often gives results in good (but not perfect) agreement with experiment. Why? Because absolute homogeneity is not necessary for good agreement, only *optical homogeneity*, i.e., homogeneity on the scale of the wavelength. The wavelengths of visible light are around 500 nm, whereas the molecules in a glass of water are approximately 0.5 nm in diameter and separated by comparable distances. A cube of water one (visible) wavelength on a side therefore contains about 10^9 molecules. The square root of 10^9 is approximately 3×10^4 . We therefore expect, strictly on statistical grounds,¹ that the errors introduced by continuum electromagnetic theory applied to water are of the order of one part in 10^4 . The linear dimensions of wood screws are of the order of a centimeter, so to (legitimately) apply continuum electromagnetic theory to a jumble of such screws simply requires scaling the wavelength upward to perhaps a meter or so.

Continuum electromagnetic theory is also based on another myth, namely that interfaces between different media, at which boundary conditions are imposed, are perfectly smooth. Although no such interfaces exist in nature, if they are optically smooth (smooth on the scale of the wavelength), we can be reasonably confident that continuum electromagnetic theory is an adequate approximation.

It is well to keep in mind the limitations of continuum electromagnetic theory because it *always* predicts results that are not in accord with experiment. Often the discrepancies between theory and experiment are negligible, sometimes a nuisance, but sometimes an opportunity. For example, according to the Fresnel equations, which describe reflection because of an infinite, perfectly smooth planar interface between perfectly homogeneous mediums, reflection occurs in one and only one direction. Yet some people make their living measuring the light reflected in directions for which continuum theory predicts *no* light. If these people had believed in the literal truth of the Fresnel equations they would have missed an opportunity.

¹The ratio of the standard deviation of N values of a variable distributed according to the binomial (Poisson) distribution to the mean of that variable is proportional to the square root of N .

Because chiral materials are so intimately linked with transformations of polarized light, the next section is devoted to such light. Treatments of polarization in physics textbooks are often misleading, almost always incomplete, sometimes downright wrong or incomprehensible. And as one moves farther from physics, into chemistry, biology and geology, what is bad becomes progressively worse. I once surveyed geology textbooks and reference works for discussions of the colored patterns produced when polycrystalline birefringent materials are seen through crossed polarizing filters. What I found was almost entirely gibberish.

If you think that I am exaggerating about the general wretchedness of textbook treatments of polarization, compare the following section with what you will find in a textbook chosen at random, especially in fields such as biology and geology.

2 Polarization: the simple truth

An electric wave of the form

$$\mathbf{E}_0 \exp(ikz - i\omega t), \quad (1)$$

together with its attendant magnetic field, is a possible solution to the Maxwell equations. Here, the vector $\mathbf{E}_0$ is constant in time and lies perpendicular to the z axis (direction of propagation), k is the wavenumber, and ω is the circular frequency. For simplicity, the propagating medium can be taken to be free space (i.e., vacuum). Such a wave is necessarily completely (or 100%) polarized in the sense that, in a fixed plane perpendicular to the z axis, the field traces out in time a definite and fixed ellipse called the *vibration ellipse*.

To show this, set up orthogonal coordinate axes arbitrarily designated as perpendicular ($\perp$) and parallel ($\parallel$) lying in a plane containing the field.² Unit vectors along the orthogonal coordinate axes are denoted by $\mathbf{e}_\perp$ and $\mathbf{e}_\parallel$, and form a right-handed system in the sense that $\mathbf{e}_\perp \times \mathbf{e}_\parallel$ is in the direction of propagation. Without loss of generality we can consider the field in the plane $z = 0$, because we can always shift the origin of time so that the spatial argument in (1) vanishes.

With these basis vectors, the electric field is

$$\mathbf{E} = E_\perp \mathbf{e}_\perp + E_\parallel \mathbf{e}_\parallel, \quad (2)$$

where

$$E_\perp = a_\perp \exp[i(\theta_\perp - \omega t)], \quad (3)$$

$$E_\parallel = a_\parallel \exp[i(\theta_\parallel - \omega t)], \quad (4)$$

and $a_\perp$ and $a_\parallel$ are real-valued amplitudes. Equation (2) gives the complex-valued representation of a real-valued field, and so to determine the curve traced out by

²The terms *vertical* and *horizontal* are best avoided, lest they convey that polarization has something to do with gravity.

this field we need to take the real part:

$$\operatorname{Re}(E_{\perp}) = a_{\perp} \cos(\theta_{\perp} - \omega t), \quad \operatorname{Re}(E_{\parallel}) = a_{\parallel} \cos(\theta_{\parallel} - \omega t). \quad (5)$$

Equations (5) describe a closed curve in the $\operatorname{Re}(E_{\perp})$ – $\operatorname{Re}(E_{\parallel})$ plane. When the phases $\theta_{\perp}$ and $\theta_{\parallel}$ of the two orthogonal components are equal, this curve is a straight line with slope equal to the ratio of amplitudes. When the phases differ by $\pi/2$, the curve is an ellipse with its principal axes aligned along the coordinate axes, where the lengths of the two semi-axes of the ellipse are $a_{\perp}$ and $a_{\parallel}$. A circle, of course, results when the two amplitudes are equal. In general, (5) describes an arbitrarily oriented ellipse of arbitrary *ellipticity* (not to be confused with eccentricity), defined as the ratio of the minor to major axis lengths. If you need a rigorous proof that (5) describes an ellipse, solve it for $\sin \omega t$ and $\cos \omega t$, square and add, and out pops the equation of an arbitrary ellipse.

There is one more *ellipsometric parameter* of the vibration ellipse: its *handedness*, by which is meant the sense in which it is traced out in time. One must be careful when discussing handedness, because there is no universal convention for what is meant by right- and left-handed. As a consequence, what often happens is that parishioners in a particular scientific church assume that all right-thinking people use their convention and hence there is no need to state it. Here I adopt the convention of calling polarized light right-handed if the vibration ellipse, imagined to be viewed looking toward the source, is traced out clockwise. The general state of complete polarization is elliptical, special cases of which are circular and linear.

What makes polarized light worthy of study, indeed the only reason we are even aware of it at all, is that two beams of light identical in all respects except in one or more ellipsometric parameters (azimuth of the major axis relative to a reference direction, ellipticity, and handedness) may interact with matter in observably different ways.

Many textbooks and, even more so, popularizations convey the notion that by polarization is meant linear polarization, no other kind being conceivable. Moreover, to make matters worse, linearly polarized light is sometimes called plane polarized, especially in older works. This is an infelicitous choice of terminology on several grounds. The wave described by (1) is often called a *plane wave*, and if it is also linearly polarized we would then have the awkward (if not ugly) designation *plane plane-polarized wave*, the first plane being the surface of constant phase, the second being either the plane defined by the electric field vector and the direction of propagation or, equivalently, by the plane the electric field vector traces out as it propagates. To be consistent with this designation, we would then have to describe elliptically polarized light as *elliptically helicoidally polarized light* because the electric field vector of such light traces out an elliptical helicoid in space. Yet this is unnecessary because the polarization state of a plane wave is specified completely by the ellipsometric parameters, which have nothing essential to do with surfaces.

My experience has been that people who learned at an impressionable age that light is *plane polarized* find it difficult to understand elliptically polarized light and

nearly impossible to understand partially polarized light. Indeed, they sometimes confuse unpolarized with circularly polarized light. And yet, partially polarized light is easy to understand, beginning with a firm grasp of completely polarized light. The essential property of completely polarized light is complete correlation between two orthogonal components of the electric field vector. They may fluctuate in time, but if they do so synchronously (i.e., the ratio of amplitudes is constant as is the difference in phases), the vibration ellipse has a definite and fixed form. Partially polarized light results when there is partial correlation between the two orthogonal components; unpolarized light results when there is no correlation.

3 Circular birefringence and circular dichroism

Because the defining property of isotropic chiral materials is mirror asymmetry, we expect polarized waves of one handedness in such materials to behave differently in some ways from waves of the opposite handedness. For treating such waves, it is sometimes more convenient to use as basis vectors those corresponding to circular rather than to linear polarization. Recall from Section 2 that the electric field of a circularly polarized plane wave is the superposition of two orthogonal linear fields of equal amplitude but $\pi/2$ out of phase. From this, it should be clear that the (complex) basis vector

$$\mathbf{e}_R = \frac{1}{\sqrt{2}}(\mathbf{e}_{\parallel} + i\mathbf{e}_{\perp}) \quad (6)$$

corresponds to a right-circularly polarized (RCP) plane wave, and

$$\mathbf{e}_L = \frac{1}{\sqrt{2}}(\mathbf{e}_{\parallel} - i\mathbf{e}_{\perp}) \quad (7)$$

corresponds to a left-circularly polarized (LCP) plane wave. These basis vectors are orthonormal in the sense that

$$\mathbf{e}_R \cdot \mathbf{e}_R^* = 1, \quad \mathbf{e}_L \cdot \mathbf{e}_L^* = 1, \quad \mathbf{e}_R \cdot \mathbf{e}_L^* = 0, \quad (8)$$

the asterisk denoting the complex conjugate. The relation between field components in the circular and linear bases can be written succinctly in matrix notation as follows:

$$\begin{pmatrix} E_R \\ E_L \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -i \\ 1 & i \end{pmatrix} \begin{pmatrix} E_{\parallel} \\ E_{\perp} \end{pmatrix}, \quad (9)$$

$$\begin{pmatrix} E_{\parallel} \\ E_{\perp} \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ i & -i \end{pmatrix} \begin{pmatrix} E_R \\ E_L \end{pmatrix}. \quad (10)$$

Suppose that plane waves are propagating (in any direction) in an isotropic chiral medium, with different phase velocities depending on their handedness. This is equivalent to saying that the refractive indexes (strictly, their real parts) are different for RCP and LCP waves. Consider the coherent superposition of a RCP wave and a LCP wave, each with different phase velocities, propagating in a certain direction. At each point along the direction of propagation there is, in general, a phase difference between these two waves. Suppose that this phase difference is zero at $z = 0$ and that the wave is linearly polarized along the perpendicular axis: $E_{\perp} = E$ and $E_{\parallel} = 0$. The corresponding RCP and LCP components are obtained from (9). Propagation results in a phase difference Φ , so that at $z > 0$ the components are

$$E_R = -\frac{i}{\sqrt{2}}E, \quad E_L = \frac{i}{\sqrt{2}}E \exp(i\Phi). \quad (11)$$

Now, using (10), find the corresponding linear components and take the ratio of their real parts:

$$\frac{\text{Re}(E_{\parallel})}{\text{Re}(E_{\perp})} = \frac{\sin \Phi}{1 + \cos \Phi} = \tan \frac{\Phi}{2}. \quad (12)$$

This ratio of field components is the tangent of the angle between the electric field vector and $\mathbf{e}_{\perp}$, and hence the field vector is rotated through an angle equal to one-half the phase shift. This capability of anisotropic chiral medium to rotate the electric field of linearly polarized light is called *optical activity*, and such a medium is said to be *optically active*. These are yet more examples of infelicitous terms. Optical activity, taken by itself, gives no clues to its meaning. Indeed, any nonvacuous medium can be said to be optically active in the sense that it is activated or excited by an external source of light. This is what scattering in all of its manifestations is about. Moreover, a medium is sometimes said to be active if waves propagated in it grow in amplitude in the direction of propagation.

A more transparent term is *circular birefringence*, which not only conveys that the refractive indexes (real parts) are different for RCP and LCP plane waves, but it also betrays its kinship to *linear birefringence* (different refractive indexes for orthogonal linearly polarized waves). Circular birefringence gives rise to *optical rotation*.

If the real parts of the refractive indexes for RCP and LCP plane waves are different, so must be their imaginary parts, and hence their amplitudes must spatially attenuate at different rates. Let us assume, for simplicity, that the phase shift between these two kinds of waves is zero, and take the wave to be linearly polarized along the axis parallel to $\mathbf{e}_{\parallel}$ at $z = 0$. Its two circularly polarized components propagate without change of phase but with change in amplitude. As a consequence of propagation, linearly polarized light is transformed into elliptically polarized light with ellipticity

$$\frac{a_R - a_L}{a_R + a_L}. \quad (13)$$

The ellipticity is zero (i.e., $a_L = a_R = a$) at $z = 0$, but the two amplitudes decrease at different rates:

$$a_R = a \exp\left(-\frac{\alpha_R z}{2}\right), \quad a_L = a \exp\left(-\frac{\alpha_L z}{2}\right). \quad (14)$$

The factor $1/2$ in the exponent is a consequence of taking α to be the absorption coefficient for power (Poynting vector), which is twice the absorption coefficient for amplitude. By combining (13) and (14), the ellipticity becomes

$$\tanh \frac{(\alpha_L - \alpha_R)z}{4} = \tanh \Theta. \quad (15)$$

Note that we have expressed ellipticity as a signed quantity, positive for a right-handed ellipse ($\alpha_R < \alpha_L$), negative for a left-handed ellipse ($\alpha_R > \alpha_L$). In the limit of $z \rightarrow \infty$, light linearly polarized at $z = 0$ becomes completely circularly polarized, although with vanishing amplitude.

A medium for which absorption coefficients (or, equivalently, imaginary parts of the refractive index) are different for circularly polarized waves of opposite handedness is said to be *circularly dichroic* or to exhibit *circular dichroism*. Its cousin is *linear dichroism*: different absorption coefficients for orthogonally polarized linear states of polarization. The sheet polarizing filters in some sunglasses are linearly dichroic: the absorption coefficient for light linearly polarized along one (invisible) axis in the sheet is considerably greater than that along an orthogonal axis (called the transmission axis). Moreover, the magnitude of the absorption coefficients is such that one component is almost entirely transmitted through a thin sheet, the other almost entirely absorbed. The beauty of sheet polarizing filters is that they are compact. But, to the best of my knowledge no one has ever constructed a true sheet circularly polarizing filter. Circularly polarized light is obtained from unpolarized light by transmission through a sandwich composed of a linear polarizing filter and a linear retarder with its (invisible) axes oriented at 45° to those of the filter and with a retardance (phase shift) of $\pi/2$. But, such a sandwich is only a one-way circularly polarizing filter: light transmitted in one direction through the sandwich becomes circularly polarized, but in the opposite direction becomes linearly polarized.

Because phase differences are determined by differences in the real parts of (complex) wave vectors for circularly polarized waves of opposite handedness and amplitude ratios are determined by differences in their imaginary parts, we can summarize optical rotation ϕ and circular dichroism θ as

$$\phi + i\theta = \frac{k_R + k_L}{2}, \quad (16)$$

where k_R and k_L are the complex-valued wavenumbers for RCP and LCP plane waves, respectively. A bit of care is needed in interpreting ϕ and θ . The rate of

change of azimuth with distance along the direction of propagation, or optical rotation per unit distance, is indeed ϕ ; but the rate of change of ellipticity is

$$\frac{d}{dz} \tanh(z\theta), \quad (17)$$

which is (approximately) θ only if $\Theta = \theta z \ll 1$. Rotation is endless, whereas ellipticity must eventually end up as ± 1 .

Optical rotation and circular dichroism are frequency-dependent quantities that are not completely independent, being linked by the Kramers–Kronig (or dispersion) relations

$$\phi(\omega) = \frac{2\omega^2}{\pi} \mathcal{P} \int_0^\infty \frac{\theta(\Omega)}{\Omega(\Omega^2 - \omega^2)} d\Omega, \quad (18)$$

$$\theta(\omega) = -\frac{2\omega}{\pi} \mathcal{P} \int_0^\infty \frac{\phi(\Omega)}{\Omega^2 - \omega^2} d\Omega, \quad (19)$$

where $\mathcal{P}$ indicates the principal part of the integral. With the vast amount of publicity given to helical molecules in living organisms, most notably the double helix of DNA, it should come as no surprise that measurement of optical rotation and circular dichroism spectrums is an important tool in organic chemistry and biochemistry.

We now can state succinctly the salient optical characteristics of an isotropic chiral material. Regardless of direction, the only plane waves that can propagate in such a medium without change of polarization state are circularly polarized of either handedness. An arbitrary elliptically polarized plane wave continuously changes azimuth (i.e., the major axis of its vibration ellipse rotates) and its ellipticity increases. But plane time-harmonic waves are not the only kind of waves allowed by the Maxwell equations. Indeed, strictly speaking, no waves of the form (1) exist (because such a wave occupies all space). So we need to turn to general electromagnetic fields in isotropic chiral materials. Before doing so, however, I digress briefly on the nature of vectors.

4 A digression on vectors

Vectors are usually introduced to students by saying that they are entities with magnitude and direction and depicting them geometrically as arrows. This is a good start but by itself leads nowhere. Equipped solely with this naive view of a vector (in three-dimensional Euclidean space), one would be hard-pressed to extend the vector concept to spaces of higher dimensions and to tensors.

The symbol **A** stands for a single entity (vector) with an infinite number of representations, one for each possible coordinate system. This implies that the representation in one coordinate system must be related to that in all others. Indeed,

how the representations of a mathematical entity change under coordinate transformations determines whether it is or is not a vector.

Consider the ordered set of numbers $\{A_\ell(x)\}$, $(\ell = 1, 2, \dots, m)$, defined in a coordinate system x (with coordinates $x^1, x^2, \dots, x^m$). Suppose that we transform to an arbitrary (within reason) coordinate system y :

$$y^\ell = y^\ell(x^1, x^2, \dots, x^m). \quad (20)$$

If the corresponding ordered set of numbers $\{B_\ell(y)\}$ in this coordinate system is given by

$$B_\ell(y) = \sum_{j=1}^m \frac{\partial x^j}{\partial y^\ell} A_j(x), \quad (\ell = 1, 2, \dots, m), \quad (21)$$

then the entire set of quantities $\{A(x), B(y), \dots\}$ defines a *covariant* vector; here, $A(x)$ denotes the set $\{A_\ell(x)\}$, etc. If the ordered set $\{B_\ell(y)\}$, is given by

$$B_\ell(y) = \sum_{j=1}^m \frac{\partial y^\ell}{\partial x^j} A_j(x), \quad (\ell = 1, 2, \dots, m), \quad (22)$$

the entire set of quantities $\{A(x), B(y), \dots\}$ defines a *contravariant* vector. The distinction between covariant and contravariant disappears if the coordinate transformation is orthogonal.

Now, the cross product of two vectors, often passed off as a vector, is not strictly a vector because it does not have the proper transformation properties under reflections of coordinate systems. To show this, consider the cross product $\mathbf{c} = \mathbf{a} \times \mathbf{b}$. The components of $\mathbf{a}$ in a rectangular Cartesian coordinate system are (a_1, a_2, a_3) , those of $\mathbf{b}$ are (b_1, b_2, b_3) , and according to the rule for forming cross products those of $\mathbf{c}$ are

$$c_1 = a_2 b_3 - a_3 b_2, \quad c_2 = a_3 b_1 - a_1 b_3, \quad c_3 = a_1 b_2 - a_2 b_1. \quad (23)$$

Now suppose that we transform to a reflected coordinate system (denoted by prime): two coordinates are equal but one is changed to its negative. According to the transformation law (21) or (22), the components of $\mathbf{a}$ and $\mathbf{b}$ in the reflected coordinate system are

$$\begin{aligned} a'_1 &= a_1, \quad a'_2 = a_2, \quad a'_3 = -a_3, \\ b'_1 &= b_1, \quad b'_2 = b_2, \quad b'_3 = -b_3. \end{aligned} \quad (24)$$

If we substitute these transformed components in (23), we obtain the components of $\mathbf{c}$ in the reflected coordinate system as

$$c'_1 = -c_1, \quad c'_2 = -c_2, \quad c'_3 = c_3. \quad (25)$$

But wait! According to the transformation law for vectors the components of $\mathbf{c}$ in the reflected coordinate system should have been

$$c'_1 = c_1, \quad c'_2 = c_2, \quad c'_3 = -c_3, \quad (26)$$

just as in (24). Thus the cross product is not a true vector, and indeed is sometimes called a *pseudovector*.

All is well with the cross product unless reflections inadvertently or intentionally enter our analysis. Then we have to be careful. We never get into trouble if we confine ourselves to rotations of coordinate systems or to transformations to orthogonal coordinate systems. Given the inherent mirror asymmetry of the cross product, we might expect to see it playing a role in the electrodynamics of isotropic chiral materials. This expectation is borne out in the following section.

5 Electromagnetic fields in a chiral material

With the assumption of harmonic time-dependence $\exp(-i\omega t)$, the Maxwell equations can be written as

$$\nabla \cdot \left(\mathbf{D} + \frac{i}{\omega} \mathbf{J} \right) = 0, \quad (27)$$

$$\nabla \cdot \mathbf{B} = 0, \quad (28)$$

$$\nabla \times \mathbf{E} - i\omega \mathbf{B} = \mathbf{0}, \quad (29)$$

$$\nabla \times \mathbf{H} + i\omega \left(\mathbf{D} + \frac{i}{\omega} \mathbf{J} \right) = \mathbf{0}. \quad (30)$$

The vectors in these equations are phasors that may be interpreted as the temporal Fourier transforms of the actual fields. These field equations are cast in a form slightly different from that usually encountered in order to emphasize that the electric displacement $\mathbf{D}$ and current density $\mathbf{J}$ are not separable: it is not possible to distinguish between the real (imaginary) part of $\mathbf{D}$ and the imaginary (real) part of $\mathbf{J}$, because only their sums appear in the field equations. Mostly for historical reasons separate constitutive relations between $\mathbf{D}$ and $\mathbf{E}$ and between $\mathbf{J}$ and $\mathbf{E}$ usually are assumed, but this is not necessary.³

To proceed, we need constitutive relations for a (linear) isotropic chiral material. Such relations must reduce to the usual constitutive relations in the limit of no chirality and must somehow incorporate mirror asymmetry. The simplest set of such relations is the Drude–Born–Fedorov relations:

$$\mathbf{D} = \epsilon \mathbf{E} + \epsilon \beta \nabla \times \mathbf{E}, \quad \mathbf{B} = \mu \mathbf{H} + \mu \beta \nabla \times \mathbf{H}, \quad (31)$$

³The relation between $\mathbf{J}$ and $\mathbf{E}$ is called Ohm's law.

where the constitutive parameters ϵ , μ and β are frequency-dependent phenomenological coefficients that can be determined by measurements or by appeal to a higher-order (molecular) theory. Although only $\mathbf{D}$ appears in these relations, it should be interpreted as the linear combination of $\mathbf{D}$ and $\mathbf{J}$ in (27) and (29). When the chirality parameter β vanishes, these constitutive relations reduce to familiar ones for isotropic dielectric-magnetic materials. And the appearance of the cross product (curl) should come as no surprise, given the discussion in Section 4 about the mirror asymmetry of cross products.⁴

Equations (27)–(31) can be combined into a set of equations in which the electromagnetic fields $\mathbf{E}$ and $\mathbf{H}$ appear together as a single entity:

$$\nabla \times \begin{pmatrix} \mathbf{E} \\ \mathbf{H} \end{pmatrix} = \underline{\underline{\tilde{K}}} \cdot \begin{pmatrix} \mathbf{E} \\ \mathbf{H} \end{pmatrix}, \quad \nabla \cdot \begin{pmatrix} \mathbf{E} \\ \mathbf{H} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad (32)$$

where

$$\underline{\underline{\tilde{K}}} = \frac{i\omega}{1 - \beta^2\epsilon\mu\omega^2} \begin{pmatrix} -i\beta\epsilon\mu\omega \underline{\underline{I}} & \mu \underline{\underline{I}} \\ -\epsilon \underline{\underline{I}} & -i\beta\epsilon\mu\omega \underline{\underline{I}} \end{pmatrix} \quad (33)$$

is a 6×6 matrix (such objects are denoted in this chapter by double underlining and a tilde symbol) while $\underline{\underline{I}}$ is the standard 3×3 identity matrix. Equations (32) can be manipulated to yield

$$\nabla^2 \begin{pmatrix} \mathbf{E} \\ \mathbf{H} \end{pmatrix} + \underline{\underline{\tilde{K}}}^2 \cdot \begin{pmatrix} \mathbf{E} \\ \mathbf{H} \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \quad (34)$$

which has the familiar form of the Helmholtz equation.

The transformation

$$\begin{pmatrix} \mathbf{E} \\ \mathbf{H} \end{pmatrix} = \underline{\underline{\tilde{A}}} \cdot \begin{pmatrix} \mathbf{Q}_L \\ \mathbf{Q}_R \end{pmatrix} \quad (35)$$

of the electromagnetic field diagonalizes $\underline{\underline{\tilde{K}}}$:

$$\underline{\underline{\tilde{A}}} = \underline{\underline{\tilde{A}}}^{-1} \cdot \underline{\underline{\tilde{K}}} \cdot \underline{\underline{\tilde{A}}}, \quad (36)$$

where

$$\underline{\underline{\tilde{A}}} = \begin{pmatrix} k_L \underline{\underline{I}} & \underline{\underline{0}} \\ \underline{\underline{0}} & -k_R \underline{\underline{I}} \end{pmatrix}, \quad (37)$$

⁴These constitutive equations are also consistent with Lorentz-reciprocity; i.e., no change is recorded if the positions of a transmitter and a receiver inside a space occupied by a homogeneous, isotropic chiral material are interchanged.

$$\underline{\tilde{A}} = \begin{pmatrix} \underline{I} & -i\sqrt{\mu/\epsilon}\underline{I} \\ -i\sqrt{\epsilon/\mu}\underline{I} & \underline{I} \end{pmatrix}. \quad (38)$$

The wavenumbers k_R and k_L are related to the constitutive parameters by

$$\beta = \frac{1}{2} \left(\frac{1}{k_R} - \frac{1}{k_L} \right), \quad \omega\sqrt{\epsilon\mu} = \frac{2}{(1/k_R) + (1/k_L)}; \quad (39)$$

equivalently,

$$k_R = \frac{\omega\sqrt{\epsilon\mu}}{1 + \beta\omega\sqrt{\epsilon\mu}}, \quad k_L = \frac{\omega\sqrt{\epsilon\mu}}{1 - \beta\omega\sqrt{\epsilon\mu}}. \quad (40)$$

The fields $\mathbf{Q}_L$ and $\mathbf{Q}_R$ independently satisfy the equations

$$\nabla \times \mathbf{Q} = q \mathbf{Q}, \quad \nabla \cdot \mathbf{Q} = 0, \quad \nabla^2 \mathbf{Q} + q^2 \mathbf{Q} = \mathbf{0}, \quad (41)$$

where $q = k_L$ for $\mathbf{Q} = \mathbf{Q}_L$, and $q = -k_R$ for $\mathbf{Q} = \mathbf{Q}_R$.

To show that $\mathbf{Q}_L$ can be looked upon as a left-handed field and $\mathbf{Q}_R$ as a right-handed field, consider a plane wave propagating in a direction parallel to the unit vector $\mathbf{e}$. From (41) we have $i\mathbf{e} \times \mathbf{Q} = \pm \mathbf{Q}$, where the plus sign holds when $\mathbf{Q} = \mathbf{Q}_L$ and the minus sign when $\mathbf{Q} = \mathbf{Q}_R$. In addition, the transversality condition $\mathbf{e} \cdot \mathbf{Q} = 0$ emerges from (41). We can therefore write

$$\mathbf{Q} = Q_{\parallel} \mathbf{e}_{\parallel} + Q_{\perp} \mathbf{e}_{\perp}, \quad (42)$$

where $\mathbf{e} \times \mathbf{e}_{\perp} = \mathbf{e}_{\parallel}$ and $\mathbf{e} \times \mathbf{e}_{\parallel} = -\mathbf{e}_{\perp}$. This yields

$$\mathbf{Q} = Q_{\parallel} (\mathbf{e}_{\parallel} \pm i\mathbf{e}_{\perp}). \quad (43)$$

From (6) and (7), it follows that $\mathbf{Q}_L$ and $\mathbf{Q}_R$ are LCP and RCP plane waves. Note that (32)–(41) are true not only for plane waves, but also for any time-harmonic electromagnetic fields in an isotropic chiral material.

The determination of electromagnetic fields in linear, isotropic achiral materials ($\beta = 0$) could begin with (41): solve these equations to obtain $\mathbf{Q}_L$, substitute $-q$ for q in this solution to obtain $\mathbf{Q}_R$, obtain electromagnetic fields ($\mathbf{E}$, $\mathbf{H}$) from (35), and require the tangential components of these fields to be continuous across interfaces between dissimilar mediums. If one or more materials are chiral, substitute $-k_R$ for k_L in solutions to (41), which is no more difficult than substituting $-q$ for q . We therefore arrive at what appears to be a startling realization but is in fact merely a statement of the obvious: *Every single time-harmonic problem in electromagnetic theory that has ever been solved for linear, isotropic, achiral materials can also be solved for linear, isotropic, chiral materials.* This includes reflection because of interfaces, reflection and transmission by layers, scattering by particles

of any shape, and so on *ad infinitum*. So there is no need to solve from scratch many problems for isotropic chiral materials. All we need do is rearrange existing solutions for achiral materials and examine the consequences. But rather than do what is not much more than bookkeeping, let us take a global view of the expected differences between chiral and achiral materials.

Fundamentally, the only difference is that a chiral material has two complex-valued refractive indexes, one for each handedness. Thus, the magnitude of any difference between a chiral material and its achiral analog must depend on the magnitude of the difference between these two refractive indexes. From (16) it follows that the total optical rotation of a linearly polarized wave transmitted a distance h is

$$\phi = \frac{\pi}{\lambda_0} (n_L - n_R) h, \quad (44)$$

where $n_R = \text{Re}(k_R/k_0)$ and $n_L = \text{Re}(k_L/k_0)$ are the respective refractive indexes (real parts) for RCP and LCP plane waves, λ_0 is the free-space wavelength, and $k_0 = 2\pi/\lambda_0$ is the free-space wavenumber. For sake of argument take the wavelength to be in the visible part of the electromagnetic spectrum and $h = 1$ cm, the size of a laboratory cuvette. Measuring rotations of a few degrees is not a heroic task and requires a refractive index difference of only about one part in a million. Thus, a tiny difference in refractive indexes for RCP and LCP plane waves can lead to readily observable consequences.

But note that the foregoing discussion is essentially *qualitative*, a change in the nature of an electromagnetic field rather than its magnitude because of chirality. What about *quantitative* changes? The simplest way to answer this is to consider reflection because of a smooth interface between an achiral medium (take it to be negligibly absorbing for convenience) with refractive index n and air. The normal-incidence reflectivity R (for power) is given by the Fresnel equation

$$R = \left| \frac{n - 1}{n + 1} \right|^2. \quad (45)$$

Differentiate both sides of this equation with respect to n and approximate derivatives by ratios of differences to obtain

$$\frac{\Delta R}{R} \approx \frac{4\Delta n}{n^2 - 1}. \quad (46)$$

Except near $n = 1$ (when $R = 0$), a small change in n yields a small relative change in R . On physical grounds we expect the same to hold for chiral materials: small differences in refractive indexes for the two circular polarization states yield small changes in the magnitudes of fields. And small differences are the norm. One can easily postulate hypothetical chiral materials with huge refractive index differences, but finding them in nature or making them is not so easy. Thus, making

an antenna or a reflector or an absorber or a cavity out of a chiral material is unlikely to yield anything greatly different. Chiral materials are fascinating, but they are not magic.

6 Essential reading

For a sweeping survey of the roles that mirror asymmetry plays in a surprising variety of scientific fields, written at the popular level by a master of exposition, see Gardner [1].

Scattering by rough surfaces (which often are exceedingly smooth) is treated by Stover [2] as well as Bennet and Mattsson [3].

Two excellent treatises on polarized light were written by Shurcliff [4] and Clarke and Grainger [5]. Facsimile reproductions of seminal papers on polarization and its application are available in two anthologies compiled by Billings [6-7].

Optical rotation and circular dichroism from the viewpoint of chemists is discussed by Lowry [8], Djerassi [9], and Charney [10].

The digression on vectors in Section 4 relies heavily on Sokolnikoff's classic textbook [11].

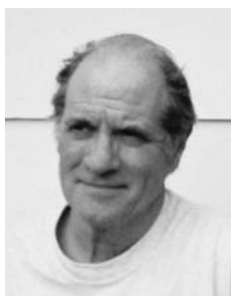
The treatment of electromagnetic fields in isotropic chiral materials in Section 5 relies heavily on Chapter 8 of Bohren and Huffman [12]. A book devoted exclusively to electromagnetic fields in chiral materials was written by Lakhtakia [13], who also published a review in 2000 [14]. For an anthology of papers on chirality, including seminal papers by Pasteur and other notables, see Lakhtakia [15].

Constitutive equations of linear materials are discussed at length in this volume by Post [16] and Weiglhofer [17]. Complex materials are often realized as composite materials, which can be treated as effectively homogeneous materials under certain conditions [18]. Certain materials, such as some sculptured thin films [19-20], are chiral at macroscopic length scales. These materials are macroscopically inhomogeneous and anisotropic, and must be distinguished from the isotropic chiral materials characterized by (31).

References

- [1] M. Gardner, *The new ambidextrous universe: symmetry and asymmetry from mirror reflections to superstrings*, 3rd ed, Freeman, New York, NY, USA, 1990.
- [2] J.C. Stover, *Optical scattering: measurement and analysis*, 2nd ed, SPIE Press, Bellingham, WA, USA, 1995.
- [3] J.M. Bennet and L. Mattsson, *Introduction to surface roughness and scattering*, 2nd ed, Optical Society of America, Washington, DC, USA, 1999.
- [4] W.A. Shurcliffe, *Polarized light*, Harvard University Press, Cambridge, MA, USA, 1962.
- [5] D. Clarke and J.F. Grainger, *Polarized light and optical measurements*, Pergamon Press, Oxford, UK, 1971.

- [6] B.H. Billings (ed), *Selected papers on polarization*, SPIE Press, Bellingham, WA, USA, 1990.
- [7] B.H. Billings (ed), *Selected papers on applications of polarized light*, SPIE Press, Bellingham, WA, USA, 1992.
- [8] T.M. Lowry, *Optical rotatory power*, Dover Press, New York, NY, USA, 1964 (reprinted from Longman, Greens, London, UK, 1935).
- [9] C. Djerassi, *Optical rotatory dispersion*, McGraw-Hill, New York, NY, USA, 1960.
- [10] E. Charney, *The molecular basis of optical activity: Optical rotatory dispersion and circular dichroism*, Krieger, Malabar, FL, USA, 1985 (reprinted from Wiley, New York, NY, USA, 1979).
- [11] I.S. Sokolnikoff, *Tensor analysis: theory and applications to geometry and mechanics of continua*, 2nd ed, Wiley, New York, NY, USA, 1964.
- [12] C.F. Bohren and D.R. Huffman, *Absorption and scattering of light by small particles*, Wiley, New York, NY, USA, 1983.
- [13] A. Lakhtakia, *Beltrami fields in chiral media*, World Scientific, Singapore, 1994. [A Beltrami field, in case you are wondering, is one that satisfies $\nabla \times \mathbf{Q} = q \mathbf{Q}$.]
- [14] A. Lakhtakia, A mini-review on isotropic chiral mediums, *Electromagnetic fields in unconventional materials and structures* (O.N. Singh and A. Lakhtakia, eds), Wiley, New York, NY, USA, 2000.
- [15] A. Lakhtakia (ed), *Selected papers on natural optical activity*, SPIE Press, Bellingham, WA, USA, 1996.
- [16] E.J. Post, Separating field and constitutive equations in electromagnetic theory, *Introduction to complex mediums for optics and electromagnetics* (W.S. Weiglhofer and A. Lakhtakia, eds), SPIE Press, Bellingham, WA, USA, 2003.
- [17] W.S. Weiglhofer, Constitutive characterization of simple and complex mediums, *Introduction to complex mediums for optics and electromagnetics* (W.S. Weiglhofer and A. Lakhtakia, eds), SPIE Press, Bellingham, WA, USA, 2003.
- [18] T.G. Mackay, Homogenization of linear and nonlinear complex composite materials, *Introduction to complex mediums for optics and electromagnetics* (W.S. Weiglhofer and A. Lakhtakia, eds), SPIE Press, Bellingham, WA, USA, 2003.
- [19] A. Lakhtakia and R. Messier, The past, the present, and the future of sculptured thin films, *Introduction to complex mediums for optics and electromagnetics* (W.S. Weiglhofer and A. Lakhtakia, eds), SPIE Press, Bellingham, WA, USA, 2003.
- [20] M.W. McCall, Towards optoelectronic applications of chiral sculptured thin films, *Introduction to complex mediums for optics and electromagnetics* (W.S. Weiglhofer and A. Lakhtakia, eds), SPIE Press, Bellingham, WA, USA, 2003.



Craig F. Bohren is Distinguished Professor Emeritus of Meteorology at The Pennsylvania State University. He received a doctorate in physics from the University of Arizona and is the coauthor (with Donald R. Huffman) of *Absorption and Scattering of Light by Small Particles* as well as the author of two popular scientific books, *Clouds in a Glass of Beer* (for which he received the first Louis J. Battan Author's Award of the American Meteorological Society) and *What Light Through Yonder Window Breaks?*

He edited a Milestone Volume entitled *Selected Papers on Atmospheric Scattering* for SPIE; and he contributed the chapter on scattering by particles to the *Handbook of Optics* and the chapter on atmospheric optics to the *Encyclopedia of Applied Physics*. His most recent book is *Atmospheric Thermodynamics* (coauthored with Bruce A. Albrecht). In 1988, he was elected a Fellow of the Optical Society of America. During the academic year 1986–87, he was a Visiting Professor of Physics and Astronomy at Dartmouth College, Hanover, New Hampshire, USA; in 1993, he was the Selby Fellow of the Australian Academy of Sciences, and in 1994 was the Mitchell Distinguished Visiting Professor in the Physics Department at Trinity University, San Antonio, Texas, USA.

Point Group Symmetries

Daniel B. Litvin

Abstract

The form of macroscopic physical property tensors of a crystalline structure can be determined from its magnetic or nonmagnetic point group symmetry. A ferroic crystal contains two or more equally stable domains of the same structure but of different spatial orientation. Macroscopic tensorial physical properties that are different in the domains provide a tensor distinction of the domains. The use of point group symmetries in this tensor distinction is reviewed in this chapter: Point group symmetry-based classifications of domains have been defined to determine if specific macroscopic tensorial physical properties can provide a tensor distinction of all or some domains that arise in a phase transition. For pairs of domains, the tensor distinction is determined from a point group symmetry relationship, called a twin law. Recent work on domain average engineering in ferroics that focuses on the averaged point group symmetry and averaged physical properties of subsets of domains is also discussed.

1 Point groups

Traditionally, a crystalline medium is defined as a solid medium whose macroscopic physical properties can be characterized by associating with it a symmetry group that belongs to one of the 32 types of nonmagnetic crystallographic point groups [1]. If one considers only the purely geometrical characterization of a crystalline medium, the medium can also be considered trivially invariant under time inversion, which we denote by $1'$. It follows that if a medium is invariant under an element R of a nonmagnetic crystallographic point group $\mathbf{R}$, it is then also invariant under the product of an element R and time inversion $1'$. The product $R1'$ is denoted by R' . From a purely geometrical point of view, one might be tempted to then say that the medium and its physical properties are invariant under the group $\mathbf{R}1'$, the group consisting of all the elements R of $\mathbf{R}$ and the elements R of $\mathbf{R}$ multiplied by time inversion, a view taken up by Zocher and Török [2]. However, this leads to the erroneous prediction that the linear magnetoelectric effect is impossible in all crystalline media (see [3] for a history of the belief in the impossibility of this effect prior to its discovery in 1960).

Landau and Lifshitz [4] stressed that the absence of time inversion symmetry is a necessary condition for the existence of magnetically ordered crystalline materials, as for example, ferromagnetic and antiferromagnetic mediums. For such materials, the medium and its physical properties are not invariant under a point group $\mathbf{R}1'$. Dyaloshinski [5], using the magnetic crystallographic point group of antiferromagnetic Cr_2O_3 , correctly predicted the existence of the magnetoelectric effect in this material, an effect that was experimentally verified soon after [6-7].

Consequently, it is not just the 32 types of nonmagnetic crystallographic point groups $\mathbf{R}$ and the 32 types of groups $\mathbf{R}1'$ that one needs to consider when predicting the physical properties of crystalline mediums. Instead, it is the 122 types of magnetic crystallographic point groups that are to be used. These groups have been classified into *magnetic superfamilies* in the following manner [8]: Let $\mathbf{R}$ denote one of the 32 types of nonmagnetic crystallographic point groups. The magnetic superfamily of crystallographic groups of type $\mathbf{R}$ consists of

- Groups of the type $\mathbf{R}$.
- Groups of the type $\mathbf{R}1'$, where the time inversion group $1'$ consists of the identity 1 and time inversion $1'$.
- Groups of the type $\mathbf{R}(\mathbf{D}) = \mathbf{D} + (\mathbf{R} - \mathbf{D})1'$, where $\mathbf{D}$ is a subgroup of index two of $\mathbf{R}$; i.e., $\mathbf{D}$ is a group containing one-half the number of elements of $\mathbf{R}$, and the remaining elements $\mathbf{R} - \mathbf{D}$ of $\mathbf{R}$ are each combined with time inversion $1'$.

For example, the magnetic superfamily of $\mathbf{R} = 2/\mathbf{m}$ consists of

- Groups of the type $2/\mathbf{m} = \{1, 2, \bar{1}, m\} = 2/\mathbf{m}$
- Groups of the type $2/\mathbf{m}1' = \{1, 2, \bar{1}, m, 1', 2', \bar{1}', m'\} = 2/\mathbf{m}1'$

- Groups of the type $2/\mathbf{m}(2) = \{1, 2, \bar{1}', m'\} = 2/\mathbf{m}'$
 $2/\mathbf{m}(\bar{1}) = \{1, 2', \bar{1}, m'\} = 2'/\mathbf{m}'$
 $2/\mathbf{m}(\mathbf{m}) = \{1, 2', \bar{1}', m\} = 2'/\mathbf{m}$

On the extreme right of this tabulation are the symbols of these types of groups where the $\mathbf{R}(\mathbf{D})$ notation has been replaced with a *primed* notation.

A complete listing of the 122 types of magnetic crystallographic point groups in groupings of magnetic superfamilies is given in Table 1. How to interpret the meaning of the point group symbols is given in Appendix A. A computerized tabulation of group theoretical properties of the magnetic crystallographic point groups has recently been published [9]. The notation used for the point groups in Table 1

Table 1 The 122 types of magnetic crystallographic point groups. The first column lists the 32 types $\mathbf{R}$ of nonmagnetic point groups and the second, the 32 types $\mathbf{R}1'$. The remaining columns list the remaining 58 types $\mathbf{R}(\mathbf{D})$ of magnetic point groups in primed notation.

1	11'			
$\bar{1}$	$\bar{1}1'$	$\bar{1}'$		
2	21'	2'		
$\mathbf{m}$	$\mathbf{m}1'$	$\mathbf{m}'$		
2/m	2/m1'	2'/m	2/m'	2'/m'
222	2221'	2'2'2		
mm2	mm21'	m'm2'	m'm'2	
mmm	mmm1'	m'mm	m'm'm	m'm'm'
4	41'	4'		
$\bar{4}$	$\bar{4}1'$	$\bar{4}'$		
4/m	4/m1'	4'/m	4/m'	4'/m'
422	4221'	4'22'	42'2'	
4mm	4mm1'	4'm'm	4m'm'	
$\bar{4}2\mathbf{m}$	$\bar{4}2\mathbf{m}1'$	$\bar{4}'2'\mathbf{m}$	$\bar{4}'2\mathbf{m}'$	$\bar{4}2'\mathbf{m}'$
4/mmm	4/mmm1'	4/m'mmm	4'/mmm'm	4'/m'm'm
		4/mmm'm'	4/m'm'm'	
3	31'			
$\bar{3}$	$\bar{3}1'$	$\bar{3}'$		
32	321'	32'		
3m	3m1'	3m'		
$\bar{3}\mathbf{m}$	$\bar{3}\mathbf{m}1'$	$\bar{3}'\mathbf{m}$	$\bar{3}'\mathbf{m}'$	$\bar{3}\mathbf{m}'$
6	61'	6'		
$\bar{6}$	$\bar{6}1'$	$\bar{6}'$		
6/m	6/m1'	6'/m	6/m'	6'/m'
622	6221'	6'2'2	62'2'	
6mm	6mm1'	6'm'm	6m'm'	
6m2	6m21'	6'm'2	6'm2'	$\bar{6}\mathbf{m}'2'$
6/mmm	6/mmm1'	6/m'mmm	6'/mmm'm	6'/m'm'm
		6/mmm'm'	6/m'm'm'	
23	231'			
$\mathbf{m}\bar{3}$	$\mathbf{m}\bar{3}1'$	$\mathbf{m}'\bar{3}'$		
432	4321'	4'32'		
$\bar{4}3\mathbf{m}$	$\bar{4}3\mathbf{m}1'$	$\bar{4}'3\mathbf{m}'$		
$\mathbf{m}\bar{3}\mathbf{m}$	$\mathbf{m}\bar{3}\mathbf{m}1'$	$\mathbf{m}'\bar{3}'\mathbf{m}$	$\mathbf{m}\bar{3}\mathbf{m}'$	$\mathbf{m}'\bar{3}'\mathbf{m}'$

and throughout this paper is referred to as the *short international* notation. Note that we conform to the newest version of the International Tables for Crystallography [10], where symbols $\mathbf{m3}$ and $\mathbf{m3m}$ have been replaced by $\mathbf{m}\bar{3}$ and $\mathbf{m}\bar{3}\mathbf{m}$. Other notations exist, e.g., the group type denoted by $2/\mathbf{m}(\mathbf{m}) = 2'/\mathbf{m}$ is denoted by $\mathbf{C}_{2h}$ ($\mathbf{C}_s$) in Schonflies notation [11] and $\underline{2}:\mathbf{m}$ in Shubnikov notation [12].

2 Physical property tensors

The derivation and tabulation of physical property tensors invariant under the magnetic and nonmagnetic crystallographic point groups have been considered by many authors [13]–[22] and references contained in these sources. Tables of a wide variety of physical property tensors invariant under nonmagnetic crystallographic point groups have been given in the literature [23–24] and will, for both magnetic and nonmagnetic point groups, appear in the forthcoming Volume D of the *International Tables for Crystallography* [25].

The form of the physical property tensors invariant under magnetic crystallographic point groups can be derived from the existing tables of physical property tensors invariant under nonmagnetic crystallographic point groups: Let $\mathbf{V}$ denote a polar vector and $V^n = V \times V \times \dots \times V$ the n th-ranked product of $\mathbf{V}$, and let e and a denote zero-rank tensors that change sign under spatial inversion $\bar{1}$ and time inversion $1'$, respectively. It has been shown [26] that the form of a physical property tensor transforming as a tensor V_n , aV_n , eV^n , or aeV^n invariant under a magnetic point group $\mathbf{M}$ is the same as the form of a physical property tensor transforming as a tensor V^n or eV^n invariant under a related nonmagnetic point group $\mathbf{R}$. Detailed tables listing the groups $\mathbf{R}$ corresponding to all magnetic point groups $\mathbf{M}$ and all tensor types have been tabulated [26].

From the Taylor expansion of the density of stored free enthalpy, the polarization P , magnetization M , and mechanical deformation s can be written in a third-order expansion in terms of the electric field E , magnetic field H , and stress tensor L as [27]:

$$P_i = \kappa_i^o + \kappa_{ij}E_j + \frac{1}{2}\kappa_{ijk}E_jE_k + \alpha_{ij}H_j + \alpha_{ijk}H_jE_k + \frac{1}{2}\beta_{ijk}H_jH_k + d_{ijk}L_{jk}, \quad (1)$$

$$M_i = \chi_i^o + \chi_{ij}H_j + \frac{1}{2}\chi_{ijk}H_jH_k + \alpha_{ij}E_j + \beta_{ijk}E_jH_k + \frac{1}{2}\alpha_{ijk}E_jE_k + g_{ijk}L_{jk}, \quad (2)$$

$$s_{ij} = \mathbf{s}_{ij}^o + d_{ijk}E_k + g_{ijk}H_k, \quad (3)$$

where (1)–(3) have been written with the summation convention, i.e., summation is implied if, in the same term, an index appears twice. In Table 2, the name of each coefficient, corresponding phenomenon, and tensor type are given. A survey of these and higher-order magnetoelectric effects is provided in the chapter of Schmid [28].

Table 2 Physical property tensors.

Physical property	Tensor type	Symbol	Phenomenons
spontaneous polarization	V	κ_i^o	pyro-, ferro-, ferrielectricity
spontaneous magnetization	aeV	χ_i^o	pyro-, ferro-, ferrimagnetism
spontaneous deformation	[V ²]	s_{ij}^o	pyro-, ferroelasticity
electric susceptibility	[V ²]	κ_{ij}	induced polarization, Brillouin, Raman, Raleigh scattering
magnetic susceptibility	[V ²]	χ_{ij}	induced magnetization
magnetoelectric susceptibility	aeV ²	α_{ij}	magnetoelectric effect
piezoelectric coefficient	V[V ²]	d_{ijk}	piezoelectricity
piezomagnetic coefficient	aeV[V ²]	g_{ijk}	piezomagnetism
nonlinear electric susceptibility	[V ³]	κ_{ijk}	electro-optic effect, hyper Raman effect
nonlinear magnetic susceptibility	ae[V ³]	χ_{ijk}	magneto-optic effect
magnetobielectric coefficient	aeV[V ²]	α_{ijk}	second-order magnetoelectric effect
electrobimagnetic coefficient	V[V ²]	β_{ijk}	second-order magnetoelectric effect

For the magnetoelectric effect, the magnetoelectric susceptibility α_{ij} transforms as a tensor of the type aeV². From published tables [22], the form of the magnetoelectric effect tensor invariant under the magnetic point group $\bar{4}'2'\mathbf{m}$ is the same as the form of a tensor of the type eV² invariant under the nonmagnetic point group **4mm**. The latter, for the point group $4_z\mathbf{m}_x\mathbf{m}_{xy}$ is [23]

$$\begin{bmatrix} 0 & A & 0 \\ -A & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \tag{4}$$

which agrees with the form of the magnetoelectric physical property tensor invariant under $\bar{4}'_z2'_x\mathbf{m}_{xy}$ given in [15].

3 Tensor distinction of domains in ferroic crystals

A ferroic crystal contains two or more equally stable *domains*, volumes of the same homogeneous crystalline structure in different spatial orientations. These domains can coexist in a crystal and may be distinguished by the values of components of certain macroscopic tensorial physical properties of the domains. Crystals in which the domains may be distinguished by spontaneous polarization, magnetization or strain are called primary ferroic crystals [29-30]. Crystals whose domains are characterized by differences in the piezoelectric tensor [31-32] is an example of a secondary ferroic crystal [33]-[35].

Consider a ferroic phase transition, that is, a phase transition of a crystalline structure from a phase of higher point group symmetry **G** to a phase of lower point group symmetry **F**. In the lower symmetry phase there are $n = |\mathbf{G}|/|\mathbf{F}|$ *single domain states* $S_1, S_2, \dots, S_n$, where $|\mathbf{G}|$ and $|\mathbf{F}|$ denote the number of elements in **G** and **F**, respectively. Single domain states have the same crystalline structure

and differ only in their orientation in space. *Domain states* refer to the bulk structures, with their specific orientations in space, of domains in a polydomain sample. Several disconnected domains can have the same domain state. Domain states represent then the structures that appear in a polydomain sample, irrespective of which domain they are in.

In *nonferroelastic* polydomain phases, the orientation of each domain state coincides with the orientation of a single domain state. The number of domain states is therefore the same as the number of single domain states. In *ferroelastic* polydomain phases, because of disorientations, i.e., rotations of domains that arise as a result of the requirement that neighboring domains in the polydomain sample must meet along a coherent boundary, domain states in general differ in their orientation from single domain states. The number of domain states is then, in general, greater than the number of single domain states. In distinction to domains in nonferroelastic polydomain phases, the orientations are then, in general, not related by the rotational parts of the elements of $\mathbf{G}$. We consider here ferroelastic polydomain phases in the *parent-clamping approximation* [36-37], which disregards the disorientations. We disregard these disorientations; the number and orientation of domain states in a ferroelastic polydomain sample then also coincide with the number and orientations of single domain states, as in nonferroelastic polydomain samples.

The domain states, denoted also by $S_1, S_2, \dots, S_n$, are related by elements of $\mathbf{G}$ not found in $\mathbf{F}$: We subdivide the group $\mathbf{G}$ into a left coset decomposition with respect to $\mathbf{F}$:

$$\mathbf{G} = \mathbf{F} + g_2\mathbf{F} + g_3\mathbf{F} + \dots + g_n\mathbf{F}, \quad (5)$$

where the elements g_i , $i = 1, 2, \dots, n$, $g_1 = 1$, are called the *coset representatives* of the coset decomposition of $\mathbf{G}$ with respect to $\mathbf{F}$. Denoting the elements of the group $\mathbf{F}$ by $\{1, f_2, \dots, f_s\}$, each coset $g_i\mathbf{F}$ denotes a subset of elements $\{g_i, g_i f_2, \dots, g_i f_s\}$ of $\mathbf{G}$. The choice of coset representatives is not unique because the coset representative g_i can be replaced by $g_i f$, where f is any element of $\mathbf{F}$. Defining the domain state S_1 as the domain invariant under $\mathbf{F}$, the orientations of the remaining domain states are related to S_1 by the coset representatives of (5), i.e., $S_i = g_i S_1$, $i = 2, 3, \dots, n$. In addition, each domain state S_i , $i = 1, 2, \dots, n$, is invariant under the group $\mathbf{F}_i \equiv g_i \mathbf{F} g_i^{-1}$.

For example, if $\mathbf{G} = 4_z / \mathbf{m}_z \mathbf{m}'_x \mathbf{m}'_{xy}$ and $\mathbf{F} = 2'_{xy} / \mathbf{m}'_{xy}$, then

$$\begin{array}{llll} \mathbf{F} = \{1, \bar{1}, 2'_{xy}, \mathbf{m}'_{xy}\} & g_1 = 1 & & \mathbf{F}_1 \equiv g_1 \mathbf{F} g_1^{-1} = 2'_{xy} / \mathbf{m}'_{xy} \\ g_2 \mathbf{F} = \{2_z, m_z, 2'_{\bar{x}y}, \mathbf{m}'_{\bar{x}y}\} & g_2 = 2_z & S_2 = 2_z S_1 & \mathbf{F}_2 \equiv g_2 \mathbf{F} g_2^{-1} = 2'_{xy} / \mathbf{m}'_{xy} \\ g_3 \mathbf{F} = \{2'_y, 4_z, \bar{4}_z, \mathbf{m}'_y\} & g_3 = 2'_y & S_3 = 2'_y S_1 & \mathbf{F}_3 \equiv g_3 \mathbf{F} g_3^{-1} = 2'_{\bar{x}y} / \mathbf{m}'_{\bar{x}y} \\ g_4 \mathbf{F} = \{2'_x, 4_z^3, \mathbf{m}'_x, \bar{4}_z^3\} & g_4 = 2'_x & S_4 = 2'_x S_1 & \mathbf{F}_4 \equiv g_4 \mathbf{F} g_4^{-1} = 2'_{\bar{x}y} / \mathbf{m}'_{\bar{x}y}. \end{array}$$

We are interested here in what is referred to as *tensor distinction*, i.e., the distinction of the domains in a polydomain phase of a ferroic phase transition by

macroscopic tensorial physical properties of tensor types $\mathbf{T}$. As the set of domain states represents the structure of all domains in a polydomain phase, we consider the tensor distinction of the domain states.

We denote by T_i , $i = 1, 2, \dots, n$, the tensors of a tensor type $\mathbf{T}$ in the set of domain states $S_1, S_2, \dots, S_n$ of a polydomain sample. The tensors are all given in a single coordinate system, e.g., the coordinate system of the parent phase structure or of one of the domain states. A tensor type $\mathbf{T}$ is said to be able to distinguish between two domain states, with corresponding tensors T_i and T_j of the type $\mathbf{T}$, if $T_i \neq T_j$. In particular, we consider two types of tensor distinction problems separately:

a) *Global tensor distinction*: We consider whether a tensor of type $\mathbf{T}$ can distinguish among all domain states.

b) *Domain pair tensor distinction*: For each pair of domain states, we consider whether a tensor of type $\mathbf{T}$ can distinguish between the domain states of the domain pair.

3.1 Global tensor distinction

Consider a ferroic phase transition of a crystalline structure from a phase of higher symmetry $\mathbf{G}$ to a phase of lower symmetry $\mathbf{F}$. Let $S_1, S_2, \dots, S_n$ denote the domain states of the lower symmetry phase, $\mathbf{T}$ a tensor type, and tensors T_i , $i = 1, 2, \dots, n$, the form of the tensor type $\mathbf{T}$ in the domain states $S_1, S_2, \dots, S_n$. Following the terminology of Aizu [38], if the set of tensors T_i , $i = 1, 2, \dots, n$, are all distinct, then we say that the tensor $\mathbf{T}$ provides a *full* distinction of the domain states S_i , $i = 1, 2, \dots, n$, and the transition is a *full* ferroic phase transition with respect to tensor type $\mathbf{T}$. Each domain state is then characterized by a distinct form of tensor type $\mathbf{T}$, and macroscopic physical properties of this tensor type can distinguish all domain states. If the set of tensors T_i , $i = 1, 2, \dots, n$, are not all distinct, but are also not all identical, then we say that the tensor $\mathbf{T}$ provides a *partial* distinction of the domain states, and the transition is a *partial* ferroic phase transition with respect to tensor type $\mathbf{T}$. A tensor of type $\mathbf{T}$ can then distinguish among some but not all of the domain states. If the set of tensors T_i , $i = 1, 2, \dots, n$, are all identical, then the tensor type $\mathbf{T}$ provides no distinction, we shall say a *null* distinction, of the domain states. The transition is referred to as a *null* ferroic phase transition with respect to the tensor type $\mathbf{T}$.

The *null* case has been subdivided into two [39]: The case where the set of tensors T_i , $i = 1, 2, \dots, n$, is identically zero, is referred to as *zero* distinction, the case where the set of tensors are all identical and *nonzero* is referred to as *null* distinction.

Concomitant with the classification of ferroic phase transitions into full, partial, null, and zero classes with respect to a specific tensor type $\mathbf{T}$, is the classification of ferroic phase transitions with respect to the higher and lower symmetry phase groups $\mathbf{G}$ and $\mathbf{F}$. Using a classification introduced by Aizu [40], 247 classes of nonmagnetic ferroic phase transitions and a general method for determining the

global tensor distinction for classes of ferroic phase transitions were derived [39]. A tabulation of the global tensor distinction of all 247 classes of nonmagnetic ferroic phase transitions for all nonmagnetic tensor types $\mathbf{T}$ of rank $n \leq 4$ has been given [41]. There has not been an analogous listing of the classes of magnetic ferroic phase transitions.

3.2 Domain pair tensor distinction

Consider a ferroic phase transition of a crystalline structure from a phase of higher symmetry $\mathbf{G}$ to a phase of lower symmetry $\mathbf{F}$. Let $S_1, S_2, \dots, S_n$ denote the domain states of the lower symmetry phase, $\mathbf{T}$ a tensor type, and tensors $T_i, i = 1, 2, \dots, n$, the form of tensor type $\mathbf{T}$ in the domain states $S_1, S_2, \dots, S_n$. We now consider all *ordered domain pairs* $\{S_i, S_j\}, i, j = 1, 2, \dots, n, i \neq j$. The tensors T_i and T_j of the two domain states in the domain pair $\{S_i, S_j\}$ can be determined as follows: if T_i is the form of tensor type $\mathbf{T}$ that is invariant under the point group $\mathbf{F}_i$, the point group of the domain state S_i . T_j can be determined from $T_i, T_j = g_{ij}T_i$, where g_{ij} is an element of $\mathbf{G}$ that transforms domain state S_i into domain state S_j ; i.e., $S_j = g_{ij}S_i$. $T_j = g_{ij}T_i$ represents the equation that relates the components of the tensors. The actual equation depends on the transformational properties of the tensor type $\mathbf{T}$ and its rank [14]. The element g_{ij} is not unique, as any element of the coset $g_{ij}\mathbf{F}_i$ can be used. Consequently, in a ferroic phase transition from $\mathbf{G}$ to $\mathbf{F}$, the tensor distinction of a domain pair $\{S_i, S_j\}$ can be determined by the point group $\mathbf{F}_i$ and the element g_{ij} of $\mathbf{G}$.

Let us classify all possible domain pair $\{S_i, S_j\}$ into classes where all domain pairs in a single class are distinguished by the same set of tensor types. To this end we introduce the following tensor classification of domain pairs: Two domain pairs, $\{S_i, S_j\}$ whose tensor distinction is determined by the point group $\mathbf{F}_i$ and element g_{ij} , and $\{S_p, S_q\}$ whose tensor distinction is determined by the point group $\mathbf{F}_p$ and element g_{pq} , are said to be in the same class of domain pair if there exists an element r of the full rotation group $\mathcal{R}$ such that

$$r\mathbf{F}_i r^{-1} = \mathbf{F}_p, \quad (6)$$

$$r g_{ij} r^{-1} = g_{pq} f_p, \quad (7)$$

where f_p is an element of $\mathbf{F}_p$. The appearance of the element f_p in (7) is due to the nonuniqueness of the choice of the coset representative g_{pq} . To remove any nonuniqueness in (7), we can replace (7) with the condition

$$r g_{ij} \mathbf{F}_i r^{-1} = g_{pq} \mathbf{F}_p. \quad (8)$$

All elements of the coset $g_{ij}\mathbf{F}_i$ —i.e., all elements of $\mathbf{G}$ that transform the domain state S_i into the domain state S_j —are taken by (8) into all the elements of $\mathbf{G}$ that transform the domain state S_p into the domain state S_q . If two domain pairs belong to the same class of domain pair, then if a tensor type can or cannot

distinguish between the first pair of domains, then it can or cannot distinguish between the second pair of domains [42]. We refer to these classes of domain pair as *equivalence tensor classes of domain pair*, or simply as *tensor classes of domain pair*.

This tensor classification of domain pairs is equivalent to the following classification [43] of ordered domain pairs: Two domain pairs, $\{S_i, S_j\}$ and $\{S_p, S_q\}$, are said to be in the same class of domain pair if there exists an element r of the full rotation group $\mathcal{R}$ such that $\{rS_i, rS_j\} = \{S_p, S_q\}$. This is a classification of *ordered* domain pairs. That is, the *unordered* pair of domain pairs $\{S_i, S_j\}$ and $\{S_j, S_i\}$ are not automatically placed in the same tensor class. While if a tensor of type $\mathbf{T}$ can or cannot distinguish between domain state S_i and S_j it is trivial to conclude it can or cannot distinguish between the domain states S_j and S_i ; we do not use a classification of unordered domain pairs.

There is a second classification of domain pair [43]. Because of its relationship to the *double coset decomposition* of $\mathbf{G}$ with respect to $\mathbf{F}$, we refer to it as the classification of domain pair into *double coset classes of domain pair*. The double coset decomposition of $\mathbf{G}$ with respect to $\mathbf{F}$ is written as

$$\mathbf{G} = \mathbf{F}g_1^{dc}\mathbf{F} + \mathbf{F}g_2^{dc}\mathbf{F} + \mathbf{F}g_3^{dc}\mathbf{F} + \cdots + \mathbf{F}g_m^{dc}\mathbf{F}, \quad (9)$$

where g_i^{dc} , $i = 1, 2, \dots, m$ are the double coset representatives. A symbol $\mathbf{F}g_i^{dc}\mathbf{F}$ means the set of distinct elements $fg_i^{dc}\hat{f}$, where f and $\hat{f}$ are elements of $\mathbf{F}$. Sets of cosets constitute a single double coset. For example, if $\mathbf{G} = \mathbf{4}_z/\mathbf{m}_z\mathbf{m}'_x\mathbf{m}'_{xy}$ and $\mathbf{F} = \mathbf{2}'_{xy}/\mathbf{m}'_{xy}$, then

$$\begin{aligned} \mathbf{F}g_1^{dc}\mathbf{F} &= \{1, \bar{1}, 2'_{xy}, m'_{xy}\} & g_1^{dc} &= 1 \\ \mathbf{F}g_2^{dc}\mathbf{F} &= \{2_z, m_z, 2'_{\bar{x}y}, m'_{\bar{x}y}\} & g_2^{dc} &= 2_z \\ \mathbf{F}g_3^{dc}\mathbf{F} &= \{2'_y, 4_z, \bar{4}_z, m'_y; 2'_x, 4_z^3, m'_x, \bar{4}_z^3\} & g_3^{dc} &= 2'_y \end{aligned}$$

where there are three double cosets, the third consisting of a set of two cosets.

Two domain pairs $\{S_i, S_j\}$ and $\{S_p, S_q\}$ belong to the same double coset class of domain pair if there exists an element g of $\mathbf{G}$ such that $\{gS_i, gS_j\} = \{S_p, S_q\}$. All domain pairs belonging to the same double coset class also belong to the same tensor class of domain pair. However, domain pairs that belong to *different* double coset classes may also belong to the same tensor class of domain pair. For example, in a phase transition between $\mathbf{G} = 2_x 2_y 2_z$ and $\mathbf{F} = 1$, where $S_2 = 2_x S_1$, $S_3 = 2_y S_1$, and $S_4 = 2_z S_1$, the two domain pairs $\{S_1, S_2\}$ and $\{S_1, S_3\}$ belong to two different double coset classes of domain pair. In this case, in (6)–(8) we have $F_i = F_p = 1$, $g_{ij} = 2_x$, and $g_{pq} = 2_y$. With $r = 2_{xy}$, (6)–(8) are satisfied and these two domain pairs belong to the same tensor class of domain pair.

Computer software exists to calculate the double coset classification of domain pairs for both nonmagnetic [44] and magnetic [9] point groups. These tabulations

also provide indexing and point group symmetry of the domain states, permutations of the domain states under the elements of $\mathbf{G}$, and domain pair characterizing groups, as well as the double coset classification of the domains. The number of double coset classes of domain pairs $\{S_i, S_j\}$ is equal to the number of double cosets in the double coset decomposition of $\mathbf{G}$ with respect to $\mathbf{F}$. A list of one representative double coset from each class of double cosets is $\{S_1, g_i^{dc} S_1\}$, $i = 1, 2, \dots, m$, where m is the number of double cosets in (9). For example, if $\mathbf{G} = 4_z/\mathbf{m}_z\mathbf{m}'_x\mathbf{m}'_{xy}$ and $\mathbf{F} = 2'_{xy}/\mathbf{m}'_{xy}$, then there are three classes whose representative domain pairs are $\{S_1, S_1\}$, $\{S_1, S_2\}$, and $\{S_1, S_3\}$.

3.3 Domain pair symmetry and twinning groups

Pairs of domains $\{S_i, S_j\}$ are characterized by two groups: the *symmetry group* and *twinning group* of the domain pair. The symmetry group $\mathbf{J}_{ij}$ of the domain pair $\{S_i, S_j\}$ is defined by

$$\mathbf{J}_{ij} = \mathbf{F}_{ij} + g_{ij}^* \mathbf{F}_{ij}. \quad (10)$$

Therein, $\mathbf{F}_{ij} = \mathbf{F}_i \cap \mathbf{F}_j$ consists of all elements that simultaneously leave both domain states invariant. g_{ij}^* switches the two domain states; i.e., $g_{ij}^* S_i = S_j$ and $g_{ij}^* S_j = S_i$. The twinning group $\mathbf{K}_{ij}$ of a domain pair $\{S_i, S_j\}$ is defined by [45]

$$\mathbf{K}_{ij} = \langle \mathbf{F}_i, g_{ij} \rangle, \quad (11)$$

where $\mathbf{F}_i$ is the point group of S_i and $g_{ij} S_i = S_j$, and is the group generated by g_{ij} and the elements of the group $\mathbf{F}_i$.

For every magnetic point group $\mathbf{G}$, subgroup $\mathbf{F}$ and representative domain pair $\{S_1, g_i^{dc} S_1\}$, $i = 2, 3, \dots, m$, there exist tabulations of the domain pair's symmetry group and twinning group [9], [44]. The case $i = 1$ is not considered as the corresponding domain pair $\{S_1, S_1\}$ consists of identical domain states. We consider only one domain pair from each class because the relative spatial orientations is the same for the two domain states in each domain pair of a single class [46]. For example, for $\mathbf{G} = 4_z/\mathbf{m}_z\mathbf{m}'_x\mathbf{m}'_{xy}$ and $\mathbf{F} = 2'_{xy}/\mathbf{m}'_{xy}$, for the domain pairs $\{S_1, S_2\}$ and $\{S_1, S_3\}$, one has [44]

$\{S_i, S_j\}$	$\mathbf{J}_{ij} = \mathbf{F}_{ij} + g_{ij}^* \mathbf{F}_{ij}$	$\mathbf{K}_{ij} = \langle \mathbf{F}_i, g_{ij} \rangle$
$\{S_1, S_2\}$	$\mathbf{m}_z\mathbf{m}'_{xy}\mathbf{m}'_{\bar{x}y} = 2'_{xy}/\mathbf{m}'_{xy} + 2_z 2'_{xy}/\mathbf{m}'_{xy}$	$\mathbf{m}_z\mathbf{m}'_{xy}\mathbf{m}'_{\bar{x}y} = \langle 2'_{xy}/\mathbf{m}'_{xy}, 2_z \rangle$
$\{S_1, S_3\}$	$2_y/\mathbf{m}_y = \mathbf{1} + 2'_y \mathbf{1}$	$\mathbf{m}_z\mathbf{m}'_{xy}\mathbf{m}'_{\bar{x}y} = \langle 2'_{xy}/\mathbf{m}'_{xy}, 2_z \rangle$

Note that the tensor distinction of a domain pair $\{S_i, S_j\}$ is determined by the domain pair's twinning group (11). The form T_i of a tensor of type $\mathbf{T}$ in the first domain S_i is that form of $\mathbf{T}$ invariant under $\mathbf{F}_i$, the symmetry group of the first domain. In the second domain S_j , the form T_j of the tensor $\mathbf{T}$ is given by $T_j = g_{ij} T_i$. If the element g_{ij} leaves T_i invariant, then the tensor cannot distinguish between

the two domains of the domain pair. If the element g_{ij} does not leave T_i invariant, then the tensor can distinguish between the two domains of the domain pair. Another way of interpreting this condition for the distinction of two domains is: If the form of a tensor of type **T** invariant under $\mathbf{F}_i$ is (is not) also invariant under $\mathbf{K}_{ij}$, then the tensor of type **T** cannot (can) distinguish between the domains of a domain pair $\{S_i, S_j\}$.

To tabulate all tensor classes of domain pair $\{S_i, S_j\}$, one tabulates the group $\mathbf{F}_i$ and element g_{ij} of one domain pair from each tensor class of domain pairs. This has been done for nonmagnetic point groups where one finds 139 tensor classes of nonmagnetic domain pair $\{S_i, S_j\}$ [47]. In Table 3 we list the 22 tensor classes of domain pair where $\mathbf{F}_i$ is a point group of the type **222** or **mm2**. An asterisk following the sequential numbering denotes one of the 43 classes of nonferroelastic domain pairs, domain pairs with the same (zero) spontaneous deformation. This number of classes differs from the 48 classes of nonferroelastic domain pairs given in [48] because of the tensor classification scheme used here. Those without an asterisk are tensor classes of ferroelastic domain pairs. Also listed is the twinning group $\mathbf{K}_{ij}$, and for each of these classes the tensor distinction for seven types of physical property tensor is also given: e enantiomorphism; V spontaneous polarization; $e[V^2]$ optical activity; $V[V^2]$ piezoelectricity; $eV[V^2]$ electrogyration; $[[V^2]^2]$ and $[V^2]^2$, respectively.

Table 3 Tensor classification of domain pairs and tensor distinction where $\mathbf{F}_i$ is a point group of the type **222** or **mm2**. The indexing in the first column is that of [47]. Columns numbered 1 through 7 are for tensors of types e, V, $e[V^2]$, $V[V^2]$, $eV[V^2]$, $[[V^2]^2]$ and $[V^2]^2$, respectively.

	$\mathbf{F}_i$	g_{ij}	$\mathbf{K}_{ij}$	1	2	3	4	5	6	7
53)*	$2_x 2_y 2_z$	$\bar{1}$	$\mathbf{m}_x \mathbf{m}_y \mathbf{m}_z$	Y	Z	Y	Y	N	N	N
54)	$2_x 2_y 2_z$	2_{xy}	$4_z 2_x 2_{xy}$	N	Z	Y	Y	Y	Y	Y
55)	$2_x 2_y 2_z$	m_{xy}	$4_z 2_x \mathbf{m}_{xy}$	Y	Z	Y	Y	Y	Y	Y
56)	$2_x 2_y 2_z$	3_{xyz}	$2_z 3_{xyz}$	N	Z	Y	Y	Y	Y	Y
57)	$2_x 2_y 2_z$	$\bar{3}_{xyz}$	$\mathbf{m}_z \bar{3}_{xyz}$	Y	Z	Y	Y	Y	Y	Y
58)	$2_{xy} 2_y 2_z$	2_z	$4_z 3_{xyz} 2_{xy}$	N	Z	Y	Y	Y	Y	Y
59)	$2_{xy} 2_y 2_z$	m_z	$\mathbf{m}_z 3_{xyz} \mathbf{m}_{xy}$	Y	Z	Y	Y	Y	Y	Y
60)	$2_x 2_2 2_z$	2_1	$6_z 2_x 2_1$	N	Z	Y	Y	Y	Y	Y
61)	$2_x 2_2 2_z$	m_1	$6_z / \mathbf{m}_z \mathbf{m}_x \mathbf{m}_1$	Y	Z	Y	Y	Y	Y	Y
62)*	$\mathbf{m}_x \mathbf{m}_y 2_z$	$\bar{1}$	$\mathbf{m}_x \mathbf{m}_y \mathbf{m}_z$	Z	Y	Y	Y	N	N	N
63)	$\mathbf{m}_x \mathbf{m}_y 2_z$	m_{xy}	$4_z \mathbf{m}_x \mathbf{m}_{xy}$	Z	N	Y	Y	Y	Y	Y
64)	$\mathbf{m}_x \mathbf{m}_y 2_z$	2_{xy}	$4_z 2_{xy} \mathbf{m}_x$	Z	Y	N	Y	Y	Y	Y
65)	$\mathbf{m}_x \mathbf{m}_y 2_z$	2_z	$4_x / \mathbf{m}_x \mathbf{m}_y \mathbf{m}_{yz}$	Z	Y	Y	Y	Y	Y	Y
66)	$\mathbf{m}_x \mathbf{m}_y 2_z$	3_{xyz}	$\mathbf{m}_z \bar{3}_{xyz}$	Z	Y	Y	Y	Y	Y	Y
67)	$\mathbf{m}_{xy} \mathbf{m}_y 2_z$	m_z	$4_z 3_{xyz} \mathbf{m}_{xy}$	Z	Y	Y	Y	Y	Y	Y
68)	$\mathbf{m}_{xy} \mathbf{m}_y 2_z$	2_z	$\mathbf{m}_z \bar{3}_{xyz} \mathbf{m}_{xy}$	Z	Y	Y	Y	Y	Y	Y
69)	$\mathbf{m}_{\bar{x}y} \mathbf{m}_z 2_{xy}$	$m_{\bar{x}z}$	$\mathbf{m}_z \bar{3}_{xyz} \mathbf{m}_{xy}$	Z	Y	Y	Y	Y	Y	Y
70)	$\mathbf{m}_{\bar{x}y} \mathbf{m}_z 2_{xy}$	$2_{\bar{y}z}$	$\mathbf{m}_z \bar{3}_{xyz} \mathbf{m}_{xy}$	Z	Y	Y	Y	Y	Y	Y
71)	$\mathbf{m}_z \mathbf{m}_y 2_1$	2_2	$6_z \mathbf{m}_x 2_1$	Z	Y	Y	Y	Y	Y	Y
72)	$\mathbf{m}_z \mathbf{m}_2 2_x$	2_1	$6_z / \mathbf{m}_z \mathbf{m}_x \mathbf{m}_1$	Z	Y	Y	Y	Y	Y	Y
73)	$\mathbf{m}_x \mathbf{m}_2 2_z$	m_1	$6_z \mathbf{m}_x \mathbf{m}_1$	Z	N	Y	Y	Y	Y	Y
74)	$\mathbf{m}_x \mathbf{m}_2 2_z$	2_1	$6_z / \mathbf{m}_z \mathbf{m}_x \mathbf{m}_1$	Z	Y	Y	Y	Y	Y	Y

$[[V^2]^2]$ linear elasticity; and $[V^2]^2$ piezo-optics. In Table 3, “Y” denotes that the physical property represented by a tensor type can distinguish between the domains of domain pairs of that class; “N” denotes that the tensor cannot distinguish between the domains and “Z” denotes that the tensor is identically zero in both domains.

The purpose of the foregoing classification of domain pairs into tensor classes is to provide a classification in which one can determine whether a tensor of a specific tensor type can or cannot distinguish between the domains of the domain pair. If a tensor of a specific type can distinguish between the domains, then subsequently one would wish to know which components are the same and which are different in the two domains. This additional problem we refer to as *tensor component distinction*. While we do not intend to focus on this problem here, the above classification of domain pairs into tensor classes has been chosen to take the tensor component distinction problem into account. For two pairs of domain pairs, belonging to the same tensor class of domain pair, there exist coordinate systems for each pair where the tensor component distinction is the same. That is, if a specific component is the same (different) within the domains of the first domain pair, the identical component is the same (different) within the domains of the second pair. It is for this reason that a tensor classification has been defined where domain pairs belonging to classes 69 and 70 from Table 3 are in distinct classes even though the identical groups $\mathbf{F}_i$ and $\mathbf{K}_{ij}$ are associated with them. Consequently, the tensor distinction of domain pairs belonging to both of these classes is identical. The tensor *component* distinction of domain pairs of these two classes is different. Consider the polarization tensor $\mathbf{P}$ and two domain pairs, a domain pair of class 69, where $\mathbf{F}_i = \mathbf{m}_{\bar{x}y} \mathbf{m}_z \mathbf{2}_{xy}$, $g_{ij} = m_{\bar{x}z}$, $\mathbf{K}_{ij} = \mathbf{m}_z \bar{\mathbf{3}}_{xyz} \mathbf{m}_{xy}$, $P_i = (P, P, 0)$ and $P_j = (0, P, P)$ and a domain pair of tensor class 70 where $\mathbf{F}_i = \mathbf{m}_{\bar{x}y} \mathbf{m}_z \mathbf{2}_{xy}$, $g_{ij} = 2_{\bar{y}z}$, $\mathbf{K}_{ij} = \mathbf{m}_z \bar{\mathbf{3}}_{xyz} \mathbf{m}_{xy}$, $P_i = (P, P, 0)$ and $P_j = (-P, 0, -P)$. While polarization does distinguish between the domains in both domain pairs, since both have the same $\mathbf{F}_i$ and $\mathbf{K}_{ij}$, the tensor component distinction is distinct. Comparing the polarization tensors in the domains of the domain pair of tensor class 69 one finds that one component remains the same while the remaining two interexchange, while in the domains of the domain pair of tensor class 70 all three components change.

3.4 Completely transposable twinning groups

Although a complete list of the 139 tensor classes of nonmagnetic domain pairs has been derived, given in terms of 139 nonmagnetic twinning laws, there is no complete list in the case of magnetic domain pairs. Only special cases have been considered. One such case is when the twinning group consists of two coset,

$$\mathbf{K}_{ij} = \langle \mathbf{F}_i, g_{ij} \rangle = \mathbf{F}_i + g_{ij} \mathbf{F}_i, \quad (12)$$

where the element g_{ij} not only transforms S_i into S_j of the domain pair $\{S_i, S_j\}$ but also transforms S_j into S_i . This element interexchanges the two domains of the

domain pair and consequently could be denoted as in the symmetry group, (10), as g_{ij}^* . Such twinning groups are called *completely transposable* twinning groups or twinning laws [49]. In the past, these were referred to as ambivalent twin laws [50]. The point groups of the two domains S_i and S_j of a domain pair $\{S_i, S_j\}$ having a completely transposable twinning group are the same; i.e.,

$$\mathbf{F}_j \equiv g_{ij}\mathbf{F}_i g_{ij}^{-1} = \mathbf{F}_i.$$

The completely transposable twinning group is uniquely characterized by the group $\mathbf{K}_{ij}$ and the group $\mathbf{F}_i$, a subgroup of index two of $\mathbf{K}_{ij}$.

Completely transposable nonmagnetic twin laws have been used in determining macroscopic tensorial physical properties that distinguish domains of a domain pair in the cases of nonferroelastic and ferroelastic domains [48-49]. Because $\mathbf{F}_i$ is a subgroup of index two of $\mathbf{K}_{ij}$, the mathematical structure of nonmagnetic completely transposable twinning groups is the same as that of black and white (dichromatic, antisymmetry) groups and magnetic groups [8], [12]. For example, consider the nonmagnetic completely transposable twinning group

$$4_z 2_x 2_{xy} = 4_z + 2_x 4_z,$$

where the elements of $\mathbf{F}_i = 4_z = \{1, 4_z, 2_z, 4_z^{-1}\}$ leave both of the domains of the domain pair $\{S_i, S_j\}$ invariant, and the elements of the second coset $g_{ij}\mathbf{F}_i = 2_x 4_z = \{2_x, 2_y, 2_{xy}, 2_{\bar{x}y}\}$ interchange the two domains. This nonmagnetic completely transposable twinning group has been given the symbol $4_z 2'_x 2'_{xy}$, where an unprimed symbol such as 4_z denotes that the element 4_z leaves both domains invariant, and a primed symbol such as $2'_x$ denotes that the element 2_x interchanges the two domains of the domain pair [48]. This symbol is not used here to avoid confusing it with that of a magnetic point group. Instead, one replaces the prime with an asterisk, e.g., $4_z 2_x^* 2_{xy}^*$, where the asterisk now denotes that the symbol interchanges the two domains.

Using a double group notation $\mathbf{K}_{ij} [\mathbf{F}_i]$ for completely transposable twinning groups, the 380 classes of magnetic completely transposable twinning groups have been tabulated [51]. If $\mathbf{Q}$ denotes one of the 32 types of nonmagnetic point groups, then there are six types of magnetic completely transposable twinning groups:

$$\begin{array}{lll} 1 : \mathbf{Q}[\mathbf{H}] & 2 : \mathbf{Q}1'[\mathbf{Q}] & 3 : \mathbf{Q}1'[\mathbf{H}1'] = \mathbf{Q}[\mathbf{H}]1' \\ 4 : \mathbf{Q}1'[\mathbf{Q}(\mathbf{H})] = \mathbf{Q}(\mathbf{H})1'^* & 5 : \mathbf{Q}(\mathbf{H})[\mathbf{H}] & 6 : \mathbf{Q}(\mathbf{H})[\mathbf{K}(\mathbf{N})], \end{array} \quad (13)$$

where $\mathbf{H}$ and $\mathbf{K}$ are subgroups of index 2 of $\mathbf{Q}$, and $\mathbf{N}$ is a subgroup of index 2 of $\mathbf{K}$. Further, a prime denotes time inversion and an asterisk denotes that an element interexchanges the two domains. Representative twinning groups of classes

belonging to the family of $\mathbf{Q} = 222$ are

6.1	1	$2_x 2_y 2_z [2_z]$	$2_x^* 2_y^* 2_z$	$2_x 2_y 2_z \{2_z\}$
6.2	2	$2_x 2_y 2_z 1' [2_x 2_y 2_z]$	$2_x 2_y 2_z 1'^*$	$2_x 2_y 2_z 1'^*$
6.3	3	$2_x 2_y 2_z 1' [2_z 1']$	$2_x^* 2_y^* 2_z 1'$	$2_x 2_y 2_z \{2_z\} 1'$
6.4	4	$2_x 2_y 2_z 1' [2'_x 2'_y 2_z]$	$2'_x 2'_y 2_z 1'^*$	$2_x 2_y 2_z (2_z) 1'^*$
6.5	5	$2'_x 2'_y 2_z [2_z]$	$2_x^* 2_y^* 2_z$	$2_x 2_y 2_z (2_z) \{2_z\}$
6.6	6	$2'_x 2'_y 2_z [2'_x]$	$2'_x 2_y^* 2_z^*$	$2_x 2_y 2_z (2_z) \{2_x(1)\}$

(14)

The first column in (14) lists a sequential numbering and the second column the type of the twinning group using the numbering shown in (13). The third column lists the double group notation $\mathbf{K}_{ij} [\mathbf{F}_i]$ and the fourth column a single group notation $\mathbf{K}_{ij}$.

The mathematical structure of magnetic completely transposable twinning groups is the same as that of the so-called double antisymmetry groups introduced by Zamorzaev and Sokolov [52]–[54], where the prime and asterisk here are the analogous double antisymmetries. Double antisymmetry point groups are defined in the following context: All points of a finite object are assigned two properties, each of which can take one of two values, usually interpreted as a plus or minus sign. In addition to the point group transformations of the unsigned finite object, one defines transformations of the signs, a transformation $1'$, which reverses the value of the first sign, and 1^* , which reverses the value of the second sign. A double antisymmetry point group is an invariance group of such a signed finite object, i.e., the group of those point group transformations and point group transformations coupled with $1'$, 1^* , or $1'^*$ that leave the signed finite object invariant. By interpreting time inversion $1'$ and interchanging of domains 1^* as sign-reversing transformations, a magnetic completely transposable twinning group becomes a double anti-symmetry group. In the fifth column in (14), we list the standard double anti-symmetry notation for the groups. The format of this notation is $\mathbf{A}(\mathbf{B})\{\mathbf{C}\}$, where $\mathbf{B}$ is the subgroup of index 2 of $\mathbf{A}$ of elements that are not primed, and $\mathbf{C}$ is the subgroup of index 2 of $\mathbf{A}$ of elements that do not have an asterisk.

Of the 380 magnetic completely transposable twinning groups, 141 are nonferroelastic magnetoelectric twinning groups [55]; that is, twinning groups of domain pairs where the two domains have the same (zero) spontaneous deformation tensor (they are nonferroelastic) and distinct magnetoelectric tensors. For a domain pair with the nonferroelastic magnetoelectric twinning group $\mathbf{K}_{ij}[\mathbf{F}_i] = 4_z 2'_x 2'_{xy} [4_z]$, in Table 4 we give the form of eight tensor types in the two domains. The format in which the tensor components are listed is explained in Appendix B. Note that the spontaneous deformation tensor $[V^2]$ is the same in both domains and the magnetoelectric tensor aeV^2 is distinct.

4 Domain tensors and tensor invariants

In describing the physical properties of crystals by tensors, the components of the tensors are usually given in a Cartesian coordinate system [14]. However, in relat-

Table 4 For the completely transposable magnetic twinning law $K_{ij}[F_i] = 4_z 2'_x 2'_{xy}[4_z]$ and the seven tensor types listed in the first column, the forms of the tensors T_i and $T_j = 2'_x T_i$ are given in the second and third columns, respectively.

Tensor type	Domain state S_i T_i						Domain state S_j T_j					
V	0						0					
	0						0					
	A						$-A$					
aeV	0						0					
	0						0					
	A						A					
[V ²]	A	0	0				A	0	0			
	0	A	0				0	A	0			
	0	0	B				0	0	B			
aeV ²	A	C	0				$-A$	C	0			
	$-C$	A	0				$-C$	$-A$	0			
	0	0	B				0	0	$-B$			
V[V ²]	0	0	0	B	A	0	0	0	0	B	$-A$	0
	0	0	0	A	$-B$	0	0	0	0	$-A$	$-B$	0
	C	C	D	0	0	0	$-C$	$-C$	D	0	0	0
aeV[V ²]	0	0	0	B	A	0	0	0	0	$-B$	A	0
	0	0	0	A	$-B$	0	0	0	0	A	B	0
	C	C	D	0	0	0	C	C	D	0	0	0
[V ³]	0	0	0	0			0	0	0	0		
	0	0	0				0	0	0			
	A	A	B				$-A$	$-A$	$-B$			

ing these components to parameters that drive phase transitions, it is more appropriate and revealing to relate these Cartesian components first to so-called *tensorial* covariants [56]. Tensorial covariants are linear combinations of the cartesian components of tensors that transform as sets of basis functions of irreducible representations of the point group $\mathbf{G}$ in a phase transition from a higher symmetry phase with point group $\mathbf{G}$ to a lower symmetry phase of point group $\mathbf{F}$ [57]. Extensive tables [58] enable one to calculate the Cartesian components of physical property tensors in domains arising in a phase transition from $\mathbf{G}$ to $\mathbf{F}$, in terms of tensor covariants.

For example, in the phase transition from $\mathbf{G} = \mathbf{m}_z \bar{3}_{xyz} \mathbf{m}_{xy}$ to $\mathbf{F} = \mathbf{m}_x \mathbf{m}_y \mathbf{m}_z$, which corresponds to phase transitions in lead zirconate [59] and cesium lead chloride [60], the Cartesian components of the spontaneous deformation tensor, u in the notation [57], in terms of tensorial covariants in the domain invariant under $\mathbf{F} = \mathbf{m}_x \mathbf{m}_y \mathbf{m}_z$, are given by

$$\begin{aligned} u_{xx} &= u_1^+/3 - u_{3x}^+/3 + u_{3y}^+/\sqrt{3} & u_{xy} &= u_{yx} = 0 \\ u_{yy} &= u_1^+/3 - u_{3x}^+/3 - u_{3y}^+/\sqrt{3} & u_{xz} &= u_{zx} = 0 \\ u_{zz} &= u_1^+/3 + u_{3x}^+/3 & u_{yz} &= u_{zy} = 0. \end{aligned}$$

u_1^+ is invariant under $\mathbf{G} = \mathbf{m}_z \bar{\mathbf{3}}_{xyz} \mathbf{m}_{xy}$ and transforms as a basis function of the identity representation of $\mathbf{G}$. $\{u_{3x}^+, u_{3y}^+\}$ transforms as a set of basis functions for the Γ_3^+ irreducible representation of $\mathbf{G}$. Knowing how the tensorial covariants transform under elements of $\mathbf{m}_z \bar{\mathbf{3}}_{xyz} \mathbf{m}_{xy}$ allows one to calculate the components of this tensor in each of the six domains which arise in this phase transition. The nonzero components are

	u_{xx}	u_{yy}	u_{zz}
S_1	$u_1^+/3 - u_{3x}^+/3 + u_{3y}^+/\sqrt{3}$	$u_1^+/3 - u_{3x}^+/3 - u_{3y}^+/\sqrt{3}$	$u_1^+/3 + u_{3x}^+/3$
S_2	$u_1^+/3 + u_{3x}^+/3$	$u_1^+/3 - u_{3x}^+/3 + u_{3y}^+/\sqrt{3}$	$u_1^+/3 - u_{3x}^+/3 - u_{3y}^+/\sqrt{3}$
S_3	$u_1^+/3 - u_{3x}^+/3 - u_{3y}^+/\sqrt{3}$	$u_1^+/3 + u_{3x}^+/3$	$u_1^+/3 - u_{3x}^+/3 + u_{3y}^+/\sqrt{3}$
S_4	$u_1^+/3 - u_{3x}^+/3 - u_{3y}^+/\sqrt{3}$	$u_1^+/3 - u_{3x}^+/3 + u_{3y}^+/\sqrt{3}$	$u_1^+/3 + u_{3x}^+/3$
S_5	$u_1^+/3 + u_{3x}^+/3$	$u_1^+/3 - u_{3x}^+/3 - u_{3y}^+/\sqrt{3}$	$u_1^+/3 - u_{3x}^+/3 + u_{3y}^+/\sqrt{3}$
S_6	$u_1^+/3 - u_{3x}^+/3 + u_{3y}^+/\sqrt{3}$	$u_1^+/3 + u_{3x}^+/3$	$u_1^+/3 - u_{3x}^+/3 - u_{3y}^+/\sqrt{3}$

In general [56], it can be shown that in this phase transition only tensors of the type u , with tensorial covariants which transform as a set of basis functions for the Γ_3^+ , can distinguish between all six domains. That is, only these tensors are fully ferroic in Aizu's classification [38]. One can also find general relationships among the Cartesian components from such tables; e.g., from the table above,

$$\begin{aligned}
 u_{xx}(S_1) &= u_{yy}(S_2) = u_{zz}(S_3) = u_{yy}(S_4) = u_{zz}(S_5) = u_{xx}(S_6) \\
 u_{yy}(S_1) &= u_{zz}(S_2) = u_{xx}(S_3) = u_{xx}(S_4) = u_{yy}(S_5) = u_{zz}(S_6) \\
 u_{zz}(S_1) &= u_{xx}(S_2) = u_{yy}(S_3) = u_{zz}(S_4) = u_{xx}(S_5) = u_{yy}(S_6).
 \end{aligned}$$

If the phase transition is driven by a parameter that does not transform as or couple with a basis function of Γ_3^+ , then in the lower symmetry phase $u_{3x}^+ = u_{3y}^+ = 0$ and $u_{xx}(S_i) = u_{yy}(S_i) = u_{zz}(S_i)$ for $i = 1, 2, \dots, 6$.

5 Domain average engineering of ferroics

Multidomain samples of ferroics (ferroelectrics, ferroelastics and related materials) can give rise to new macroscopic properties. In *domain average engineered* samples of ferroic crystals the specimen is subdivided into a very large number of domains, representing μ domain states where μ is smaller than the theoretically allowed maximum number n , see (5), and forming a regular or irregular pattern. Its response to external fields is described roughly by tensorial properties averaged over all involved domain states. An example of domain average engineering is the piezoelectric properties of PZN-PT single crystals poled along one of the $\{001\}$ directions [61]–[63]. Assuming that the material went through the phase transition from $\mathbf{G} = \mathbf{m}\bar{\mathbf{3}}\mathbf{m}$ to $\mathbf{F} = \mathbf{3m}$, poling along $[100]$ supports the coexistence of four domain states with spontaneous polarization along the directions $[111]$, $[\bar{1}\bar{1}1]$, $[11\bar{1}]$ and $[\bar{1}1\bar{1}]$ with equal probability. The effective symmetry of a domain-average-engineered system is given by a point group.

The classification of subsets of domains and the determination of the effective symmetry point group of a domain-average-engineered sample has been given by Fousek *et al.* [64]: We consider a phase transition from a phase of higher point group symmetry $\mathbf{G}$ to a phase of lower point group symmetry $\mathbf{F}$. Two subsets of domains $\{S_1, S_2, \dots, S_\mu\}$ and $\{S'_1, S'_2, \dots, S'_\mu\}$ are said to belong to the same class of subsets of domains if there exists an element g of $\mathbf{G}$ such that $g\{S_1, S_2, \dots, S_\mu\} = \{gS_1, gS_2, \dots, gS_\mu\} = \{S'_1, S'_2, \dots, S'_\mu\}$. The symmetry group $\mathbf{H}$ of a subset of domains $\{S_1, S_2, \dots, S_\mu\}$ is defined as the group of all elements g of $\mathbf{G}$ that leave the set invariant, i.e., $g\{S_1, S_2, \dots, S_\mu\} = \{S_1, S_2, \dots, S_\mu\}$. The group $\mathbf{H}$ represents the effective symmetry of the domain average engineered system consisting of the subset of domains $\{S_1, S_2, \dots, S_\mu\}$.

As an example, we consider the phase transition from $\mathbf{G} = \mathbf{m}\bar{3}\mathbf{m}$ to $\mathbf{F} = 3_{xyz}\mathbf{m}_{\bar{x}y}$. Here, $n = 8$. The indexing of the domain states, the corresponding coset representatives of the coset decomposition of $\mathbf{G}$ with respect to $\mathbf{F}$, the symmetry groups, and the corresponding polarizations in each domain state are given in Table 5. All subsets of these domain states have been classified into classes as defined above. In Table 6 we list one subset of domain states from each class. Each subset is denoted by listing between square brackets the indexes of the domain states contained in that subset; e.g., the subset $\{S_1, S_3, S_5\}$ is denoted by [135]. In the right-hand column is the subgroup $\mathbf{H}$ of elements of $\mathbf{G}$ that leave the corresponding subset invariant. This table, in fact, represents the list of domain-average-engineered systems which can arise in a material undergoing a phase transition from $\mathbf{m}\bar{3}\mathbf{m}$ to $3\mathbf{m}$.

In Table 6 only one domain is listed from each class. In the class of 4 domains whose representative domain is [1368] are a total of 6 sets of 4 domains:

Subset of Domains	Symmetry $\mathbf{H}$ of Subset
[1278]	$4_x\mathbf{m}_y\mathbf{m}_{yz}$
[1368]	$4_z\mathbf{m}_x\mathbf{m}_{xy}$
[1467]	$4_y\mathbf{m}_x\mathbf{m}_{xz}$
[2358]	$4_y\mathbf{m}_x\mathbf{m}_{xz}$
[2457]	$4_z\mathbf{m}_x\mathbf{m}_{xy}$
[3456]	$4_x\mathbf{m}_y\mathbf{m}_{yz}$

Table 5 Domain state index, coset representative, symmetry group and polarization.

Index i	Coset representative g_i	$\mathbf{F}_i = g_i\mathbf{F}_1g_i^{-1}$	$\mathbf{P}_i = g_i\mathbf{P}_1$
1	1	$3_{xyz}\mathbf{m}_{\bar{x}y}$	(A, A, A)
2	2_x	$3_{\bar{x}yz}\mathbf{m}_{\bar{y}z}$	(A, -A, -A)
3	2_z	$3_{xy\bar{z}}\mathbf{m}_{\bar{x}y}$	(-A, -A, A)
4	2_y	$3_{x\bar{y}z}\mathbf{m}_{\bar{x}z}$	(-A, A, -A)
5	$\bar{1}$	$3_{xyz}\mathbf{m}_{\bar{x}y}$	(-A, -A, -A)
6	m_x	$3_{\bar{x}yz}\mathbf{m}_{\bar{y}z}$	(-A, A, A)
7	m_z	$3_{xy\bar{z}}\mathbf{m}_{\bar{x}y}$	(A, A, -A)
8	m_y	$3_{x\bar{y}z}\mathbf{m}_{\bar{x}z}$	(A, -A, A)

Table 6 Representative subsets of domain states for the species $\bar{\mathbf{m}}\mathbf{3m}$ to $\mathbf{3m}$ and the subgroups of $\bar{\mathbf{m}}\mathbf{3m}$ that leave them invariant.

Representative subset	Symmetry H of the subset
[1] or [2345678]	$\mathbf{3}_{xyz}\bar{\mathbf{m}}_{\bar{x}y}$
[13] or [245678]	$\bar{\mathbf{m}}_{xy}\bar{\mathbf{m}}_{\bar{x}y}\mathbf{2}_z$
[15] or [234678]	$\bar{\mathbf{3}}_{xyz}\bar{\mathbf{m}}_{\bar{x}y}$
[16] or [234578]	$\bar{\mathbf{m}}_x\bar{\mathbf{m}}_{\bar{y}z}\mathbf{2}_{yz}$
[123] or [45678]	$\mathbf{3}_{x\bar{y}z}\bar{\mathbf{m}}_{\bar{x}z}$
[135] or [24678]	$\bar{\mathbf{m}}_{\bar{x}y}$
[136] or [24578]	$\bar{\mathbf{m}}_{xy}$
[1234]	$\mathbf{43m}$
[1235]	$\bar{\mathbf{m}}_{\bar{x}z}$
[1238]	$\mathbf{3}_{x\bar{y}z}\bar{\mathbf{m}}_{\bar{x}z}$
[1356]	$\mathbf{2}_{\bar{x}z}$
[1357]	$\bar{\mathbf{m}}_{xy}\bar{\mathbf{m}}_{\bar{x}y}\bar{\mathbf{m}}_z$
[1368]	$\mathbf{4}_z\bar{\mathbf{m}}_x\bar{\mathbf{m}}_{xy}$

In this class of subsets of four domains is the subset [1278] with corresponding polarizations along the directions [111], [$\bar{1}\bar{1}\bar{1}$], [$\bar{1}\bar{1}\bar{1}$] and [$\bar{1}\bar{1}\bar{1}$], the set of polarization directions of the example given above. The point group of this subset is $\mathbf{4}_x\bar{\mathbf{m}}_y\bar{\mathbf{m}}_{yz}$. A computer program has been developed [65] to calculate the properties of the classification of subsets of domains, including the listing of all subsets of each class.

6 Conclusions

In this review we have surveyed the use of point group symmetry in the distinction of domains by their physical property tensors. Various domains, when referred to a common coordinate system, may have different tensor components. An analysis of the relationship among the tensor components enables one to design experiments for observing and investigating domains, to relate the properties of domains to the parameters that drive phase transitions and to determine the average physical properties of aggregates of domains. General references to these applications of the material surveyed are provided here [14], [15], [25], [30], [57].

Related mathematical topics beyond the purview of this review are the concepts of group theory, the derivation of point groups and how one determines the form of the physical property tensors given the point group symmetry of the crystalline structure. An elementary yet instructive introduction to group theory is [66], and more detailed classical introductions can be found in [8], [11], [67]. The derivation of the magnetic and/or nonmagnetic point groups can be found in [11], [68], [69], and how one mathematically determines the form of the physical property tensors in [10], [14], [15], [17], [18], [23], [70].

Appendix A: Point group symbols

The symbol of a nonmagnetic point group, a magnetic point group of the type **R**, is made up of a combination of letters and numbers. A number *N* represents a

rotation of $360^\circ/N$ and $\bar{1}$ spatial inversion, i.e., inversion through a point. $\bar{N} \equiv N\bar{1}$ represents a rotation of $360^\circ/N$ coupled with spatial inversion $\bar{1}$ and $m \equiv 2\bar{1}$ a reflection across a plane, which is equivalent to a rotation of 180° about an axis perpendicular to the plane coupled with spatial inversion. A specific rotation N is denoted by N_i , where i denotes the axis of rotation, and a specific reflection m is denoted by $m_i \equiv 2_i\bar{1}$, where i denotes the direction perpendicular to the reflection plane.

The symbol of a point group contains the letters and numbers representing the smallest number of elements of the group from which the group can be constructed. For example, the point group $4_z\mathbf{m}_x\mathbf{m}_{xy}$ contains the three elements 4_z , m_x , and m_{xy} . By repetitively multiplying these elements among themselves one can obtain all elements 1 , 4_z , 2_z , 4_z^3 , m_x , m_y , m_{xy} , and $m_{\bar{x}y}$ of the point group, e.g., $4_z \times 4_z \times m_x = m_y$. In the symbol for a specific point group, e.g., $4_z\mathbf{m}_x\mathbf{m}_{xy}$, the relative orientation of the elements represented by the individual symbols is given by the subindexes. In the symbol for a point group type, e.g., $4mm$ as given in Table 1, the relative orientation of the elements shown is implied by the position of the individual symbols [10]. For the 32 types of nonmagnetic crystallographic point groups, we explicitly show these relative orientations of the elements of the point group given in the group symbol by listing one specific group from each type:

1	$\mathbf{m}_x\mathbf{m}_y\mathbf{m}_z$	$4_z/\mathbf{m}_z\mathbf{m}_x\mathbf{m}_{xy}$	$\bar{6}_z$	$\mathbf{m}_z\bar{3}_{xyz}$
$\bar{1}$	4_z	3_z	$6_z/\mathbf{m}_z$	$4_z\bar{3}_{xyz}2_{xy}$
2_z	$\bar{4}_z$	$\bar{3}_z$	$6_z2_x2_y$	$\bar{4}_z\bar{3}_{xyz}\mathbf{m}_{xy}$
$\mathbf{m}_z$	$4_z/\mathbf{m}_z$	3_z2_x	$6_z\mathbf{m}_x\mathbf{m}_y$	$\mathbf{m}_z\bar{3}_{xyz}\mathbf{m}_{xy}$
$2_z/\mathbf{m}_z$	$4_z2_x2_{xy}$	$3_z\mathbf{m}_x$	$\bar{6}_z\mathbf{m}_x2_y$	
$2_x2_y2_z$	$4_z\mathbf{m}_x\mathbf{m}_{xy}$	$\bar{3}_z\mathbf{m}_x$	$6_z/\mathbf{m}_z\mathbf{m}_x\mathbf{m}_y$	
$\mathbf{m}_x\mathbf{m}_y2_z$	$\bar{4}_z2_x\mathbf{m}_{xy}$	6_z	$2_z\bar{3}_{xyz}$	

For magnetic point groups, there are two addition types of symbols. The symbol of the type $\mathbf{R1}'$ is made up of the symbol for the nonmagnetic point group $\mathbf{R}$ and the time inversion symbol $1'$. The elements of such a group contain all the elements of $\mathbf{R}$ and all the elements of $\mathbf{R}$ multiplied by time inversion, e.g., the group $4_z\mathbf{m}_x\mathbf{m}_{xy}1'$ contains the elements 1 , 4_z , 2_z , 4_z^3 , m_x , m_y , m_{xy} , $m_{\bar{x}y}$ and the elements $1'$, $4'_z$, $2'_z$, $4_z'^3$, m'_x , m'_y , m'_{xy} , and $m'_{\bar{x}y}$. The remaining magnetic point groups use the symbol for a nonmagnetic point group $\mathbf{R}$ where individual numbers or letters are primed representing that the element is now multiplied with time inversion. For example, the group $4_z\mathbf{m}'_x\mathbf{m}'_{xy}$ symbol denotes the elements 4_z , $m'_x = m_x1'$ and $m'_y = m_y1'$. The group elements are 1 , 4_z , 2_z , 4_z^3 , m'_x , m'_y , m'_{xy} , $m'_{\bar{x}y}$.

Appendix B: Form of tensors

Tensors of the type V and aeV are tensors of rank 1, i.e., the tensor's components T_i have one index which, in a Cartesian coordinate system, takes the values $i = x$,

y and z. The form of presenting all components of tensors of rank 1 is, as in Table 4 [14], [23], [70]:

$$\begin{matrix} T_x \\ T_y \\ T_z. \end{matrix}$$

Tensors of the type $[V^2]$ and aeV^2 in Table 4 and eV^2 in (4) are second-rank tensors, i.e., having components T_{ij} with two indexes. Type $[V^2]$ are also symmetric, i.e., $T_{ij} = T_{ji}$. The components of these tensors are listed in the format:

$$\begin{matrix} T_{xx} & T_{xy} & T_{xz} \\ T_{yx} & T_{yy} & T_{yz} \\ T_{zx} & T_{zy} & T_{zz} \end{matrix}$$

Tensors of the type $V[V^2]$ and $aeV[V^2]$ are third-rank tensors, i.e., having components T_{ijk} with three indexes, which are also symmetric with respect to two indexes, i.e., $T_{ijk} = T_{ikj}$. Because these tensors are symmetric, not all tensor components are listed. The components listed and the format of listing them is

$$\begin{matrix} T_{xxx} & T_{xxy} & T_{xxz} & T_{xyy} & T_{xyz} & T_{xzz} \\ T_{yxx} & T_{yyx} & T_{yxz} & T_{yyy} & T_{yyz} & T_{yzz} \\ T_{zxx} & T_{zxy} & T_{zxz} & T_{zyy} & T_{zyz} & T_{zzz} \end{matrix}$$

Tensors of the type $[V^3]$ are third-rank tensors symmetric with respect to all three indexes; i.e., $T_{ijk} = T_{ikj} = T_{jki} = T_{jik} = T_{kij} = T_{kji}$. The components listed and the format of listing them is

$$\begin{matrix} T_{xxx} & T_{xyy} & T_{xzz} & T_{xyz} \\ T_{yxx} & T_{yyy} & T_{yzz} \\ T_{zxx} & T_{zyy} & T_{zzz} \end{matrix}$$

References

- [1] W. Opechowski, Magnetoelectric symmetry, *Magnetoelectric interaction phenomena in crystals* (A.J. Freeman and H. Schmid, eds), Gordon and Breach, London, UK, 1975.
- [2] H. Zocher and C. Török, About space-time asymmetry in the realm of classical, general and crystal physics, *Proc Nat Acad Sci USA* **30** (1953), 681–686.
- [3] T.H. O'Dell, *The electrodynamics of magneto-electric media*, North-Holland, Amsterdam, The Netherlands, 1970.
- [4] L.L. Landau and E.M. Lifshitz, *Statistical physics*, Pergamon Press, London, UK, 1958.

- [5] I.E. Dyaloshinskii, On the magneto-electric effect in antiferromagnetics, *Zh Exp Teor Fiz* **37** (1959), 881–882 [*Sov Phys JETP* **10** (1960), 628–629].
- [6] D.N. Astrov, The magnetoelectric effect in antiferromagnetics, *Zh Exp Teor Fiz* **38** (1960), 984–985 [*Sov Phys JETP* **11** (1960), 708–709].
- [7] V.J. Folen, G.T. Rado and E.W. Stalder, Anisotropy of the magnetoelectric effect in Cr_2O_3 , *Phys Rev Lett* **6** (1961), 607–608.
- [8] W. Opechowski, *Crystallographic and metacrystallographic groups*, North-Holland, Amsterdam, The Netherlands, 1986.
- [9] J. Schlessman and D.B. Litvin, Coset and double coset decomposition of the magnetic point groups, *Acta Cryst A* **57** (2001), 114–115.
- [10] Th. Hahn (ed), *International tables for crystallography, Vol A: Space group symmetry*, Kluwer, Dordrecht, The Netherlands, 1983 [revised editions: 1987, 1989, 1993, 1995].
- [11] M. Hamermesh, *Group theory and its application to physical problems*, Addison-Wesley, Reading, MA, USA, 1962.
- [12] A.V. Shubnikov, Symmetry and antisymmetry of finite figures, *Colored symmetry* (W.T. Hoser, ed), Pergamon Press, New York, NY, USA, 1964.
- [13] H.A. Jahn, Note on the Bhagavantam-Suryanarayana method of enumerating the physical constants of crystals, *Acta Cryst* **2** (1949), 30–33.
- [14] J.F. Nye, *Physical properties of crystals*, Clarendon Press, Oxford, UK, 1957.
- [15] R.R. Birss, *Symmetry and magnetism*, North-Holland, Amsterdam, The Netherlands, 1964.
- [16] V. Kopsky, A simplified calculation and tabulation of tensorial covariants for the magnetic point groups belonging to the same Laue class, *Acta Cryst A* **35** (1979), 95–101.
- [17] D.E. Sands, *Vectors and tensors in crystallography*, Addison-Wesley, Reading, MA, USA 1982.
- [18] J. Brandmuller and F.X. Winter, Influence of symmetry on the static and dynamic properties of crystals, *Z Kristallogr* **172** (1985), 191–231.
- [19] S. Bhagavatam, Crystal symmetry and magnetic properties, *Proc Indian Acad Sci* **34** (1965), 72–74.
- [20] S. Bhagavatam and P.V. Pantulu, Crystal symmetry and physical properties: Application of group theory, *Proc Indian Acad Sci* **59** (1964), 1–10.
- [21] J. Tenenbaum, Tabulation of the number of independent components of physical tensors for the magnetic classes, *Proc Indian Acad Sci* **64** (1966), 74–83.
- [22] H. Grimmer, General connections for the form of property tensors in the 122 Shubnikov point groups, *Acta Cryst A* **47** (1991), 226–232.
- [23] Yu. Sirotnin and M.P. Shaskolskaya, *Fundamentals of crystal physics*, Mir, Moscow, USSR, 1982.
- [24] J. Brandmuller, H. Bross, W.G. Bauer and F.X. Winter, Tables to accompany J. Brandmuller and F.X. Winter, Influence of symmetry on the static and dynamic properties of crystals, *Z Kristallogr* **172** (1985), 191–231, second corrected edition with extension to noncrystallographic point groups (1986).

- [25] A. Authier (ed), *International tables for crystallography, Vol D: Physical properties of crystals*, Kluwer, Dordrecht, The Netherlands, in press.
- [26] D.B. Litvin, Magnetic physical property tensors, *Acta Cryst A* **50** (1994), 406–408.
- [27] H. Schmid, On a magnetoelectric classification of materials *Magnetoelectric interaction phenomena in crystals* (A.J. Freeman and H. Schmid, eds), Gordon and Breach, London, UK, 1975.
- [28] H. Schmid, Magnetoelectric effects in insulating magnetic materials, *Introduction to complex mediums for optics and electromagnetics* (W.S. Weiglhofer and A. Lakhtakia, eds), SPIE Press, Bellingham, WA, USA, 2003.
- [29] R.E. Newnham, Domains in minerals, *Am Mineral* **59** (1974), 906–918.
- [30] V.K. Wadhawan, *Introduction to ferroic materials*, Gordon and Breach, New York, NY, USA, 2000.
- [31] K. Aizu, Second order ferroic state shifts, *J Phys Soc Jpn* **34** (1973), 121–128.
- [32] R.E. Newnham and L.E. Cross, Secondary ferroics and domain-divided piezoelectrics, *Ferroelectrics* **10** (1976), 269–276.
- [33] R.E. Newnham and L.E. Cross, Symmetry of secondary ferroics I, *Mater Res Bull* **9** (1974), 927–934.
- [34] R.E. Newnham and L.E. Cross, Symmetry of secondary ferroics II, *Mater Res Bull* **9** (1974), 1021–1032.
- [35] R.E. Newnham and D.P. Skinner Jr., Polycrystalline secondary ferroics, *Mater Res Bull* **11** (1976), 1273–1284.
- [36] Z. Zikmund, Symmetry of domain pairs and domain twins, *Czech J Phys B* **34** (1984), 932–949.
- [37] J. Janovec, W. Schranz, H. Warhanek and Z. Zikmund, Symmetry analysis of domain structure in KSCN crystals, *Ferroelectrics* **98** (1989), 171–189.
- [38] K. Aizu, Possible species of ferromagnetic, ferroelectric, and ferroelastic crystals, *Phys Rev B* **2** (1970), 754–772.
- [39] D.B. Litvin, Tensorial classification of nonmagnetic ferroic crystals, *Acta Cryst A* **40** (1984), 255–261.
- [40] K. Aizu, Comprehensive tabulation of the four categories of ferroic point groups derived from each of the 31 prototype point groups, *J Phys Soc Jpn* **46** (1979), 1716–1725.
- [41] D.B. Litvin, Tensorial classification of nonmagnetic ferroic crystals: tensors of rank $n \leq 4$, *Acta Cryst A* **41** (1985), 89–90.
- [42] D.B. Litvin, Tensor distinction of domains in ferroic crystals, *Acta Cryst A* **55** (1999), 884–890.
- [43] J. Janovec, Group analysis of domains and domain pairs, *Czech J Phys B* **22** (1972), 974–994.
- [44] J. Schlessman and D.B. Litvin, Nonmagnetic twin laws, *Acta Cryst A* **51** (1995), 947–949.
- [45] J. Janovec, D.B. Litvin and J. Fuksa, Transposable domain pairs and domain distinction, *Ferroelectrics* **172** (1995), 351–359.

- [46] D.B. Litvin and T.R. Wike, Tensorial classification of domain pairs, *Ferroelectrics* **98** (1989), 157–166.
- [47] D.B. Litvin and V. Janovec, Classification of domain pairs and tensor distinction, *Ferroelectrics* **222** (1999), 87–93.
- [48] J. Janovec, L. Richterova and D.B. Litvin, Nonferroelastic twin laws and distinction of domains in nonferroelastic phases, *Ferroelectrics* **140** (1993), 95–100.
- [49] J. Janovec, D.B. Litvin and L. Richterova, Tensor distinction of ferroelastic domain states in completely transposable domain pairs, *Ferroelectrics* **157** (1994), 75–80.
- [50] J. Janovec, Symmetry and structure of domain walls, *Ferroelectrics* **35** (1981), 105–110.
- [51] D.B. Litvin, S.Y. Litvin and V. Janovec, Magnetic completely transposable twin laws and tensor distinction, *Acta Cryst A* **51** (1995), 534–529.
- [52] A.M. Zamorzaev and E.I. Sokolov, Symmetry and various kinds of antisymmetry of finite bodies, *Kristallografiya* **2** (1957), 9–14 [*Sov Phys Crystallogr* **2** (1957), 5–9].
- [53] A.M. Zamorzaev, *The theory of simple and multiple antisymmetry*, Shtintsa, Kishinev, Moldova, 1976.
- [54] A.M. Zamorzaev and A.F. Palistrant, Antisymmetry, its generalizations and geometrical applications, *Z Kristallogr* **151** (1980), 231–248.
- [55] D.B. Litvin, V. Janovec and S.Y. Litvin, Nonferroelastic magnetoelectrin twin laws, *Ferroelectrics* **162** (1994), 275–280.
- [56] V. Kopsky and D.B. Litvin, Tensorial covariants and domain state tensors, *Ferroelectrics* **252** (2001), 59–67.
- [57] V. Kopsky, Tensor parameters of ferroic phase transitions I. Theory and tables, *Phase Transitions* **73** (2001), 1–422.
- [58] V. Kopsky, *Group lattices, subduction of basis, and fine domain structures for magnetic crystal point groups*, Akademia, Prague, Czechoslovakia, 1982.
- [59] K.H. Hellwege and A.M. Hellwege (eds), *Landolt-Boernstein—Condensed Matter III/16b: Ferroelectrics and related substances—nonoxides*, Springer, Berlin, Germany, 1982.
- [60] M. Chabin and F. Gilletta, Theoretical investigation of the ferroelastic domain structure in cesium lead chloride in the monoclinic phase, *J Appl Cryst* **13** (1980), 533–538.
- [61] S.E. Park and T.R. Shrout, Ultrahigh strain and piezoelectric behavior in relaxor based ferroelectric single crystals, *J Appl Phys* **82** (1997), 1804–1811.
- [62] P. Wada, S.E. Park, L.E. Cross and T.R. Shrout, Engineered domain configuration in rhombohedral PZN-PT single crystals and their ferroelectric related properties, *Ferroelectrics* **221** (1999), 147–155.
- [63] J. Yin and W. Cao, Domain configurations in domain engineered $0.955\text{Pb}(\text{Zn}_{1/3}\text{Nb}_{2/3})\text{O}_3-0.045\text{PbTiO}_3$ single crystals, *J Appl Phys* **87** (2000), 7438–7441.

- [64] J. Fousek, D.B. Litvin and L.E. Cross, Domain geometry engineering and domain average engineering in ferroics, *J Phys: Condens Matter* **13** (2001), L33–L38.
- [65] B. Shaparenko, J. Schlessman and D.B. Litvin, Domain average engineering in ferroics, *Ferroelectrics* **269** (2002), 9–14.
- [66] I. Adler, *Groups in the new mathematics*, John Day, New York, NY, USA, 1967.
- [67] E.P. Wigner, *Group theory and its application to the quantum mechanics of atomic spectra*, Academic Press, New York, NY, USA, 1959.
- [68] V.K. Vainshtein, *Modern crystallography I: Symmetry of crystals*, Springer, Berlin, Germany, 1981.
- [69] E.H. Lockwood and R.H. Macmillan, *Geometric symmetry*, Cambridge University Press, Cambridge, UK, 1978.
- [70] S.V. Popov, Yu. P. Svirko and N.I. Zheludev, *Susceptibility tensors for non-linear optics*, Institute of Physics Publishing, Bristol, UK, 1995.



Daniel B. Litvin is a Distinguished Professor of Physics in the Eberly College of Science at the Pennsylvania State University. He has published more than 90 journal articles and two books. He is co-editor and co-author of a volume of the International Tables for Crystallography and is a member of the Commission on Crystallographic Nomenclature and the Commission on International Tables of the International Union of Crystallography. He was awarded an Amoco award for excellence in teaching at the undergraduate level and a Pennsylvania State University—Berks Campus Outstanding Teaching award.

Nonlinear Optics Using Semiconductor Quantum Wells

John M. Arnold

Abstract

Semiconductor quantum wells and superlattices are nanostructured metamaterials with nonlinear optical properties that can differ from those of the bulk semiconductors of which they are composed. In addition, it is possible to alter these properties on a spatially local scale by planar microstructuring techniques, such as quantum-well intermixing. Here, an example of these technologies is considered, the generation of optical second harmonic in a guided-wave environment, in which the nonlinear function is provided by a semiconductor superlattice and the quasi-phase-matching grating is synthesized by spatially selective quantum-well intermixing.

1 Introduction

In this chapter we discuss some applications of semiconductor quantum-well materials to nonlinear optics, in particular the applications that are likely to arise in future all-optical communications networks. Already these networks are capable of operating in the region of 40 Gb s^{-1} individual channel capacities, and wavelength division multiplex (WDM) technology multiplexes many channels together into composite signals that can carry many Tb s^{-1} . Electronic-based signal processing functions are virtually extinct at these data rates, and all-optical realizations of signal processing are essential. Examples of such signal processing functions are wavelength commutation, ultrafast optical pulse generation, modulation and detection, channel switching and routing, packet header identification, digital clock extraction and many others. All of these involve some kind of nonlinear processing; that is to say, they cannot be realized by purely linear filtering of the optical signal. The nonlinearity of all these processes, in conjunction with the requirement that they have to operate ultimately on subpicosecond time scales, suggests that the development of nonlinear optical devices is highly desirable for further technological progress to meet the increasing demands of optical networks.

In prospecting for suitable materials in which to carry out ultrafast nonlinear optics at reasonable levels of optical power, semiconductors present a quite powerful case for consideration. In particular, semiconductors such as gallium arsenide have very well developed growth and fabrication technologies, they can be equipped with current injection to form lasers that are the fundamental sources of coherent optical energy, and they can in turn be assembled with other components on a single substrate to produce integrated devices of high functionality.

An additional capability of semiconductor materials that is not accessible in wide-bandgap transparent insulators is the formation of *quantum wells*, which are essentially thin layers of a low-bandgap material (wells) enclosed by regions of a higher-bandgap material (barriers). A quantum well has the unique capability to trap electrons of a certain range of energies in the well region, creating a metamaterial with effectively two-dimensional properties embedded in a fully three-dimensional medium. The quantum-mechanical wavefunctions of such trapped electrons extend slightly outside the wells by tunnelling into the barriers, but only over very small distances of a few nanometers. If many quantum wells are stacked together, sufficiently closely that significant tunnelling of the electron from one well to the next nearest can occur, the coupled well states behave as an extended system, and the resulting composite material is known as a *superlattice*.

Quantum wells and superlattices are very well established features of optoelectronic devices. They can be grown by precise epitaxial processes with monolayer control on well thicknesses and barrier spacings, and the technology of selective-area quantum-well intermixing (QWI) has provided a reliable means of spatially patterning integrated devices with feature resolutions around $1 \mu\text{m}$. Quantum-well intermixing is a planar fabrication technology that permits the quantum wells to be destroyed in chosen regions of a semiconductor wafer so that the intermixed regions revert to being locally like bulk barrier material, with higher bandgap. Lasers

made from quantum-well materials operate with lower threshold currents than the earlier heterojunction lasers. In addition, selective-area QWI can be used to monolithically integrate passive components with active laser regions, since the higher-bandgap intermixed regions do not absorb significantly the light emitted by the lower-bandgap laser regions. Using this type of technology, QWI has been used to produce many forms of integrated multifunction devices such as a laser integrated with a modulator and connecting waveguide. Selective-area QWI can be achieved by several alternative methods, including cap-induced vacancy disordering, implantation of impurity atoms at high energies, laser absorption in the quantum well, and several others [1].

Nonlinear optics is also of fundamental importance in the functioning of lasers themselves, when they are operating modelocked or Q-switched to produce short-pulse modulation directly on the laser light output. Semiconductor quantum-well lasers, integrated with one or more intracavity saturable absorbers, have produced light with envelope pulse modulations in the region of several hundred gigahertz and pulse durations down to 1 ps, the current record being an envelope modulation of 2.1 THz produced from a harmonic-cavity modelocked laser in our own laboratories in the University of Glasgow [2]. The action of nonlinearities in the quantum-well saturable absorbers and the dynamics of the quantum-well gain medium are fundamental to the operation of these devices.

Even in bulk-optics lasers using solid-state gain mediums, optical nonlinearities in semiconductor quantum wells have recently played a crucial role in driving the pulse widths of modelocked lasers to limits around the duration of one optical cycle or less. A solid-state bulk-optics laser containing a quantum-well saturable absorber has produced pulses as short as 6.5 fs [3], and studies are active with the goal of reducing this to 1 fs or less. In the widely used method of Kerr-lens modelocking, intensity-dependent self-focusing in a nonlinear medium followed by aperturing of the self-focused beam can be used to mimic the behavior of a saturable absorber, but with response times that are on femtosecond time scales rather than the tens of picoseconds typical of saturable absorption, leading again to modelocked pulses with durations < 100 fs. A very compact way of realizing this kind of pulse shaping is by means of a spatial soliton in a planar semiconductor waveguide such that the spatial soliton is channelled through an aperture integrated in the waveguide. The spatial soliton diffracts in the plane of the planar waveguide by an amount that depends on the peak intensity of the incident beam, so that the transmission through the aperture depends on the intensity, higher intensity having higher transmission through the aperture than lower intensity. This resembles the behavior of a saturable absorber, but no absorption is actually taking place in the optical medium.

In further developments of these concepts, materials have been studied in which the electrons in the semiconductor are confined in more dimensions than the one dimension of the quantum well. Thus, quantum wires have electron confinement in two dimensions, and quantum dots have electron states that are confined in all three spatial dimensions. Quantum dots can be grown on the planar surface of an epitax-

ially grown semiconductor by a self-organizing phenomenon known as Stranski-Krastanow growth, which is an elastic strain instability that favors the deposition of material in isolated pyramids of typical dimensions a few tens of nanometers, rather than as a homogeneous layer as in normal epitaxial crystal growth. These pyramids can act as total traps for electrons, and the associated quantum energy states of electrons trapped in the dots are completely discrete, as in simple atoms below the ionization energy. Not so much is known about the practicalities of nonlinear optics in quantum dots, yet there may be significant enhancement of optical nonlinearities available in quantum dots due to the totally confined nature of the electron states.

Quantum wells and superlattices provide means of realizing nonlinear optical phenomena in engineered devices, which is an attractive route to all-optical switching, wavelength conversion and pulse shaping. Quantum-well intermixing technology provides a means of engineering devices through basic manipulation of the materials. These structures form prototype metamaterials with highly useful optical properties in the nonlinear regime.

2 Theoretical nonlinear optics

Nonlinear optical effects in electromagnetic theory are generally described by the polarization expansion $\hat{\mathbf{P}} = \hat{P}_1\mathbf{u}_1 + \hat{P}_2\mathbf{u}_2 + \hat{P}_3\mathbf{u}_3$ induced in the dielectric in ascending powers of the electric field [4-5]

$$\hat{P}_j(\omega) = \epsilon_0 \sum_{n=1}^{\infty} (2\pi)^{-n+1} \int \cdots \int_{\omega=\omega_1+\cdots+\omega_n} \sum_{k_1, \dots, k_n} \chi_{jk_1 \dots k_n}^{(n)}(\omega; \omega_1, \dots, \omega_n) \times \delta(\omega_1 + \cdots + \omega_n - \omega) \hat{E}_{k_1}(\omega_1) \cdots \hat{E}_{k_n}(\omega_n) d\omega_1 \cdots d\omega_n, \quad (1)$$

where $\hat{\mathbf{E}} = \hat{E}_1\mathbf{u}_1 + \hat{E}_2\mathbf{u}_2 + \hat{E}_3\mathbf{u}_3$ is the macroscopic electric field in the medium, $\chi_{jk_1 \dots k_n}^{(n)}$ is the order- n susceptibility of the medium, and the delta function in the integrand picks out the combination of frequencies $\omega = \sum_{r=1}^n \omega_r$. The macroscopic field $\hat{\mathbf{E}}$ is related to the polarization density $\hat{\mathbf{P}}$ via the Maxwell equations, giving rise to the inhomogeneous wave equation

$$\nabla^2 \hat{\mathbf{E}} - \nabla(\nabla \cdot \hat{\mathbf{E}}) + \frac{\omega^2}{c^2} \hat{\mathbf{E}} = -\frac{\omega^2}{\epsilon_0 c^2} \hat{\mathbf{P}} \quad (2)$$

in the frequency domain, where the caret $\hat{}$ represents Fourier transformation of the time-domain variable:

$$f(t) = (2\pi)^{-1} \int_{-\infty}^{\infty} \hat{f}(\omega) \exp(-i\omega t) d\omega. \quad (3)$$

The polarization generally has a part linear in the electric field:

$$\hat{P}_j^{\text{linear}}(\omega) = \epsilon_0 \sum_{k=1}^3 \chi_{jk}^{(1)}(\omega; \omega) \hat{E}_k(\omega).$$

When this linear term is transferred to the left side of (2), the result in component form is

$$\begin{aligned} \nabla^2 \hat{E}_j(\omega) - \partial_j \sum_{k=1}^3 \partial_k \hat{E}_k(\omega) + \frac{\omega^2}{c^2} \left[\hat{E}_j(\omega) + \sum_{k=1}^3 \chi_{jk}^{(1)}(\omega; \omega) \hat{E}_k(\omega) \right] \\ = -\frac{\omega^2}{\epsilon_0 c^2} \hat{P}_j^{\text{nl}}(\omega), \end{aligned} \quad (4)$$

where the source term $\hat{P}_j^{\text{nl}}$ represents the polarization density arising from terms in (1) with $n > 1$.

The linear operator acting on $\hat{\mathbf{E}}$ on the left side of (4) is diagonalized in a uniform medium by plane waves such as

$$\hat{\mathbf{E}} = \mathbf{e} \exp(i\beta z), \quad (5)$$

where β is the propagation coefficient of the plane wave, $\mathbf{e}$ is a unit vector representing the eigenpolarization direction, and it is assumed that z denotes the direction of propagation of the plane wave. In general, there are two independent eigenpolarizations for any fixed direction of propagation, and these are denoted by the subscript $j = 1, 2$. Generally, the propagation coefficients β_1 and β_2 are different, a phenomenon known as *birefringence*. If the medium is weakly anisotropic, it is a good approximation to take the two eigenpolarizations $\mathbf{e}_1$ and $\mathbf{e}_2$ to be orthogonal and perpendicular to the direction of propagation.

For planewave propagation that is slowly varying along the z direction and involves only continuous waves with fixed carrier frequencies and constant (stationary) envelopes in time, we introduce the representation

$$\hat{\mathbf{E}}(x, y, z, \omega) = \frac{1}{2} \sum_{m \in \mathbf{Z}} \hat{\mathbf{E}}_m(x, y, z) \delta(\omega - \Omega_m) e^{i\kappa_m z}, \quad (6)$$

$$\hat{\mathbf{P}}(x, y, z, \omega) = \frac{1}{2} \sum_{m \in \mathbf{Z}} \hat{\mathbf{P}}_m(x, y, z) \delta(\omega - \Omega_m) e^{i\kappa_m z}, \quad (7)$$

where it is assumed that there are several waves present, with harmonic frequencies $\Omega_m = m\Omega$, ($m = 0, \pm 1, \pm 2, \dots$), and reference propagation coefficients κ_m which are in principle arbitrary, but may be chosen to have some convenient properties, such as phase matching. In the case of a linearly birefringent medium, the

Cartesian directions denoted by indexes $j, k = 1, 2$ are chosen to be aligned along the eigenvectors of the linear susceptibility tensor $\chi_{jk}^{(1)}$ on the transverse indexes $j, k = 1, 2$. We allow each eigenpolarization to have its own propagation coefficient in (6) and (7), so κ_{mj} belongs to eigenpolarization j at frequency $\Omega_m = m\Omega$. The phase-matching condition for the propagation coefficients is

$$\kappa_{mj} = \sum_r \kappa_{m_r j_r}, \quad \text{if } m = \sum_r m_r.$$

The final approximate coupled wave equation for n -wave interactions, including linear birefringence, is then

$$2i\beta_{mj}(\partial_z - i\Delta\beta_{mj})\hat{E}_{mj} = -\frac{m^2\Omega^2}{\epsilon_0 c^2}\hat{P}_{mj}^{\text{nl}}, \quad (j = 1, 2), \quad (8)$$

where β_{mj} are the true propagation constants for plane wave propagation in the linear anisotropic medium with susceptibility $\chi^{(1)}(\Omega_m; \Omega_m)$, and $\Delta\beta_{mj} = \beta_{mj} - \kappa_{mj}$ represents the deficits between the true propagation coefficients β_{mj} and the reference propagation coefficients κ_{mj} . In (8) we have neglected longitudinal field components (weak birefringence) and used the *slowly varying amplitude* approximation.

For *guided-wave* propagation, the same approximation (8) applies, but the fields E_{mj} are interpreted as the complex amplitudes of modes in the waveguide, and the indexes $j, k = 1, 2$ refer to principal polarizations of these modes, such as quasi-TE and quasi-TM in an integrated semiconductor rib-loaded guide.

Figure 1 illustrates a typical form of rib waveguide fabricated in materials based on gallium arsenide. The crystallographic axes are shown as well as a reference Cartesian coordinate system (x, y, z) . The z axis is specified as the propagation direction of the guided waves, with quasi-TM modes having a dominant electric field along the x direction and quasi-TE modes having a dominant electric field along the y direction.

3 Quantum wells

Quantum wells are specially stratified semiconducting materials consisting of thin layers of a low-bandgap material interleaved with layers of a higher-bandgap material. The effect of interleaving the low-bandgap region with a higher-bandgap medium is to create one-dimensional potential wells for carriers (electrons or holes) in the conduction and valence bands. Electrons or holes can be trapped in these potential wells with respect to the direction perpendicular to the layer planes, while remaining free to move in the directions parallel to the layer plane. An electron or hole with a definite in-plane wavevector $\mathbf{k}_{\parallel}$ has certain discrete energy levels $E_s(\mathbf{k}_{\parallel})$ due to the trapping effects of the potential well, as well as a continuum of energy states $E(\mathbf{k}_{\parallel}, k_z)$ that describe unbound particle states moving freely above

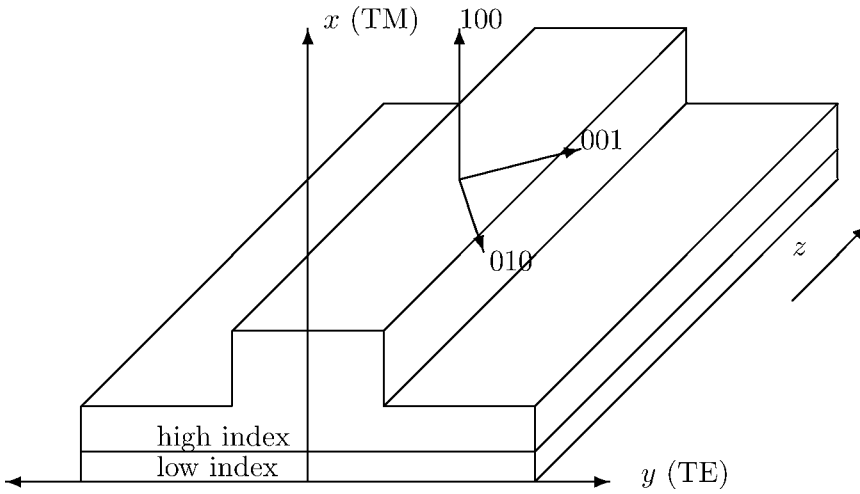


Figure 1 Geometry of an optical rib waveguide.

the potential barriers. The presence of the discrete states in the energy spectrums of the particle states has a significant effect on the electromagnetic susceptibilities of the composite quantum-well medium, since they correspond to resonances that are not present in a pure bulk medium.

The nonlinear susceptibilities of both bulk mediums and quantum wells can be computed by systematic procedures from fundamental quantum-mechanical models of the electron-hole band structure. For the second-order nonlinear susceptibility $\chi^{(2)}$ the results are

$$\begin{aligned} \chi_{jk_1k_2}^{(2)}(\omega; \omega_1, \omega_2) = & -\frac{e^3}{2\epsilon_0 m^3 \hbar^2} \frac{i}{\omega \omega_1 \omega_2} \\ & \times \sum_{b,c} \int_{\text{BZ}} \frac{\Xi_{j ab} \Xi_{k_1 bc} \Xi_{k_2 ca}}{(\Omega_{ba} - \omega_1 - \omega_2)(\Omega_{ca} - \omega_2)} d^3 \mathbf{k} \\ & + \text{perm}\{(j, -\omega), (k_1, \omega_1), (k_2, \omega_2)\} \end{aligned} \quad (9)$$

for a bulk material, and

$$\begin{aligned} \chi_{jk_1k_2}^{(2)}(\omega; \omega_1, \omega_2) = & -\frac{e^3}{2\epsilon_0 m^3 L \hbar^2} \frac{i}{\omega \omega_1 \omega_2} \\ & \times \sum_{s_1, s_2} \int_{\text{BZ}_2} \frac{\Xi_{j s s_1} \Xi_{k_1 s_1 s_2} \Xi_{k_2 s_2 s}}{(\Omega_{s_1 s} - \omega_1 - \omega_2)(\Omega_{s_2 s} - \omega_1)} d^2 \mathbf{k}_{\parallel} \\ & + \text{perm}\{(j, -\omega), (k_1, \omega_1), (k_2, \omega_2)\} \end{aligned} \quad (10)$$

for a quantum well. Here the notation “perm S ” means “permute the objects in the set S and sum the results of all permutations,” and represents the total per-

mutation symmetry required by all nonlinear susceptibilities [4]. The ground state is assumed to be numbered 1, and $a = 1$ (or $s = 1$), so $\Omega_{ba} \geq 0$ ($\Omega_{s's} \geq 0$) for all levels b (s'). The symbol BZ represents the three-dimensional first Brillouin zone of the bulk semiconductor, and BZ₂ represents the two-dimensional Brillouin zone of the quantum-well structure. The symbol $\mathcal{E}_{jab}$ represents the momentum matrix element for the spatial j -component between the energy states labelled a and b . These matrix elements have imaginary values and are antisymmetric in the energy indexes. Generally, the transition frequencies and the matrix elements are $\mathbf{k}$ -dependent in the bulk case and $\mathbf{k}_{\parallel}$ -dependent in the quantum-well case. The parameter L is the inter-well separation of a multiple quantum well (MQW), assumed to be large enough that the bound states of the wells are not coupled, but still much smaller than the optical wavelength.

The significant feature of the expressions (9) and (10) is the zeroes that occur in the denominators of the integrands when various combinations of frequencies equal interlevel transition frequencies, corresponding to resonant excitations of the medium. These resonances result in very strong nonlinear responses to waves whose temporal spectrums are concentrated near the resonant frequencies. The discrete quantum-well transition frequencies $\Omega_{s's}$ are dependent on the depth and width of quantum wells, and can therefore be engineered by variations of these two geometrical parameters of the wells.

A fundamental difference between bulk and quantum-well semiconductor mediums is the fact that the quantum-well breaks the symmetry of the bulk medium. Therefore, it is possible for quantum-well systems to exhibit nonvanishing tensor components of nonlinear susceptibilities that would be required to vanish by symmetry constraints in the bulk medium.

One method of altering the depth of quantum wells after they have been fabricated by epitaxial growth is called *quantum-well intermixing*. In this method, vacancies are generated by some means in the grown material. After annealing at a high temperature, these vacancies assist the diffusion of concentration differences through the material; as a consequence, the entire material tends over a long diffusion time to revert to a single bulk medium as the differences in atomic concentrations at boundaries between different semiconductors in the quantum-well composite are equalized. The effect on the quantum well over time is a reduction of depth and an increase of width of the potential well as the composition of the well becomes more like that of the surrounding barrier regions. These changes tend to increase the resonant transition frequencies as the anneal time increases. For many applications of nonlinear optics, the optical frequencies are chosen to lie in the transparency region for which photon energies are less than the lowest interband transition frequency; it follows that an increase of the lowest transition frequency due to intermixing of the quantum well will tend to make the interaction of the electromagnetic wave with the medium less resonant, with an attendant decrease in the $\chi^{(n)}$ susceptibility tensor coefficients. Also, intermixing tends to remove any symmetry-breaking tensor coefficients, as they cannot exist in the bulk medium that is the target of the intermixing process at infinite anneal time.

There are many ways that have been demonstrated to induce the vacancies required to promote high-temperature intermixing of the quantum wells, including [1] (i) energetic ion implantation through the planar surface of the epitaxial semiconductor; (ii) laser absorption in the quantum-well region; (iii) preferential removal, at the surface, of one atomic species from a compound semiconductor. In all of these methods, stopping the thermal annealing process at a finite time results in reduction of the susceptibility tensor coefficients as the contributions to these components from the discrete quantum-well energy states are diminished.

All of the afore-mentioned intermixing processes can be spatially controlled over the extent of a semiconductor wafer by appropriate masking techniques, so that it is possible to intermix the quantum wells in parts of the semiconductor, but to leave them intact in other parts. This permits a technology of spatial patterning of the linear and nonlinear susceptibilities due to the differences in transition resonance that occur between quantum-well and intermixed quantum-well regions. One such application of intermixed quantum wells with spatial patterning is described later in the realization of periodically patterned gratings to achieve the essential feature of phase-matching for efficient second-order interactions.

4 Second-order quasi-phase-matching

When only $n = 2$ terms are present in the nonlinear polarization expansion, the interaction is said to be *second-order*. Second-order nonlinearities are responsible for second-harmonic generation (SHG), three-wave mixing, optical rectification and many other useful effects such as the second-order cascade nonlinearity.

A fundamental aspect of second-order interactions concerns the topic of *phase-matching*. In a dispersive medium such as a semiconductor, the linear propagation coefficients are not phase-matched, meaning that the sum of the propagation coefficients for the two waves interacting at the fundamental frequency is not equal to the propagation coefficient of the generated second harmonic wave. Deviation from phase-matching reduces the efficiency of parametric interactions very sharply, and it is highly desirable to engineer phase-matching to restore the efficiency.

The crystal symmetry of gallium-arsenide is cubic. Crystals with this symmetry are linear isotropic but have a single second-order nonlinear susceptibility tensor $\chi_{xyz}^{(2)}$ [4]. This means that second-harmonic radiation will only be produced polarized along the 100-axis if a fundamental signal is present along *both* axes 010 and 001 (see Figure 1). The usual cleavage planes of gallium-arsenide are {110}-planes. Hence, if x denotes the direction of growth of the quantum-well superlattice, and yz denotes the plane of the wells, cleaved facets will be at 45 degrees with respect to the crystallographic 001 and 010 axes, and a waveguide stripe normal to the facets will be also at 45 degrees with respect to these axes. It then follows that quasi-TE-polarized modes, having principal direction of modal electric field parallel to the plane of the wells and parallel to the facets, possess equal components E_{010} and E_{001} resolved along the 010- and 001-directions, respectively. Under

these conditions a tensor of type $\chi_{xyz}^{(2)}$ produces second-harmonic in the quasi-TM-polarized mode, with a principal electric field along the 100-direction, from a fundamental pump source polarized in a TE-mode. On the other hand, a TM-polarized pump will produce no SHG in any polarization. Although the bulk material is linearly isotropic, a waveguide etched in the semiconductor will have different propagation coefficients for TE- and TM-modes at the same frequency, and since the material is strongly dispersive the phase-matching deficit $2\beta_{1\text{TE}} - \beta_{2\text{TM}}$ will not vanish. It is therefore necessary, for efficient SHG, that some form of phase-matching be devised.

In principle, a method of (partially) achieving phase-matching is the method of *quasi-phase matching* (QPM), in which the active second-order nonlinear susceptibility is spatially modulated along the propagation direction of the waveguide, with a spatial period p given by $2\pi/p = |2\beta_{1\text{TE}} - \beta_{2\text{TM}}|$. In order to illustrate how this method works, let us assume a step-wise modulation of the second-order tensor coefficient

$$\chi^{(2)} = \begin{cases} \chi_1, & np < z < \left(n + \frac{1}{2}\right)p; \\ \chi_2, & \left(n + \frac{1}{2}\right)p < z < (n+1)p. \end{cases} \quad (11)$$

It is useful to bring the equations (8) to a standard normalized form by means of appropriate scalings. To this end, a length scale L is introduced, and the propagation distance z is normalized by the transformation $z/L \rightarrow z$. Further, let

$$\psi_1 = \frac{\chi\beta_1 L}{2n_1^2} E_{1\text{TE}}, \quad \psi_2 = \frac{\chi\beta_1 L}{2n_1^2} E_{2\text{TM}} \quad (12)$$

be normalized field quantities, where $n_1 = \beta_1 c_0 / \Omega$ is the effective refractive index of the fundamental mode, Ω is the fundamental frequency, and $\chi = |\chi_1 - \chi_2|$. The propagation equations (8) can then be manipulated into the form

$$-i\partial_z \psi_1 = f(z) \psi_2 \psi_1^*, \quad (13)$$

$$-i\partial_z \psi_2 = f(z) \psi_1^2 + \Delta \psi_2, \quad (14)$$

where f is the piecewise constant function

$$f(z) = \begin{cases} \frac{1}{2}(a+1), & np < z < \left(n + \frac{1}{2}\right)p; \\ \frac{1}{2}(a-1), & \left(n + \frac{1}{2}\right)p < z < (n+1)p, \end{cases} \quad (15)$$

$a = (\chi_1 + \chi_2)/|\chi_1 - \chi_2|$, and $\Delta = (\beta_{2\text{TM}} - 2\beta_{1\text{TE}})L$ is a normalized mismatch parameter. Neglecting depletion of the fundamental pump field by the approximation

$\psi_1 \sim 1$ permits the equation for the second-harmonic amplitude ψ_2 to be integrated immediately to give

$$\psi_2(z) = i \exp(i\Delta z) \int_0^z f(z') \exp(-i\Delta z') dz', \quad (16)$$

with an initial condition $\psi_2(0) = 0$. With $f(z)$ given by (15) this evaluates at $z = Np$, an integer number of periods, to

$$\begin{aligned} \psi_2(Np) = & \frac{1}{2} i(a+1) \sum_{n=0}^{N-1} \exp[i\Delta(N-n)p] \int_0^{p/2} \exp(-i\Delta z') dz' \\ & + \frac{1}{2} i(a-1) \sum_{n=0}^{N-1} \exp[i\Delta(N-n)p] \int_{p/2}^p \exp(-i\Delta z') dz'. \end{aligned} \quad (17)$$

After some manipulation this becomes

$$\begin{aligned} \psi_2(Np) = & \frac{1}{2} i(a+1) K \exp(i\Delta Np) \\ & + \frac{1}{2} i(a-1) K \exp(i\Delta Np) \exp\left(-\frac{1}{2} i\Delta p\right), \end{aligned} \quad (18)$$

with

$$\begin{aligned} K = & \sum_{n=0}^{N-1} \exp(-i\Delta np) \int_0^{p/2} \exp(-i\Delta z') dz' \\ = & \frac{1 - \exp(-iN\Delta p)}{1 - \exp(-i\Delta p)} \frac{\exp\left(-\frac{1}{2} i\Delta p\right) - 1}{-i\Delta} \\ = & \frac{1}{\Delta} \frac{\sin\left(\frac{1}{2} N\Delta p\right)}{\cos\left(\frac{1}{4} \Delta p\right)} \exp\left[-\frac{1}{2} i\left(N - \frac{1}{2}\right) \Delta p\right]. \end{aligned} \quad (19)$$

In the special case $\Delta p = 2\pi$, corresponding to exact quasi-phase-matching when the period of the QPM grating is equal to the beat period of the phase mismatch, the coefficient K achieves its maximum value

$$K = -2i \frac{N}{\Delta}. \quad (20)$$

The conclusion of this calculation is that the intensity of the second-harmonic wave after propagating an integral number, N , of periods of the QPM grating is increasing with N^2 under exact phase-matching conditions.

A method of engineering second-order phase-matching using the technology of *quantum-well intermixing* (QWI) has recently been successfully demonstrated experimentally [6]. When a quantum-well region is intermixed it reverts to a bulk semiconductor, the interband transition energy being increased. Since the second-order susceptibilities depend on the relation of the photon energy to the bandgap energy through resonant denominators, a quantum well is closer to resonance with the optical carrier than is the bulk medium, and consequently the quantum well exhibits higher $\chi^{(2)}$ values than the bulk medium. It has been predicted theoretically [7-8] that quite large changes in $\chi^{(2)}$ of the order 100 pm V^{-1} can be achieved by intermixing a superlattice in this way. In addition, the QWI technology can be applied on a selective area basis on a semiconductor wafer, so that a periodic grating in the $\chi^{(2)}$ can be written into the material. A waveguide supporting modes at both the fundamental and second harmonic carriers is then etched into the wafer with its axis parallel to the lattice vector of the grating. The lattice period p of the $\chi^{(2)}$ grating is chosen to exactly compensate the phase-matching deficit of the relevant propagation coefficients, $2\pi/p = |\beta_{2\text{TM}} - 2\beta_{1\text{TE}}|$.

In experimental work [6] it has been verified that quasi-phase-matching gratings with modulation periods in the range $10\text{--}12 \text{ }\mu\text{m}$ will phase-match SHG from a fundamental wavelength $\sim 1550 \text{ nm}$ with second-harmonic at $\sim 775 \text{ nm}$. The planar waveguide consisted of a $0.6 \text{ }\mu\text{m}$ GaAs:AlAs 14:14 monolayers superlattice core embedded in a cladding of $\text{Al}_{0.6}\text{Ga}_{0.4}\text{As}$, the whole structure grown by molecular beam epitaxy on a GaAs substrate. The quasi-phase-matching of the $\chi^{(2)}$ grating was verified by demonstrating significant TE- to TM-polarized SHG but no TM-to TE-polarized SHG, and by exhibiting the resonance of the SHG generation as the fundamental pump wavelength was tuned through the wavelength at which the SHG was optimally phase-matched.

5 Third-order nonlinearity

Third-order nonlinearities arise from the $n = 3$ term of the nonlinear polarization expansion (1). Some third-order interactions, such as self-focusing or temporal soliton formation, are automatically phase-matched and no additional engineering is required to achieve phase-matching. These self-phase-matched interactions are typified by the optical Kerr effect, in which the intensity of light in a medium changes its refractive index.

Strong Kerr-effect interactions take place in bulk gallium-arsenide, with some enhancement obtained from the confinement of the electron in a quantum well or superlattice. Quantum-well materials exhibit much stronger exciton resonances than bulk materials, because the binding energy of the quantum-well exciton is greater than its bulk counterpart. The exciton is a discrete electron state with energy just below the valence-conduction bandgap, and has atomlike behavior when interacting with electromagnetic fields. Because the quantum-well exciton binding energy is greater, it requires more energy to thermalize into the conduction band;

consequently, it can interact more strongly with the electromagnetic field. A variety of near-resonant and resonant nonlinear optical interactions can be conducted with the QW exciton.

In addition to the enhanced exciton resonances of quantum wells, the technology of QWI again serves as a method of spatially delineating separate regions of nonlinear and linear behavior, by the same method that was discussed already for the case of second-order interactions. The intermixed regions have local bandgap energies larger than the nonintermixed regions; for photon energies lower than the nonintermixed bandgap, the nonlinear coefficients $\chi^{(3)}$ in the nonintermixed regions are larger than those in the intermixed regions because the former are closer to resonance with the band edge than the latter. Thus, by QWI it is possible to integrate laser, nonlinear and linear passive regions all on the same planar substrate, leading to high functionality of the integrated devices.

Many components for integrated optical systems have been studied by many different research groups. Of particular importance are saturable absorbers for use in modelocked lasers. In this case, the behavior of the material required for the absorber region is quite different to that of the gain medium. If the laser is to be a quantum-well laser, then the absorber and any passive waveguide sections have to be integrated into the same material system. It is not strictly accurate to describe saturable absorption as a third-order effect since it is a resonant effect and the polarization expansion (1) does not apply on-resonance due to the singular behavior of the susceptibilities at resonance. However, there is much experimental evidence, and some theoretical explanation, that refractive rather than absorptive mechanisms may play a significant role in modelocking on ultrashort timescales (<100 fs) [2-3], [9], and these appear to be of the Kerr-effect type.

The full set of mathematical expressions for paraxial guided-wave propagation in a planar waveguide with third-order self-phase-matched interactions is of the form [10]

$$-i\partial_z\psi_\nu = \Delta_\nu\psi_\nu + \frac{1}{2}\partial_x^2\psi_\nu - \frac{1}{2}d_\nu\partial_\tau^2\psi_\nu + \sum_{\nu_1,\nu_2,\nu_3} \alpha_{\nu\nu_1\nu_2\nu_3}\psi_{\nu_1}\psi_{\nu_2}^*\psi_{\nu_3}, \quad (21)$$

along with the complex conjugates of the same equations. In these equations, z and τ are scaled propagation distance and time, respectively; x is a scaled coordinate for the lateral dimension in the plane; the indexes $\nu = 1, 2$ refer to two orthogonal polarizations of modes in a waveguide; ψ_ν are the normalized complex amplitudes of the modes; Δ_ν represents linear birefringence of the modes; d_ν represents temporal dispersion parameters; and the α -coefficients are normalized tensor components that nonlinearly couple the two polarizations together in various ways.

The equations (21) exhibit a large range of dynamical effects, including spatial solitons, stable and unstable nonlinear polarization states that are different from the linear ones, intensity-dependent modulation instabilities of stationary states, and polarization rotation by change in intensity [11]-[13]. In addition, Faraday rotation of the mode polarization in the nonlinear case has been studied [14].

When combined, these effects are enough to produce devices that will switch between two stable stationary polarization states of the waveguide under control by an external magnetic field *via* the magneto-optical effect. At the present state of technology the optical intensities required are quite high, and we do not yet know of good ways to incorporate strong magnetic materials in the gallium-arsenide epitaxial growth system.

6 Conclusions

Semiconductor quantum wells and superlattices are good metamaterials for the realization of nonlinear optical functions of the kind required in nonlinear signal processing for all-optical networks. They can be grown by epitaxial growth techniques and can be spatially altered by techniques such as quantum-well intermixing to produce laterally structured devices for monolithic all-optical integration, including various nonlinear functions as well as laser sources. Much work remains to be done to realize the prospects of these metamaterials as actual devices with optimized performance, but this will happen as the basic physics and materials engineering become better understood.

Acknowledgments

Work referred to at the University of Glasgow on nonlinear interactions in quantum-well mediums has been supported by the Engineering and Physical Science Research Council of the United Kingdom, and by the European Community. This is the work of my colleagues in the Optoelectronics Research Group: J.H. Marsh, A.C. Bryce, C.N. Ironside, R.M. De La Rue, P.J.R. Laybourn, D.C. Hutchings, C.R. Stanley, J.S. Aitchison (now at the University of Toronto, Canada), T.F. Krauss (now at the University of St. Andrews, UK) and E.A. Avrutin (now at the University of York, UK).

References

- [1] J.H. Marsh, Quantum well intermixing, *Semicond Sci Technol* **8** (1993), 1136–1155.
- [2] D.A. Yanson, M.W. Street, E.A. Avrutin, S.D. McDougall, I.G. Thayne and J.H. Marsh, Passive harmonic modelocking in monolithic compound-cavity laser diodes, *Electron Lett* **36** (2000), 1930–1931.
- [3] I.D. Jung, F.X. Kartner, N. Matuschek, D.H. Sutter, F. Morier-Genoud, G. Zhang, U. Keller, V. Scheuer, M. Tilsch and T. Tschudi, Self-starting 6.5 fs pulses from a Ti:sapphire laser, *Opt Lett* **22** (1997), 1009–1011.
- [4] P.N. Butcher and D. Cotter, *The elements of nonlinear optics*, Cambridge University Press, Cambridge, UK, 1990.
- [5] N. Bloembergen, *Nonlinear optics, 4th ed*, World Scientific, Singapore, 1996.

- [6] A. Saher Helmy, D.C. Hutchings, T.C. Kleckner, J.H. Marsh, A.C. Bryce, J.M. Arnold, C.R. Stanley, J.S. Aitchison, C.T.A. Brown, K. Moutzouris and M. Ebrahimzadeh, Quasi-phase-matching in GaAs-AlAs superlattice waveguides via bandgap tuning using quantum well intermixing, *Opt Lett* **25** (2000), 1370–1372.
- [7] D.C. Hutchings, Electronic bandstructure of disordered superlattices, *Superlatt Microstruct* **26** (1999), 195–209.
- [8] D.C. Hutchings, Modulation of the second-order susceptibility in GaAs/AlAs superlattices, *Appl Phys Lett* **76** (2000), 1362–1364.
- [9] V.P. Kalosha, M. Muller and J. Hermann, Theory of solid-state laser mode locking by coherent semiconductor quantum-well absorbers, *J Opt Soc Am B* **16** (1999), 323–338.
- [10] J.M. Arnold, Anisotropic effects in nonlinear optical waveguides, *Wave Motion* **34** (2001), 339–352.
- [11] J.S. Aitchison, D.C. Hutchings, J.M. Arnold, J.U. Kang, G.I. Stegeman, E. Ostrovskaya and N. Akhmediev, Power-dependent polarization dynamics of mixed-mode spatial solitary waves in AlGaAs waveguides, *J Opt Soc Am B* **14** (1997), 3032–3037.
- [12] D.C. Hutchings, J.S. Aitchison and J.M. Arnold, Nonlinear refractive coupling and vector solitons in anisotropic cubic media, *J Opt Soc Am B* **14** (1997), 869–879.
- [13] D.C. Hutchings, J.M. Arnold and J.S. Aitchison, Theory of mixed polarization propagation in anisotropic media, *Opt Quant Electron* **30** (1998), 771–782.
- [14] D.C. Hutchings and J.M. Arnold, Modification of the linear and nonlinear polarization dynamics in semiconductors owing to the Kerr magneto-optic effect, *J Opt Soc Am B* **17** (2000), 1774–1780.



John M. Arnold is Professor of Applied Electromagnetics in the Department of Electronics and Electrical Engineering at the University of Glasgow, where he is presently Head of Department. He is a Fellow of the Institute of Physics. He has published extensively on the theory of electromagnetic and acoustic wave propagation, and has contributed to the development of many novel optoelectronic devices through the application of theory and computational modeling. His present research interests are concentrated on the nonlinear dynamics of optoelectronic devices, particularly semiconductor lasers, and in soliton phenomena in nonlinear guided-wave optics.

Organic Thin-Film Photorefractive Materials

Partha P. Banerjee

Abstract

The essential properties of photorefractive polymers are reviewed and their use in two-beam coupling, edge enhancement of images and image correlation are demonstrated. The advantages of using photorefractive polymers over photorefractive crystals, such as ease of fabrication and low cost, are discussed.

1 Introduction

The photorefractive (PR) effect is a nonlinear optical (NLO) phenomenon caused by an induced intensity-dependent refractive index in a doped material that is also an electro-optic (EO) photoconductor [1]. Under nonuniform illumination conditions, photogenerated charge carriers in a doped photoconductor migrate from the bright regions and subsequently become trapped in darker regions of the material. This charge displacement results in the buildup of a space-charge field that can induce refractive index changes in a photoconductor that also has an EO effect. A nonlinear intensity-dependent refractive index is thus achieved [2].

Another possible transport mechanism is light-assisted tunneling, or hopping, directly between trapping sites. Whatever the transport mechanism, if the charges are illuminated continuously, they eventually rearrange themselves under equilibrium conditions. If the crystal lacks inversion symmetry, then the electrostatic field surrounding each charge can cause a large, first-order change in the index of refraction of the crystal. Some of the most useful PR crystals, such as barium titanate (BaTiO_3), bismuth tin oxide (BTO) and lithium niobate (LiNbO_3), are ferroelectrics whose inversion symmetry has been removed by poling [3]. Poling is a process to reorient the axes of the crystal. This process requires the application of an electric field above a certain critical temperature and cooling the crystal in the presence of the applied electric field.

An interesting characteristic of PR materials is that if the crystal is a good insulator in the dark, its optical nonlinearity is critically dependent on the intensity of the optical beams, thereby permitting spectacular nonlinear effects even with weak laser beams. The optical intensity determines the speed and the magnitude of the nonlinearity [4].

The movement of photoexcited free carriers can be effected by the following three different mechanisms:

1. **Diffusion:** It is the movement of charges due to the gradient in the concentration of optically created free charges, such as electrons, in the material. Where light intensity is higher, more free charges are created, and these diffuse toward regions where the light intensity, and hence, free charge concentration, is lower.
2. **Drift:** In this case, the displacement of free carriers is achieved by an externally applied static electric field.
3. **Photovoltaic effect:** In some PR crystals, uniform illumination of light can generate open-circuit voltage or short-circuit current [5].

These charge-transport (CT) mechanisms lead to the physical motion of charges in the material, usually over a distance on the order of micrometers. This motion gives rise to a spatial phase shift between the incident light intensity pattern and the refractive index modulation. An important consequence of this phase shift is an energy transfer between two light beams interfering in a PR medium, called

asymmetric two-beam coupling (TBC). If the coupling is sufficiently strong, TBC gain may exceed the absorption and reflection losses of the sample, and optical amplification can occur. This cannot occur with any local mechanism of refractive index change (unless the material or the light intensity pattern is translated at a particular rate) [6].

In this chapter, readers are first introduced to PR polymers and their advantages in Section 2. Thereafter in Section 3, we briefly summarize various constituents of PR polymers and how they can be engineered to give optimum PR effects. In Section 4, we review TBC and its modeling in a photorefractive polymer. The effects of beam fanning are included in the discussions. In Sections 5 and 6, we demonstrate two applications of these polymers, viz., in edge enhancement of images and in correlation of two images.

2 Photorefractive polymers

There are several reasons for pursuing the development and applications of organic PR materials. One motivation comes from the consideration of a particular figure-of-merit that compares the index change possible in different materials (assuming equal densities of trapped charges). This figure-of-merit may be defined as $Q_e = n^3 r_e / \epsilon_r$ [7], where n is the optical index of refraction, r_e is the effective EO coefficient, and ϵ_r is the dc permittivity relative to the permittivity of free space ϵ_0 .

Table 1 [8-9] shows a comparison of the measured Q_e values of inorganics along with estimated possible values for Q_e in organic polymers. In this table, $\text{Bi}_{12}\text{SiO}_{20}$ is bismuth silicon oxide (BSO), GaAs is gallium arsenide, KNbO_3 is potassium niobate and $(\text{Sr}, \text{Ba})\text{Nb}_2\text{O}_6$ is strontium barium niobate (SBN). For inorganics, Q_e is limited by usually high values of the relative permittivity.

Doped polymeric PR composites are particularly attractive, because all important functionalities of the PR effect can be optimized simultaneously and almost independently to a degree not possible in the existing inorganic materials. The most important specific advantages of the multicomponent composites approach are as follows:

- i) The active wavelength of the material can be tuned to the desired regimes of operation by incorporating appropriate sensitizers for charge photogeneration.
- ii) The CT and trapping characteristics in such materials can also be optimized using dopants with desired ionization energy for charge transfer.
- iii) Polymeric materials can be processed into different applicational device forms.
- iv) Optical quality and linear refractive index can be optimized.
- v) The process of synthesizing PR polymers is relatively easy.
- vi) The cost of fabrication of PR polymers is much lower than that of growing doped PR crystals.

Table 1 Figure-of-merit Q_e for inorganic and organic materials.

Material	r_e (pm/V)	n	ϵ_r	Q_e (pm/V)
Bi ₁₂ SiO ₂₀	5	2.54	56	1.5
GaAs	1.43	3.4	12	4.7
BaTiO ₃	1640	2.4	3600	6.3
LiNbO ₃	31	2.2	32	10.3
KNbO ₃	380	2.3	240	19.3
(Sr, Ba)Nb ₂ O ₆	216	2.3	750	3.5
Organic Polymers (Potential)	30	1.6	4.0	31

An example of a PR polymer is PNP:PVK:ECZ:TNF. This polymer is based on the photoconducting material poly(N-vinylcarbazole) (PVK). The EO activity is introduced into the material by doping it with (S)-(-)-N-(5-nitro-2-pyridyl)prolinol (PNP) as the optically second-order nonlinear material. N-ethylcarbazole (ECZ) is added as a plasticizing agent to lower the glass transition temperature T_g . A small amount of 2-4-7 trinitro-9 fluorenone (TNF) is added to increase the photosensitivity of the material in the visible region. The composition is prepared of PNP:PVK:ECZ:TNF at ratios of 50:33:16:1. The polymer possesses an absorption coefficient α of 25 cm^{-1} at 633 nm. The TBC coefficient Γ for the material is 101 and 68 cm^{-1} for s- and p-polarized light, respectively, at a free-space wavelength $\lambda_0 = 633 \text{ nm}$ and with an applied voltage of $100 \text{ V}/\mu\text{m}$ [10].

3 Engineering photorefractive polymers

As stated before, the necessary elements of photorefractivity are photoinduced charge generation (CG), CT, trapping, and the EO effect. Three different approaches are used to engineer the PR polymers. These approaches are discussed next [7].

3.1 Nonlinear optical polymer hosts

The first approach is to use an optically nonlinear polymer with the nonlinear dopant bonded to the polymer backbone. An alternative approach is to utilize a $\chi^{(3)}$ polymer in a biasing electric field to produce the required second-order optical nonlinearity. A $\chi^{(3)}$ material is one in which the nonlinear polarization is cubically dependent on the optical field [3]. However, with the application of a dc field, the material has an effective quadratic nonlinearity, meaning that the nonlinear polarization depends on the square of the optical field. In either case, the nonlinear host polymer is doped with CT agents in concentrations high enough to ensure bulk transport of the photogenerated charge generation sites to trap sites. The CG and trapping can be provided by the nonlinear polymer itself. Alternatively, the PR effect can be optimized by including additional dopants in relatively low concentration to perform the function of CG and trapping.

Table 2 Measured parameters for first NLO polymer class.

NLO Polymer	Sensitizer	CT agent	λ_0 (nm)	α (cm ⁻¹)	Γ (cm ⁻¹)
BisA-NPDA	None	DEH	647	10	0.33
BisA-NAT	None	DEH	676	18	1.8
BisA-NAT	None	DEH	753	2.4	1.2
NNDN-NAN	None	DEH	568	5.9	0.23
PMMA-PNA	None	DEH	647	1.1	0.11
PMMA-PNA	C ₆₀	DEH	647	3.4	0.6
PMMA-PNA	p-dci	DEH	647	40	2.2

Most of the research concentrates on two types of hosts. The first type is one in which the NLO chromophore is cross-linked into an epoxy material through multiple bonds between the epoxy chain and the NLO chromophore. The second type of polymer is a copolymer of methyl methacrylate (PMMA) and acrylate which has been derivatized with the NLO chromophore, and results in the NLO chromophore appearing as a pending group attached to the polymer backbone. Table 2 summarizes the properties of these polymers [11]–[14]. Here, BisA is diglycidyl ether, PNA is p-nitroaniline, DEH is (diethylamino) benzaldehyde diphenylhydrazone, NPDA is 4-nitro-1,2-phenylenediamine, NNDN is N,N-(diglycidyl)-4-nitroaniline, NAN is N-(2-aminophenyl) nitroaniline, C₆₀ is fullerene, p-dci is N,N'-bis-(2,5-di-tert-butylphenyl)-3,4:9,10-perylenebis (dicarboximide), and NAT is nitroaminotolan.

3.2 Charge-transporting polymer hosts

The second approach to producing a PR polymer is to use a host material consisting of a charge-transporting polymer that contains a charge-transporting group on the polymer backbone. The material is made optically nonlinear by the addition of nonlinear chromophore in as high concentration as possible to maximize the optical nonlinearity upon poling. As in Section 3.1, additional dopants can be added to provide or improve CG and trapping. To date, only two charge-transporting polymers have been used as hosts for PR polymer mixtures. The first and much more thoroughly investigated of these polymers is poly(N-vinylcarbazole) (PVK), and the second is a poly(silane) derivative.

The properties of PVK and its potential as a charge-transporting host have been studied extensively [15]. Nine NLO chromophores have been combined with PVK:TNF to produce PR polymers. The absorption and TBC coefficients of these polymers are shown in Table 3. The polymer was made photoconductive in the red and near-infrared regimes by the addition of the charge generator agent, either 2,4,7-trinitro-9-fluorenone (TNF) [9], [13] or C₆₀ [13], [16].

The second charge-transporting host polymer that has been investigated is poly[(4-n-butoxyphenyl)ethylsilane] (PBPEs) [17]. To achieve optical nonlinearity in the resulting PR mixture, either coumarin-153 (C-153) or (E)- β -nitro-(Z)- β -methyl-3-fluoro-4-(N,N-diethylamino)styrene (FDEA-MNST) is added to the

Table 3 NLO chromophores that are combined with PVK:TNF.

NLO Chromophore	Acronym	α (cm ⁻¹)	Γ (cm ⁻¹)
3-Fluoro-4-N,N-diethylamino-(E)- β -nitrostyrene	FDEANST	1.4	7.0
4-Piperidinobenzylidenemalononitrile	PDCST	2.3	7.8
2-(α -Methylbenzyl)amino-5-nitropyridine	MBANP	0.9	2.6
4-Methoxy-2'-trifluoromethyl-4'-nitrostilbene	MTFNS	0.6	1.2
4-N,N-Diethylamino-(Z)- β -methyl-(E)- β -nitrostyrene	DEANMST	10	8.0
4-N,N-Diethylamino-(E)- β -nitrostyrene	DEANST	10	5.0
4-N,N-Diethylamino-(E)-cinnamonnitrile	DEACST	11	2.1
(E,E)-1-(4-N,N-Diethylaminophenyl)-4-nitrobutadiene	DEABNB	40	3.2
1,3-Dimethyl-2,2-tetramethylene-5-nitrobenzimidazoline	DTNBI	49	5.4

Table 4 Absorption and TBC coefficients of PBPEs polymers.

Polymer	Chromophore	Sensitizer	λ_0 (nm)	α (cm ⁻¹)	Γ (cm ⁻¹)
PBPES	Coumarin-153	C60	647	14	0.25
PBPES	Coumarin-153	TNF	647	7.4	0.27
PBPES	FDEAMNST	C60	753	1	1.7

polymer as a dopant. In addition to the NLO chromophore, a small amount of either TNF or C₆₀ is included to photogenerate holes at typical wavelengths of 647 nm and 753 nm. Three specific PBPES:chromophore:sensitizer combinations have been investigated and their properties are shown in Table 4.

3.3 Fully functionalized polymers

The third approach to producing a PR polymer involves covalent attachment of all the required functionalities directly to the polymer backbone, referred to as fully functionalized polymers. The advantage of such an approach is the stability of such a single-component system against phase separation and the stability of the EO coefficient. The main disadvantages are the challenge in making fully functionalized polymers and the lack of flexibility. Also, there is the possibility of interrupting the conduction pathways in such polymers.

The first example of a fully functionalized polymer that has been tested for photorefractivity is a methacrylate-based material containing both carbazole and tricyanovinylcarbazole groups linked to the polymer by an alkylene [18]. Thin films of this polymer (1.4–3.3 μ m) have been poled at high temperature (100°C, 100 V/ μ m) and found to have both high EO coefficient ($r_{33} = 6.1$ pm/V) and high photoconductivity at a wavelength of 514 nm. Gratings have been written in these thin films at wavelengths between 550 and 600 nm [19]. However, TBC measurements have revealed them to be nonphotorefractive, photochemically induced absorption gratings.

A recent report [20] describes PR measurements on a previously discussed fully functionalized polymer [21] in which the NLO chromophore, a charge-transporting molecule, and a charge generator are all attached to the same polymer backbone. The nonlinear optical chromophore is provided by a monomer with an amino-methylsulfonyl-stilbene functionality, the hole transport is provided by a similar monomer with a diphenylhydrazone related to DEH, and the CG is provided by a third monomer to which one of the two different derivatives of the dye Foron brilliant blue is attached. To test for the presence of PR grating formation, a planar waveguide geometry is used with poling normal to the plane [20]. Two-beam coupling studies were performed at 632 nm with a 3-mm interaction length. In these studies, TBC coefficients up to 2.3 cm^{-1} were reported.

Another single-component system, a 1:1 side-chain polymer of (dialkylamino) nitrostilbene on a methacrylate backbone (DANS/MMA), was studied for photorefractivity without the addition of a special photoconducting component [22]. An EO coefficient of 20 pm/V was observed.

4 Wave mixing in photorefractive polymers

Wave mixing or TBC is an inherent property of PR materials that may give rise to energy exchange between the participating waves. In the case of PR polymers, such as the kind described at the end of Section 2, a typically 25–50- μm thick sample is sandwiched between two optically transparent indium tin oxide (ITO) electrodes and a voltage is applied across the sample. To facilitate wave mixing, the direction along which the resulting interference pattern created by two incident optical waves varies must have a component along the direction of the external electric field set up in the sample through the applied voltage. Accordingly, in a typical TBC geometry, we consider two incident waves E_1 and E_{-1} onto a polymer PR medium as shown in Figure 1. Here, θ_i and θ_b are respectively the angle between the two incident waves and the angle between the bisector to the incident waves and the normal. The interference between these two waves forms different grating orders in the material. Rescattering from primary gratings with grating vector $\mathbf{K}$, as well as scattering from the higher-order $2\mathbf{K}$ gratings may diffract the incident waves into different diffracted orders. Diffracted orders above the second are ignored here for simplicity. In some experimental setups, the -2 diffraction order may exceed the critical angle and may travel in the waveguide mode [23] similar to the fanning beam of intensity I_f . Fanning in PR materials is a common phenomenon resulting from surface scattering or imperfections inside the material and is supported by energy exchange from TBC. For accounts of experimental and theoretical work (the latter based on linearizations of the material equations), the reader is referred to Refs. 23–26.

Analysis of wave mixing in such a system is often complicated and involves coupling the material equations with the optical wave propagation equation. The material equation we start from is a decoupled nonlinear equation for the space

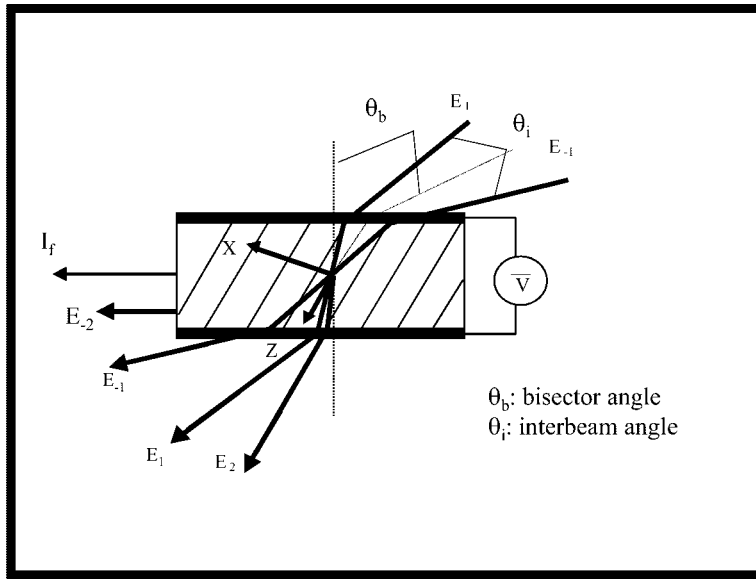


Figure 1 Two-beam coupling configuration.

charge field, assuming negligible diffusion [27]:

$$\frac{QY}{\left(1 + \frac{dY}{d\xi}\right)} = g. \quad (1)$$

In this equation,

$$\begin{aligned} Y &= \frac{E_{st}}{E_D}, & \xi &= k_D x, & Q &= 1 + \frac{I}{I_D}, \\ g &= \frac{J_c \gamma}{q \mu \nu E_D \alpha_n}, & E_D &= \frac{k_B T}{q} k_D, & k_D^2 &= \frac{e^2 N_c}{\epsilon k_B T N} (N - N_c), \\ \alpha_n &= \frac{(N - N_c)}{N_c}, & I_D &= I_b + I_d, \end{aligned} \quad (2)$$

where E_{st} is the induced electrostatic field, I is the intensity distribution, T is the absolute temperature, k_B is the Boltzmann constant, e is the electron charge, I_b and I_d are the background and dark intensities respectively, J_c is the conduction current, N and N_c are the photosensitive center and compensational charge concentrations respectively, μ is the mobility, ν is the rate of thermal excitation, ϵ is the low-frequency permittivity, and γ is the electron recombination rate.

The optical field E in the material can be written as

$$E(\xi, \zeta) = E_1(\xi, \zeta) + E_{-1}(\xi, \zeta) + E_2(\xi, \zeta) + E_{-2}(\xi, \zeta), \quad (3)$$

where $\zeta = k_D z$. We have assumed all optical fields to be polarized in the same direction, and hence treated them as scalars for simplicity. With the notation

$$E_i(\xi, \zeta) = A_{i1} \exp[-j(\kappa_{i\xi}\xi + \kappa_{i\zeta}\zeta)], \quad i = 1, -1, 2, -2, \quad (4)$$

and

$$\kappa_{i\xi, \zeta} = \frac{k_{ix, z}}{k_D}, \quad k_{ix} = \frac{K}{2} = -k_{-ix}, \quad k_{iz} = k_{-iz}, \quad k_{ix}^2 + k_{iz}^2 = n^2 k_0^2, \quad (5)$$

where k_{ix} and k_{iz} are the x and z components of the propagation vector of field i , k_0 is the magnitude of the propagation vector of light in free space and K is the magnitude of the grating vector. The normalized electrostatic field can be assumed in the form

$$Y = Y_0 + Y_1 \exp(-j\kappa\xi) + Y_1^* \exp(j\kappa\xi) + Y_2 \exp(-j2\kappa\xi) + Y_2^* \exp(j2\kappa\xi), \quad (6)$$

where $\kappa = K/k_D$ and $j = \sqrt{-1}$. Now using

- i) the definition of induced refractive index change $\Delta n = (A/2n)Y^2 E_D^2$, where A is the constant that incorporates the effect of birefringence and the EO effect that contributes to the orientational enhancement mechanism in nonlinear chromophores in PR polymers [28],
- ii) the Helmholtz equation for optical field E ,
- iii) the Fourier expansions of Y and Q , and
- iv) extensive algebra, the slowly varying envelope approximation (SVEA) and the Ewald sphere (which helps to take into account effects of phase mismatch due to propagation of higher non-Bragg orders),

we arrive at the evolution equations for E_1, E_{-1}, E_2, E_{-2} , and the beam fanning intensity I_f . These are stated as follows:

$$\begin{aligned} \frac{dA_1}{d\xi} \approx & -j\beta \{ A_1 Y_0^2 + 2A_{-1}(Y_0 Y_1 + Y_1^* Y_2) + [2A_2(Y_0 Y_1^* + Y_1 Y_2^*) \\ & + A_{-2}(2Y_0 Y_2 + Y_1^2)] \exp(j\Delta\kappa\zeta) \} - f_1 A_1 - |\beta'_1| I_f A_1, \end{aligned} \quad (7a)$$

$$\begin{aligned} \frac{dA_{-1}}{d\xi} \approx & -j\beta \{ A_{-1} Y_0^2 + 2A_1(Y_0 Y_1^* + Y_1 Y_2^*) + [2A_{-2}(Y_0 Y_1 + Y_1^* Y_2) \\ & + A_2(2Y_0 Y_2^* + Y_1^{*2})] \exp(j\Delta\kappa\zeta) \} - f_{-1} A_{-1} - |\beta'_{-1}| I_f A_{-1}, \end{aligned} \quad (7b)$$

$$\begin{aligned} \frac{dA_2}{d\xi} \approx & -j\beta \{ A_2 Y_0^2 + [2A_1(Y_0 Y_1 + Y_1^* Y_2) \\ & + A_{-1}(2Y_0 Y_2 + Y_1^2)] \exp(-j\Delta\kappa\zeta) \}, \end{aligned} \quad (7c)$$

$$\frac{dA_{-2}}{d\zeta} \approx -j\beta \{ A_{-2}Y_0^2 + [2A_{-1}(Y_0Y_1^* + Y_1Y_2^*) + A_1(2Y_0Y_2^* + Y_1^{*2})] \exp(-j\Delta\kappa\zeta) \}, \quad (7d)$$

$$\frac{dI_f}{d\zeta} \approx 2f_1|A_1|^2 + 2f_{-1}|A_{-1}|^2 + 2(|\beta'_1||A_1|^2 + |\beta'_{-1}||A_{-1}|^2)I_f, \quad (7e)$$

where

$$\beta'_i = \frac{\Gamma_{fi}}{(1 + Q_0)(I_b + I_d)k_D},$$

Γ_{fi} is the TBC strength for a fanning beam with a plane wave labeled with a subscript i , f_1 and f_{-1} are the phenomenological fanning scattering coefficients describing the fanning from the wave labeled 1 and -1 to the fanning beam. Details can be found in Matsushita *et al.* [28] and Banerjee *et al.* [29].

The first term on the right side of (7a–d) is the phase shift term caused by the linear change in the refractive index with the intensity $I_0 = I_1 + I_{-1} + I_2 + I_{-2}$. The other terms in the brackets denote the effect of scattering of a plane wave by a particular grating. Both fundamental and higher-order induced gratings play a role in the diffraction of the input waves. For example, in (7a), the term $A_{-1}Y_0Y_1$ represents the diffraction of A_{-1} by the first-order grating, the term $A_2Y_0Y_1^*$ represents the diffraction of A_2 by the first-order grating, and the term $A_{-2}Y_0Y_2$ represents the diffraction of A_{-2} by the second-order grating. The other terms in (7a) represent more complex diffractions. The same arguments can be said about the diffraction in A_{-1} , A_2 , and A_{-2} directions represented by the different terms in (7b–d). The effects of the phase mismatch are incorporated in the complex exponential. The last term in (7a) and (7b), along with (7e), are contributions from beam fanning. Notice that if the higher-order fields and beam fanning are neglected, (7a) and (7b) are reduced to the typical TBC equations in [30].

Extensive simulations on TBC, energy exchange and higher-order generation using the model equations above in the presence of and absence of beam fanning have been performed by Matsushita *et al.* [28] and Banerjee *et al.* [29]. The following parameters are taken as constants throughout the simulation, all in SI units: sample thickness $d = 25 \mu\text{m}$, step size $= 0.001 \lambda$ (used in Euler's method), $\lambda_0 = 633 \text{ nm}$, refractive index $n = 1.7$, $T = 300 \text{ K}$, orientational enhancement parameter $A = 2.89 \times 10^{-18} \text{ m}^2 \text{ V}^{-2}$, $N_c = 1 \times 10^{22} \text{ m}^{-3}$, dark current $I_d =$ background intensity $I_b = 1 \text{ W m}^{-2}$.

The simulation results, which are supported by experimental observations, are summarized as follows:

1. The power transfer from the signal beam I_{-1} to the other three beams increases with the applied voltage V , which is also in good agreement with theoretical predictions, since the coupling coefficient Γ is an increasing function of the applied voltage.

2. At higher applied voltages, the second-order intensity is proportional to the square of the applied voltage, $I_2 = aV^2$.
3. When beam fanning is considered, the beam fanning intensity I_f increases with increasing applied voltage. As the applied voltage increases, more power is transferred from I_1 and I_{-1} to beam fanning; however, there is also some power transferred to I_1 from I_{-1} . This process continues until the net power transferred to I_1 is negative; i.e., beam I_1 loses power. As a result, the power in beam I_1 starts to decrease. The turnaround point, where power in I_1 starts to decrease, is found to be around 900 V for the sample used in experiments and simulations in Banerjee *et al.* [29].

5 Real-time edge enhancement

In this section, we briefly review the results of Banerjee *et al.* [29] on real-time programmable edge enhancement simulation using organic PR materials. The configuration used is the TBC configuration where the edge enhancement occurs in the higher diffracted (non-Bragg) order E_2 . Figure 2 shows the TBC layout to perform edge enhancement in the second-order field E_2 . The original profile represents a transverse profile placed at distance d_0 from the Fourier transform (FT) lens L_1 . The second lens L_2 is used to perform the inverse FT of the edge-enhanced image. The edge-enhanced profile represents the transverse profile of the image in the higher diffracted order.

First, a metric for the edge-enhancement efficiency is defined as $\Phi = (I_h - I_l)/I_h$, where I_h and I_l are defined in Figure 3. The edge-enhancement efficiency can have a maximum value of unity.

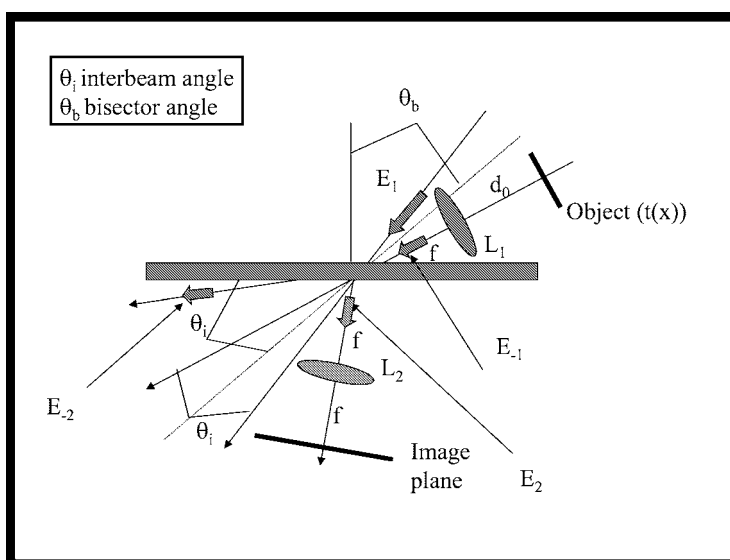


Figure 2 Edge enhancement geometry; E_2 yields the edge-enhanced image.

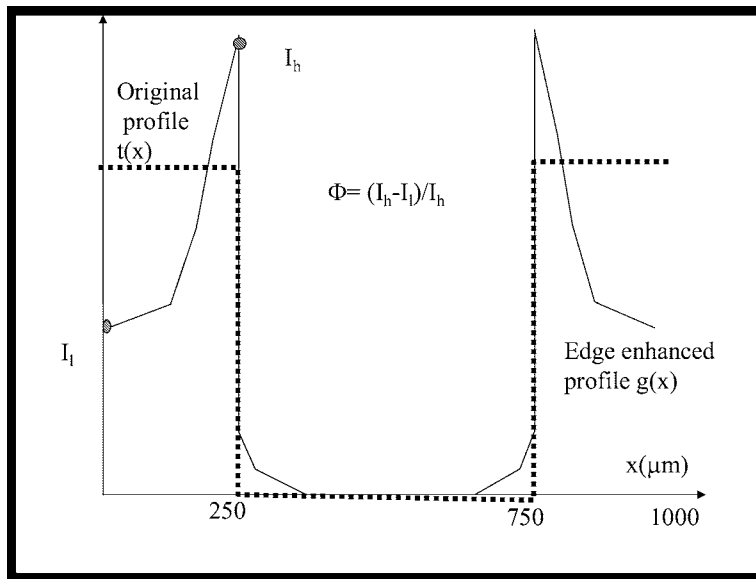


Figure 3 Edge enhancement metric Φ .

The effect of different parameters on edge enhancement in the second-order field E_2 is now investigated. The transverse profile simulates a slit of 0.5-mm width. The transverse profile is sampled at 64 points to facilitate the FFT (fast Fourier transform) operation. The optical field that carries the FT of the transverse profile interacts with the other waves in the material that is located at the back focal plane of the lens L_1 . Finally, the inverse FT of the second-order field is performed. The simulation is done on a PC with MathCad software. Euler's method is used to solve the set of evolution equations (7) for E_i .

Unless otherwise stated, the following simulation parameters are assumed for the numerical results in the remainder of this chapter: applied voltage = 1000 V, object lens focal length $f = 10$ cm, defocusing = 1 cm (will be defined later), interbeam and bisector angles (θ_i , θ_b) (shown in Figure 2) are (20° , 55°), respectively, and $I_{-1}/I_1 = 5$.

As an example, the dependences of the edge enhancement metric Φ on the defocusing $f - d_0$ and on the beam ratio I_{-1}/I_1 are studied. Accordingly, in Figure 4, we show the variation of Φ with the amount of defocusing, and with the beam ratio as a parameter. As shown, at high beam ratios the edge enhancement is caused by intensity contrast and not the defocusing. At lower beam ratios, defocusing appears to have stronger effect on the edge-enhancement metric. Note that beam fanning has a small but negative effect on edge enhancement. Edge enhancement, as defined through our metric, is therefore programmable through changing the beam ratio and the amount of defocusing.

It turns out that edge enhancement metric Φ increases with increasing applied voltage. After a critical value of the applied voltage, Φ saturates. Thereafter, any further increase in the applied voltage only increases the overall image brightness.

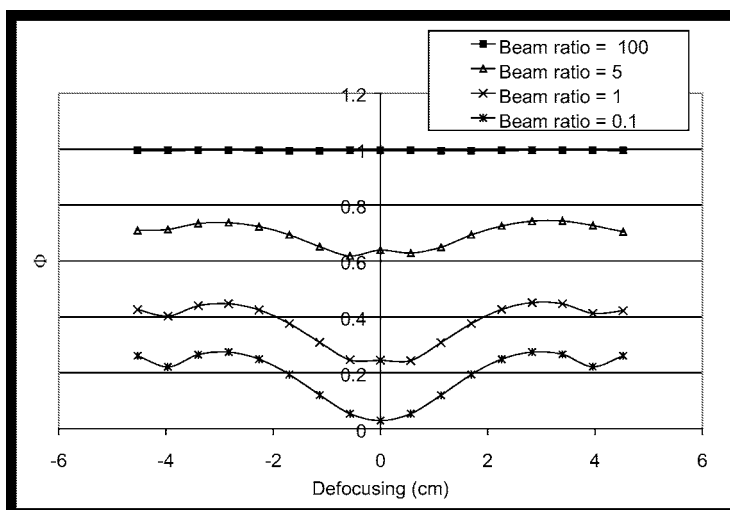


Figure 4 Effect of defocusing and beam ratio on Φ .

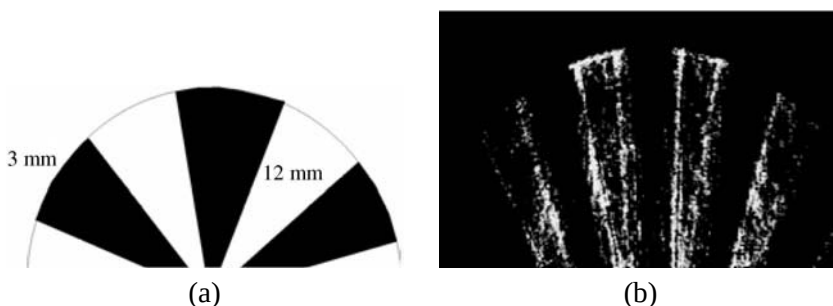


Figure 5 (a) Sketch of a checkered fan serving as an object for edge enhancement. (b) Edge enhancement for a bias voltage of 1500 V.

Simulations of edge enhancement of objects using the technique mentioned above and experiments to verify edge enhancement are detailed by Banerjee *et al.* [29].

A typical experimental result on edge enhancement is shown in Figure 5. A sketch of the original object, a checkered fan, is shown in Figure 5(a). Figure 5(b) shows edge enhancement at an applied voltage of 1500 V.

6 Edge-enhanced correlation

Image correlation is a process used to compare two images and report their degree of likeness [31]. The simple correlation operation can be performed by several techniques. Digital computers can perform discrete correlations. An alternate method of performing correlation operation is using all optical techniques. Optical correlation is usually performed using degenerate four-wave mixing (DFWM) techniques [30].

A novel approach for performing programmable all-optical edge-enhanced correlation using organic PRs, which can even be further simplified to a typical TBC configuration, is reported in Banerjee *et al.* [29]. Contrary to DFWM techniques, all beams travel nominally in the same direction, as shown in Figure 6. In this arrangement, similar to joint transform correlation [31], the horizontally traveling reading beam reconstructs the holograms written in the PR material by the FTs of the transparencies or objects t_1 and t_{-1} .

Figure 7 summarizes the various outputs in different directions when readout is performed using a beam along the bisector of the two writing beams. In this case, $t_1(x) \otimes t_1(x)$ and $t_{-1}(x) \otimes t_{-1}(x)$ are autocorrelation functions. It means that there are three components on the z axis, a delta function and two autocorrelation functions. Also, there are two cross-correlation functions, one on each side of the z axis.

It can be shown that preprocessing the object through edge enhancement gives a sharper correlation peak, which is easier to detect in the presence of other cross-correlations and noise. In our arrangement, edge enhancement during correlation is automatically performed, as evidenced from numerical simulations and experimental results.

If there is no additional reading beam, and if both t_1 and t_{-1} simultaneously read out the self-generated hologram, the various diffracted orders and the cross-correlations are shown in Figure 8.

If two different dissimilar profiles, e.g., a square wave and a triangle function with identical dc components, are assumed in the two arms of the TBC configuration (t_{-1} in -1 beam, t_1 in $+1$ beam) and the second-order intensities I_2 and I_{-2}

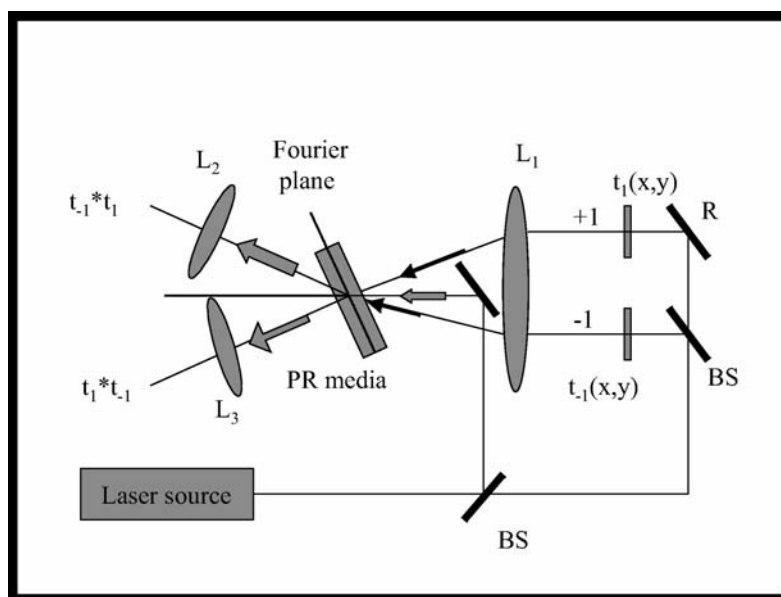


Figure 6 TBC configuration for correlation.

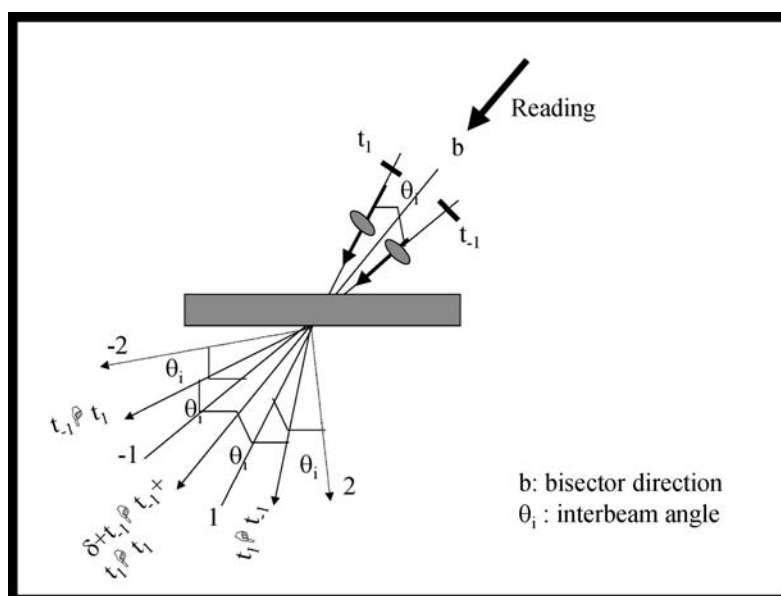


Figure 7 Output correlation with reading beam.

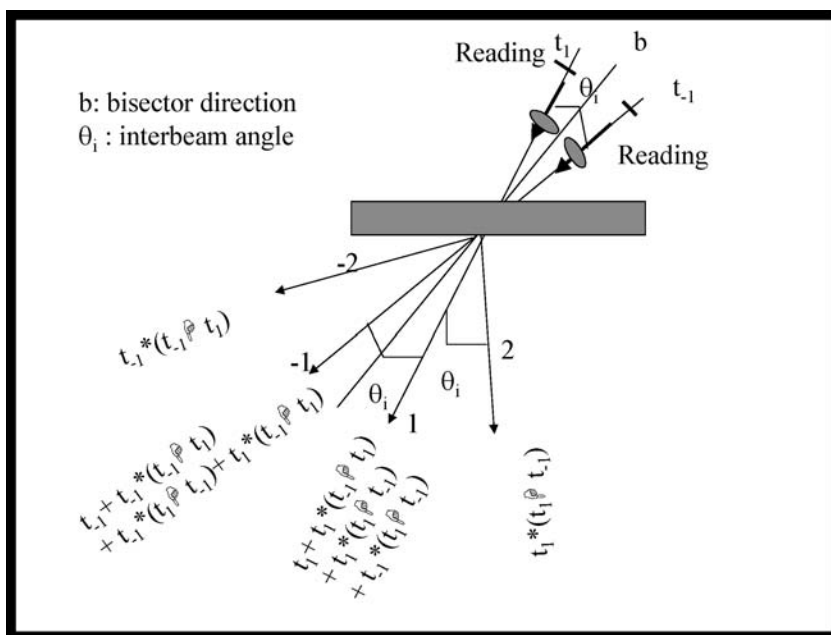


Figure 8 Output correlation when E_{-1} and E_1 beams read the grating simultaneously.

are computed, $t_1 * (t_1 \otimes t_{-1})$ should appear in the +2 direction and $t_{-1} * (t_{-1} \otimes t_1)$ in the -2 direction. The simulation is run at different values of the applied voltage (1000, 1500, 2000, and 2500 V). We find that, in general, the peak intensity is at

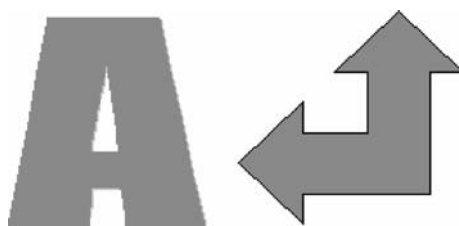


Figure 9 Typical objects used for autocorrelation and cross-correlation experiments.

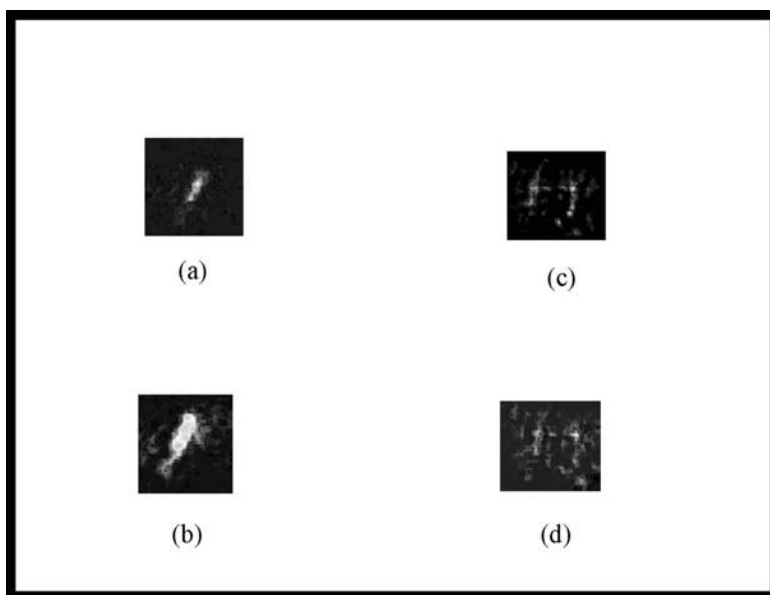


Figure 10 (a) Autocorrelation of the letter A with a bias voltage of 1000 V; (b) cross-correlation of the letter A and the arrows with a bias voltage of 1000 V; (c) Autocorrelation of the letter A with a bias voltage of 1300 V; (d) cross-correlation of the letter A and the arrows with a bias voltage of 1300 V.

least three times higher in the case of similar profiles than in the case of dissimilar profiles, as expected. Experimental results performed using letters of the alphabet as transparencies t_1 and t_{-1} are in agreement with theoretical predictions.

Typical autocorrelation and cross-correlation results of objects shown in Figure 9 are presented in Figure 10. Note that the autocorrelation intensity is much more tightly confined in space and is brighter than the cross-correlation intensity. The results are enhanced at higher bias voltages.

7 Conclusion

We have summarized the optical and EO properties of different types of PR polymers and demonstrated results of TBC in these materials. Applications in the areas of edge enhancement, and edge-enhanced correlation with and without reading

beams have been demonstrated. The main disadvantages of PR polymers are the high bias voltage and material stability. Ongoing and future research in novel polymeric PR materials should make them viable for an attractive alternative to existing PR crystals. Optical pattern recognition using photorefractives could then be achieved much cheaper and could be incorporated into real-life applications.

Acknowledgments

The author is grateful to Prof. Kenji Matsushita from Osaka City University for use of his facilities and collaboration in this work. The author wishes to thank Dr. Esam Gad from Strategic Missile Defense Command, Huntsville, AL, for performing many of the simulations and experiments and for his comprehensive review of PR polymers for his Ph.D. dissertation at the University of Alabama in Huntsville. The author also wishes to thank the ECE department at the University of Alabama in Huntsville, Drs. Hossin Abdeldayem and Don Frazier from the NASA Marshall Space Flight Center and Dr. Tracy Hudson from AMCOM for providing help with experimental equipment and resources.

References

- [1] A. Ashkin, G. Boyd, J. Dziedzic, R. Smith, A. Ballman, H. Levinstein and K. Nassau, Optically-induced refractive index inhomogeneities in LiNbO_3 and LiTaO_3 , *Appl Phys Lett* **9** (1966), 72–74.
- [2] J. Feinberg, D. Heiman, A. Tanguay and R. Hellwarth, Photorefractive effects and high induced charge migration in barium titanate, *J Appl Phys* **51** (1980), 1297–1301.
- [3] R. Sutherland, *Handbook of nonlinear optics*, Marcel Dekker, New York, USA, 1996.
- [4] J. Feinberg, Photorefractive nonlinear optics, *Phys Today* **41** (10) (1988), 46–52 (October issue).
- [5] A. Glass, D. Von der Linde and T. Negran, High-voltage bulk photovoltaic effect and the photorefractive process in LiNbO_3 , *Appl Phys Lett* **25** (1974), 233–235.
- [6] G. Valley, Two wave mixing with an applied field and a moving grating, *J Opt Soc Am B* **1** (1984), 868–873.
- [7] E. Gad, Wave mixing and image processing using photorefractive polymers. Ph.D. dissertation, University of Alabama in Huntsville, USA, 2000.
- [8] P. Gunter and J. Huignard (eds), *Photorefractive material and their applications I & II*, Springer, Berlin, Germany, 1988.
- [9] M. Doncker, S. Silence, F. Walsh, D. Burland, W. Moerner and R. Twieg, Net two-beam-coupling gain in a polymeric photorefractive material, *Opt Lett* (1993), 1044–1046.
- [10] K. Matsushita, D. Miyazaki and A. Kouchi, Photorefractive effect in polymer doped with PNP, *Photorefractive fiber and crystal devices* (F.T.S. Yu and S. Yin, eds), Proc SPIE **3137**, Bellingham, WA, USA, 1997, 59–62.

- [11] S. Ducharme, J. Scott, R. Twieg and W. Moerner, Observation of the photorefractive effect in a polymer, *Phys Rev Lett* **66** (1991), 1846–1849.
- [12] W. Moerner, S. Silence, F. Hache and G. Bjorkland, Orientationally enhanced photorefractive effect in polymers, *J Opt Soc Am B* **11** (1994), 320–330.
- [13] S. Silence, C. Walsh, J. Scott and W. Moerner, C₆₀ sensitization of a photorefractive polymer, *Appl Phys Lett* **61** (1992), 2967–2969.
- [14] S. Silence, J. Scott, F. Hache, E. Ginsberg, P. Jenkner, R. Miller, R. Twieg and W. Moerner, Poly(silane)-based high mobility photorefractive polymer, *J Opt Soc Am B* **10** (1993), 2306–2312.
- [15] S. Silence, J. Donckers, C. Walsh, D. Burland, R. Twieg and W. Moerner, Optical properties of poly(N-vinylcarbazole)-based guest-host photorefractive polymer systems, *Appl Opt* **33** (1994), 2218–2222.
- [16] Y. Zhang and P. Prasad, Observation of photorefractivity in a fullerene-doped polymer composite, *Phys Rev B* **46** (1992), 9900–9902.
- [17] S. Silence, C. Walsh, J. Scott, T. Matray, E. Twieg, F. Hache, G. Bjorklund and W. Moerner, Subsecond grating growth in photorefractive polymer, *Opt Lett* **17** (1992), 1107–1109.
- [18] K. Tamura, A. Padias, H. Hall and N. Peyghambarian, New polymeric material containing tricyanovinylcarbazole group for photorefractive applications, *Appl Phys Lett* **60** (1992), 1803–1805.
- [19] N. Peyghambarian, K. Tamura, Y. Kawabe, F. Jarka, S. Mazumdar, D. Guo, K. Hall and A. Padias, Photorefractive, electro-optical and nonlinear optical properties of conjugated polymers, *Proc Mat Res Soc* **277** (1992), 145–152.
- [20] L. Yu, Z. Bao and S. Cao, Photorefractive polymers, structure design and property characterization, *Macromolecules* **26** (1993), 2216–2221.
- [21] L. Yu, Z. Bao and S. Cao, Pure photorefractive polymer-synthesis and physical measurements, *Polym Preprints* **33** (1993), 398–399.
- [22] M. Sansome, C. Teng, A. East and M. Kwiatek, Observation of photorefractive effect in a dialkylaminonitrostilbene copolymer, *Opt Lett* **18** (1993), 1400–1402.
- [23] A. Grunnet-Jepson, C. Thompson, R. Twieg and W. Moerner, Amplified scattering in a high-gain photorefractive polymer, *J Opt Soc Am B* **15** (1998), 901–904.
- [24] W.E. Moerner and S.M. Silence, Polymeric photorefractive materials, *Chem Rev* **94** (1994), 127–155.
- [25] K. Meerholtz, B. Volodin, Sandolphson, B. Kippelen and N. Peyghambarian, A photorefractive polymer with high optical gain and diffraction efficiency near 100%, *Nature* **371** (1994), 497–500.
- [26] B. Volodin, Sandalphon, K. Meerholz, B. Kippelen, N. Kukhtarev and N. Peyghambarian, Highly efficient photorefractive polymers for dynamic holography, *Opt Eng* **34** (1995), 2213–2223.
- [27] E. DelRe, A. Ciattoni, B. Crosignani and M. Tamburrini, Approach to space-charge field description in photorefractive crystals, *J Opt Soc Am B* **15** (1998), 1469–1475.

- [28] K. Matsushita, P. Banerjee, S. Ozaki and D. Miyazaki, Multiwave coupling in high-gain photorefractive polymer, *Opt Lett* **24** (1999), 593–595.
- [29] P.P. Banerjee, E. Gad, T. Hudson, D. McMillen, H. Abdeldayem, D. Frazier and K. Matsushita, Edge enhancement and edge enhanced correlation with photorefractive polymers, *Appl Opt* **39** (2000), 5337–5346.
- [30] P. Yeh, *Introduction to photorefractive nonlinear optics*, Wiley, New York, NY, USA, 1993.
- [31] J.W. Goodman, *Introduction to Fourier optics, 2nd ed*, McGraw-Hill, New York, NY, USA, 1996.



Partha P. Banerjee is Professor and Chair of Electrical and Computer Engineering at the University of Dayton. Prior to Dayton, he was Professor of Electrical and Computer Engineering at the University of Alabama in Huntsville. His areas of interest are nonlinear optics, photorefractive materials, acousto-optics and image processing. He has authored or coauthored more than 75 refereed journal publications, more than 100 conference papers, and three textbooks, and several invited book chapters. He is the coauthor of books *Principles of Applied Optics* (Irwin/McGraw-Hill, 1991), *Computational Methods in Electromagnetic and Optical Systems* (Marcel Dekker,

2000) and *Contemporary Optical Image Processing with MATLAB* (Elsevier, 2001). He is a Fellow of SPIE and OSA, and is a senior member of IEEE. He obtained the NSF Presidential Young Investigator Award in 1987.

Optical Energy Harvesting Materials

David L. Andrews

Abstract

Optical energy harvesting signifies an integrated approach to the gathering of energy from light. The mechanisms that operate in the molecular apparatus of natural photosynthetic systems owe much of their efficiency to the highly localized pooling of energy—though that is not a feature that has been emulated in most artificial constructs. Recently, however, biomimetic energy pooling has become a theme of significant interest in the arena of photoactive materials, where applications are less restricted by a direct focus on the efficiency of solar energy harvesting. These materials present more immediate opportunities for implementation in devices for signal processing, optical computing and information technology (IT) applications. Exploiting advances in synthetic chemistry and laser photophysics, new energy pooling materials are designed to operate at much higher than ambient intensities of light, with operational characteristics geared to specific laser input wavelengths. In this chapter, it is shown how key principles learned from nature are being applied and extended in the design of optically nonlinear materials for energy pooling. Particular attention is bestowed on dendrimers, other multichromophore arrays, and rare-earth-doped materials.

1 Introduction

The modern technology of optical energy harvesting has a wide variety of operational principles, the main themes of which have evolved through pursuit of better control and economy in the global utilization of solar energy. The concept of *harvesting* signifies an integrated approach to gathering from natural or environmental resources, with centralized collection and subsequent distribution according to requirements. Over the years, a plethora of schemes have been devised for the harvesting of energy, linking sustainability with economy and, increasingly, with green issues. Prominent among recent developments from traditional methods of environmental energy harvesting are as follows: (i) devices for direct conversion of solar radiation to electric power—advanced antenna arrays; (ii) energy production from the flow of moving water in streams, currents, or pipes—technology based on advanced piezoelectric materials; (iii) microgenerators using the temperature differential between the ground and atmosphere—thermoelectric devices and materials; (iv) microhydraulic transducers producing electrical energy from environmental mechanical oscillations—combining piezoelectric with micro-electromechanical systems (MEMS) technology. Many of the newer methods still under development signify relatively local, low-scale approaches to energy harvesting.

For the scale of the environmental resource it represents, and also the extent of its geographic availability, solar power easily outstrips its competition. Each year, in excess of 10^{24} J of solar energy impinges on the earth. This exceeds by a factor of more than 100 the next largest source of environmentally available power (wind energy, 10^{22} J), while both are considerably in excess of a global demand that is not expected to reach 10^{21} J until the 2020s [1]. Natural photosynthesis represents a process whose emulation is an obvious target in energy harvesting. In fact, most of the energy reaching the earth's surface is infrared; 43% is visible and 4% ultraviolet. Infrared energy is only efficient at producing chemical change at very high intensities, well above those sustainably tolerable for biological systems. Visible light must be harvested for sustainable efficiency. However, simple consideration of the molecular energetics suggests that the mechanisms involved must be complex—and involve the collective action of two or more photons in some form of energy pooling. This is evident from the fact that, whereas photosynthesis generates oxygen through fission of water, the energy required to break any of their O—H bonds (4.8 eV, corresponding to an ultraviolet wavelength of 260 nm) is beyond any one photon in the visible spectrum.

There are many well-established nonbiological solutions to the energy problem, including photovoltaic action for the chemical storage of energy (e.g., photoelectrolysis) and photocatalytic use of metal-ligand complexes to lower the threshold energy for decomposition of water. Recently, for example, it has been shown that doping of indium tantalum oxide with nickel can reduce the bandgap to 2.3 eV, corresponding to a wavelength of 540 nm—though the mechanism is unclear and the quantum efficiency at 402 nm is only 0.66% [2]. Hybrid solar cells based on inorganic nanorods interfacing semiconducting polymers appear to hold more promise,

with prototype devices achieving efficiencies of 6.9% [3]. However, as the detailed molecular structure and chromophore layout of the photosynthetic apparatus in a variety of living organisms have begun to emerge, new avenues of research are leading to significant advances in the modeling, synthesis and operation of distinctly biomimetic energy-harvesting materials based on photon energy pooling. Whereas the purpose of manmade energy-harvesting materials is not usually biosynthesis, both artificial and natural systems share a common need to capture and efficiently channel optical energy to a suitable site of repository, and crucial lessons can be learned from the structure and operation of photobiological systems [4].

2 Precepts from photobiology

In green plants, the photosynthesis of each elementary carbohydrate unit (CH_2O) is achieved by two coupled redox reactions: one is the reduction of carbon dioxide, the other is the oxidation of water. It has long been known that each requires the energy of four photons for its completion. Natural photosystems generally comprise several non-covalently bound pigment-protein complexes. In each complex, the major protein component represents a superstructure within which the pigment molecules are held. In the visible region, the optical properties of such complexes are largely determined by the pigments, though the detailed spectral signatures of the latter are strongly influenced by their electronic surroundings.

Purple photosynthetic bacteria have light-harvesting systems whose structural and mechanistic properties [5]–[7] have been thoroughly characterized by means of ultrashort-pulsed laser studies and x-ray protein crystallography. Here the photosynthetic unit comprises pigment-protein complexes in which the key light-harvesting pigment is bacteriochlorophyll. Two of the complexes, designated LH1 and LH2, form coplanar ring structures as schematically depicted in Figure 1. In each circular aggregate, bacteriochlorophyll pigments take the form of dimers

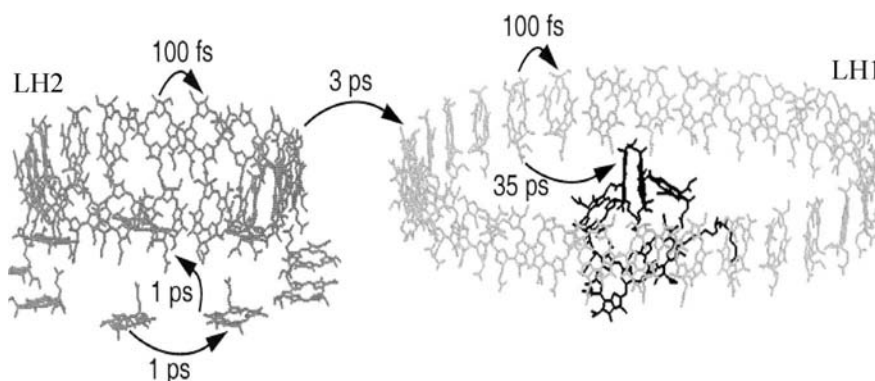


Figure 1 Key energy transfer processes in the light-harvesting complexes of purple photobacteria. Adapted from Ref. [5].

whose wavelength of maximum absorption is significantly modified by their protein environment. In common terminology, the bacteriochlorophyll dimer with an absorption maximum at 800 nm is referred to as B800. The complex LH2, on the left in Figure 1, comprises both B850 (above) and B800 rings (below); whereas LH1, on the right, has a B880 ring. In fact, many LH2 rings surround each LH1, as shown in Figure 2. Together, these complexes deliver light energy to the reaction center (RC, shown at the center of LH1) where photosynthesis ensues.

A striking feature is the very rapid transfer of energy both within and between the light-harvesting arrays. In these multichromophore complexes, the operation of a *spectroscopic gradient* ensures that multistep energy migration is not simply a random walk; each transfer from one bacteriochlorophyll species to another absorbing at a longer wavelength signifies a small but significant loss of energy, ultimately manifest as heat. For any one transfer step, a key factor in determining its efficiency is the extent of overlap between the emission spectrum of the species from which energy departs (which in this connection assumes the role of *donor*—let us call it *A*) and the absorption spectrum of the species to which energy transfers (the *acceptor*, *B*). In any subsequent migration, where the erstwhile acceptor *B* plays the role of donor, back-transfer to *A* is largely inhibited by the much poorer overlap between the emission spectrum of *B* and the absorption spectrum of *A*. For this reason, the transfer of excitation energy toward the reaction center is favored, a feature often referred to as *channeling* or *funneling* (the same principle also assists the one-way transfer of energy from carotenoids to chlorophyll species). Finally, at the reaction center, the energy of successively arriving photons is accumulated and serves to drive the redox half-reaction that initiates chemical photosynthesis.

Four key structural and mechanistic principles can be identified in the operation of such biological photosystems, and in generalized form many of the same princi-

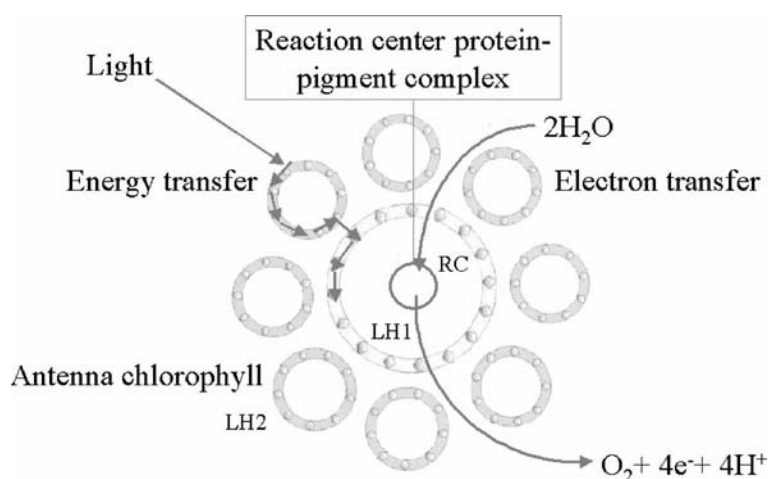


Figure 2 Energy flow in a bacterial photosystem for the oxidation of water. Adapted from Ref. [8].

ples inform the design of photoactive nanosystems and other light harvesting materials. These are as follows: (i) efficient *antennae* are required; the chromophores responsible for photon absorption need broad and intense absorption bands (and also correspondingly strong emission bands, as will emerge below), i.e., strong oscillator strength; (ii) excitation energy is conveyed through the system with high efficiency by a series of ultrafast steps. These usually involve the electromagnetic coupling mechanism known as *resonance energy transfer*, the “resonance” signifying no energy losses during the actual transfer; (iii) following each step in multistep energy migration, there may, however, be losses associated with *intramolecular relaxation*. Thus, as the energy progresses through a series of different chromophores (or identical ones with increasing bathochromic shifts due to their electronic environment), a *spectroscopic gradient* serves to ensure directionality; (iv) the convergence of optically derived energy culminates in *energy pooling*, usually a process of up-conversion (release of radiation with shorter wavelength) or physicochemical change—again, details are given later in this chapter. To develop a more complete understanding of these factors, we first turn to a more detailed appraisal of the photophysics involved in resonance energy transfer.

3 Resonance energy transfer

In considering the molecular mechanisms involved in each transfer step in a multistep process, more than one kind of excitation transfer can be expected. The spectroscopic gradient principle alone strongly suggests the possibility of a difference in behavior when the donor and acceptor species are identical, as, for example, when we consider excitation transfer *within* one of the light-harvesting ring structures. Moreover, the possibility of close proximity between donor and (neighboring) acceptor within a ring structure clearly allows for much stronger coupling than could occur between well-separated chromophores. In the following, the main principle governing the transfer of excitation between well-separated chromophores with different optical characteristics is first outlined, and the modified behavior observed in the identical-neighbor case is then identified.

The primary mechanism for energy transfer goes by a variety of names, each signifying a different aspect of the coupling: this *incoherent, radiationless, Förster*, or *resonance* energy transfer (RET) [9-10] is also designated *FRET* in connection with the ensuing fluorescence [11]. It is a mechanism that mediates energy transfer in a wide variety of materials, including rare-earth-doped crystals, Langmuir–Blodgett films and photoactive dendrimers, as well as multi-chromophore arrays such as the photosynthetic unit. The physical mechanism is *dipole-dipole interaction* between the excited donor and unexcited acceptor—the two dipoles involved being transition dipoles for the decay of the donor and the excitation of the acceptor. The physical condition for this form of energy transfer is essentially weak coupling between the donor and acceptor species. In the modern theory based on quantum electrodynamics (QED), the coupling between the donor and acceptor transitions is calculated in terms of mediation by a short-lived, non-energy conserving,

virtual photon—whose involvement can only be inferred; the energy transfer is radiationless for all practical purposes [12]. The QED theory in fact embraces both radiationless and long-range radiative energy transfer—the results are valid over the whole span of distances, ranging from the nanoscale up to laboratory-scale distances and beyond [13]–[15].

Consider the pairwise transfer of excitation between two chromophore species *A* and *B*. In this elementary process (which might be one RET component of a complex multichromophore mechanism), *A* is designated the *donor* and *B* the *acceptor*—though each may adopt another role subsequent to transfer events. Specifically, let it be assumed that prior excitation of the donor generates *A*^{*}, in which it occupies an excited state *S*_{*n*}. Onward progress of the energy is then accompanied by donor decay to a manifold of states associated with the ground electronic state *S*₀. Acquiring the energy, the acceptor *B* undergoes transition from its ground state *S*₀ to its excited manifold *S*_{*n*}, as illustrated in Figure 3. (The excited acceptor, *B*^{*}, subsequently decays either in a further transfer event or by another means such as fluorescence.)

Restricting consideration to donor-acceptor separations *R* substantially smaller than the wavelengths of visible radiation, the theory gives the following expression for the rate of pairwise energy transfer *w*, applicable for systems where the host for the donor and acceptor has refractive index *n*:

$$w = \frac{9\kappa^2 c^4}{8\pi \tau_A n^4 R^6} \int F_A(\omega) \sigma_B(\omega) \frac{d\omega}{\omega^4}. \quad (1)$$

In this expression, κ is an orientation factor given later in (3); $F_A(\omega)$ is the normalized spectrum of donor fluorescence; $\sigma_B(\omega)$ is the linear absorption cross-section of the acceptor and ω the optical frequency in radians per second ($= 2\pi\nu$, with ν

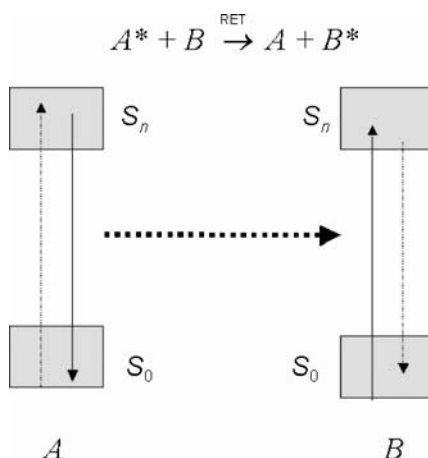


Figure 3 Energetics in one stage of resonance energy transfer. In the QED theory based on virtual photon coupling, donor decay may precede *or* follow (with ultrafast rapidity) the acceptor excitation.

in Hz); c is the speed of light. An immediately evident implication of the result is that efficient energy transfer demands significant overlap of the emission spectrum for A with the absorption spectrum of B . (A similar principle operates in microwave energy transfer, whose efficiency depends on the corresponding overlap of Fourier-mode components.) Also featuring in (1) is τ_A , the radiative decay lifetime of the donor, related to the measured fluorescence lifetime τ_{fluor} through $\tau_{\text{fluor}} = \eta\tau_A$, where η is the fluorescence quantum yield. An alternative, exactly equivalent expression for the rate of resonance energy transfer is as follows:

$$w = \frac{2\kappa^2}{3} \frac{1}{\tau_A} \left(\frac{R_0}{R} \right)^6. \quad (2)$$

Here, R_0 is the *critical* or *Förster radius*, calculable from the overlap integral of the donor fluorescence and acceptor absorption spectrums; typical values of the Förster radius range over a few nanometers.

Apparent in both (1) and (2), the inverse sixth power dependence on donor-acceptor separation R is a well-known characteristic of dipole-dipole energy transfer. It determines the fact that, with intermediary donor/acceptor species suitably positioned and electronically distinct, any process resulting in the overall migration of excitation over distances beyond the Förster radius will mostly operate through a series of short hops rather than one long hop (see below regarding exciton propagation). The κ factor in the energy transfer rate equation (which is, in fact, identical in form to the orientational dependence in static dipole coupling) depends on the orientations of the donor and acceptor, both with respect to each other and with respect to their mutual displacement unit vector $\hat{\mathbf{R}}$ as follows:

$$\kappa = (\hat{\mu}_A \cdot \hat{\mu}_B) - 3(\hat{\mathbf{R}} \cdot \hat{\mu}_A)(\hat{\mathbf{R}} \cdot \hat{\mu}_B). \quad (3)$$

Here, for each molecule, $\hat{\mu}$ designates a unit vector in the direction of the appropriate transition dipole moment. Unfavourable orientations can reduce the rate of energy transfer to zero; others, including many of those found in nature, will optimize the transfer rate. The angular disposition of chromophores is therefore a very important facet of energy transfer, and one that invites careful consideration in the design of energy-harvesting materials. In correcting a common misconception, note that transfer is not necessarily precluded when the transition moments lie in perpendicular directions—provided that neither is orthogonal to $\mathbf{R}$ ($=R\hat{\mathbf{R}}$).

The dipole-dipole coupling that leads to the foregoing results is in certain cases not applicable, as, for example, if either the donor or acceptor transition is electric dipole (E1)-forbidden or exceptionally weak. Then, the mechanism for energy transfer can involve the electric quadrupole moment (E2), higher electric multipoles (En) or even magnetic multipoles (Mn), in each case leading to an orientation and distance dependence of a different form. In the most common case of predominantly electric coupling, if (En) and (Em) are the leading nonzero moments of the donor and acceptor, the distance dependence takes the form $R^{-2(n+m+1)}$ [16-17].

Observations of energy transfer associated with dipole-forbidden interactions often reveal a degree of efficiency that precludes explanation simply on the basis of higher-order multipole coupling, however. The Dexter *exchange mechanism* [18] is based on the premise that when the electron distributions of the donor and acceptor are close enough to overlap strongly, the energy of electronic excitation can pass directly to the acceptor, essentially channeled by the overlapping electron clouds. As such it is a mechanism that comes into play only at very short distances where it exhibits an exponential dependence on R , reflecting a typical asymptote for the wavefunctions of the molecular orbitals. It is not always appreciated that the Dexter mechanism can operate over such short distances both for dipole-forbidden and also dipole-allowed transitions; for an account of how the Dexter and Förster mechanisms seamlessly merge for the dipole-allowed case, see Scholes and Ghiggino [19]. When donor and acceptor chromophores are electronically distinct, however, one should account for dipole-forbidden transitions in terms of higher multipoles.

It is known that the dielectric properties of the medium within which energy transfer takes place (the protein matrix in photosynthetic systems, for example) exercises a considerable influence on the transfer efficiency, beyond the refractive dependence exhibited in the Förster formula (1) [20]. Where electronically prominent chromophore species intervene between donor and acceptor groups, the detailed understanding of such an influence has to be sought at the molecular level. For example, one obvious factor in determining the efficiency of excitation transfer is possible mediation of the transfer through the electronic involvement of a bridge or other suitably placed species. Work on rigidly linked bichromophore and multichromophore molecules has given conclusive evidence that, under certain conditions, a through-bond super-exchange interaction can dominate intramolecular energy transport [21]–[23]. A recent analysis based on molecular QED has examined a number of the broader issues associated with the electronic influence of a neighboring chromophore on resonance energy transfer, showing how it can be tailored to effectively “switch on” the energy transfer process under circumstances where energy transfer is normally forbidden by orientational effects or symmetry grounds [24].

In the case of *identical* chromophores, closely neighboring but electronically distinct differences arise through the possible formation of excitons. It is as well to make clear the meaning of the term *exciton* in the present context, since elsewhere it is employed with a wide variety of connotations—even including the single donor- or acceptor-localized excitations discussed above. In common with much of the recent literature on energy harvesting, however, *exciton* here denotes a nonpropagating, delocalized electronic state associated with the quantum interference of separate chromophore states. At the simplest, the excited states of any isolated pair (comprising a strongly coupled donor and acceptor) can mix and form a delocalized excited state split by twice the coupling energy. With such excitonic states one can no longer specify the molecular location of the electronic excitation; by extension, in light-harvesting complexes comprising large numbers of equivalent

chromophores, excitons can spread over several equivalent donor/acceptor species and be associated with numerous, closely separated energy levels. Indeed, the separation between the lowest levels affords a useful means of gauging the extent of exciton delocalization [25]. Although any local disorder can substantially compromise the extent of excitonic delocalization and also any superexchange coupling, a surprising finding in connection with B800-B850 coupling was that site disorder can be responsible for a more rapid transfer of excitation than would occur in its absence [26]. The multiplicity of the associated exciton splitting enhances spectral overlap in multichromophore arrays and thereby accelerates energy transfer. Thus, in natural light-harvesting complexes the ring structures, which provide symmetry in the pigment arrays, enhance absorption and lead to those arrays as a whole acting as nanoscale energy traps. The mechanistic interpretations of many bacterial and other photosynthetic systems are based on this premise—see Refs. 27-29.

4 Dendrimers

Following several decades in which the science of energy-harvesting materials was pursued largely through the development and optimization of photovoltaic cells, the arrival of dendrimeric macromolecules has heralded a new opportunity for achieving more closely biomimetic methods for light harvesting [8]. Dendrimers are multibranched nanoscale structures, with essentially fractal geometry and a large number of chemically similar chromophores among which energy can transfer; as such they are also known as *nanostars* or *functional cascade molecules* [30-31]. Much current work is directed towards the use of photoactive dendrimers—not only as model systems for photosynthetic light harvesting, but also as light-emitting diodes, fluorescent sensors, frequency conversion materials and other photonic devices. Future advances may lead toward more directly biomimetic energy harvesting systems. One of the objectives is to achieve, through funneling and trapping, the kind of energy pooling that characterizes biological systems, where the energy of more than one photon is accumulated in a trap—a topic that is to be revisited later in this chapter.

In the typical example shown in Figure 4, some of the features also common to natural photosynthetic systems are immediately apparent: principally, the circularly symmetric ringlike array of chromophores about a central excitation trap. However, not only these structural motifs are important; it is also highly significant that the chromophore units are electronically distinct—here, a result of the alternate 1,3,5-substituent linkage pattern on each benzene ring. Again, the ether linkage has the effect of electronically isolating the phenyl chromophores; ethynyl linkages are also commonly employed in such dendrimers to similar effect, as detailed theory [30], [32] and spectroscopic evidence [33] have shown. As a result, there is negligible through-bond coupling between the chromophore units, and transfer of excitation between them essentially takes the form of dipole coupling.

In any dendrimer the synthetic route gives successive generations, each expanded by the peripheral addition of further functional groups. The addition of fur-

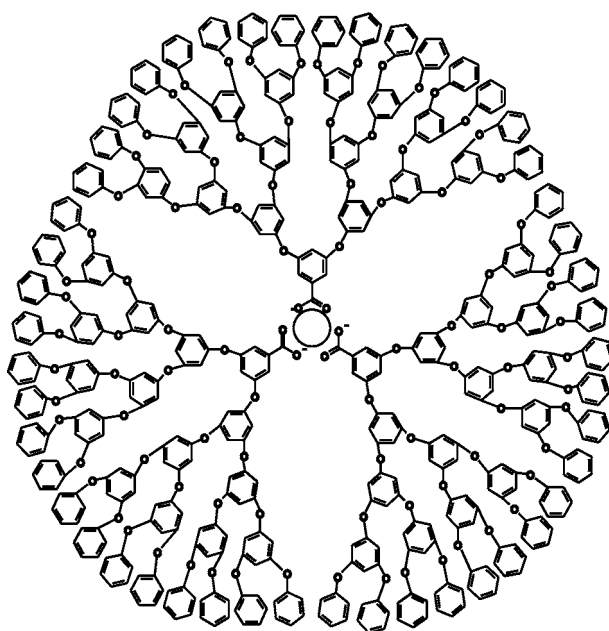


Figure 4 Polyphenylether dendrimer with a photoactive trap.

ther chromophores adds to the photon capture efficiency of the dendrimer, though the associated generation-dependence in the distribution of inter-chromophore distances can result in a reduction in efficiency of energy transfer [34]. Clearly the central unit, even when of chemically identical form to the dendrite arms, differs in its electronic environment from those other chromophores—equally, among those, the peripheral chromophores are in a still different environment. This can be played to advantage; in ideal cases the associated shifts in the corresponding absorption and fluorescence spectrums can foster exactly the kind of spectroscopic or energy gradient observed in natural systems. This principle has been neatly illustrated in work on perylene-functionalized phenylacetylene dendrimers, where the directionality of energy flow from antenna chromophores to the core is a hundred times greater than in the analogous structure with dendrons all of equal length [35–36]. As observed in Section 2, such an energy funnel represents a combination of structural and spectroscopic elements together directing energy towards the core, the latter acting as the trap [37–38]. The efficient unidirectional transfer of energy within a dendrimeric frame towards a core chromophore was in fact first reported only as recently as 1994 [39]. A new development in this area has been the design of dendrimers whose internal cavities host smaller dye molecules. Exploiting not only intramolecular but also intermolecular (dendrimer—dye) energy transfer, extremely high harvesting efficiencies (estimated at approximately 80%) have recently been reported for eosin embedded in a dendrimer with 65 chromophore groups of four different types [40].

Although not strictly dendrimeric, other light-harvesting materials with analogous structural motifs have also been investigated. A molecular square, 20-chromophore unit has recently been constructed by use of metal ion directed self-assembly methods, and shown to exhibit biomimetic energy transfer from outer to inner (pyrene to perylene) chromophores [41]. Closer to nature, multiporphyrin systems in particular appear to offer considerable promise [42]. Here, for example, it has been shown that excitation transfer, between a zinc porphyrin acting as donor and its free base counterpart as acceptor, can be significantly enhanced by exploiting suitable bridging chromophores [43]. As with the dendrimers, these structures are designed to obviate transfer between donor and acceptor species through charge conjugation or wavefunction overlap. Other recent work has forged a synthetic and conceptual link in this area, through the synthesis and characterization of multiporphyrin arrays based on a phenylethynyl protodendrimeric framework [44]. A striking success is an artificial photosynthetic antenna-reaction center complex comprising four zinc tetraarylporphyrins covalently linked, through phenylethynyl dendrites, to a free base porphyrin-fullerene acting as a “reaction center.” Following photoexcitation of the peripheral zinc porphyrin (*antenna*), energy migrates to the central zinc porphyrin (*donor*) from which it transfers to free base porphyrin (*acceptor*), initiating electron transfer to the fullerene (*reaction center*). In this system, the charge-separated excited state is generated with an impressively high quantum yield of 0.90, based on the light absorbed by the zinc porphyrin antenna [45-46].

5 Rare-earth materials for energy pooling

It may at first sight be surprising to discover several of the mechanistic principles discussed in the foregoing dendrimeric materials operating in the field of solid-state crystal optics (indeed, recognition of this has led to some highly innovative work on advanced energy pooling nanomaterials, discussed later in this chapter). Conventional energy transfer does to some extent modify the operating characteristics of laser crystals; in essence, excitation hopping competes with radiative stimulated emission as a channel of decay for electronically excited ions. However, in laser systems other novel forms of energy transfer can occur—in particular, processes of *concerted* energy pooling become possible.

The pooling that features in photobiological systems is associated with the eventual collection of energy acquired from photons that have been absorbed at various earlier times; moreover, various chromophore species experience electronic excitation and subsequent decay in the process of energy migration to the reaction center. However, the much higher levels of photon flux in laser systems makes it possible to pool the energy of photons that arrive *together* (within an ultrashort timescale). Consequently, the time-energy uncertainty principle allows intermediaries in the process of energy transport to undergo transitions to extremely short lived, non-energy-conserving virtual states. In this respect, the process has a similar basis to two-photon absorption, and indeed most other forms of optical

nonlinearity. Experimentally, such energy pooling becomes manifest in certain observations of blue-shifted fluorescence; notably, it has been shown that rare-earth (lanthanide)-doped crystals display the effect at high levels of optical excitation. In general, this fluorescence owes its origin to energy transfer mechanisms involving three chromophore/fluorophore sites, with two acting as donors and one as acceptor [47]–[49]. By exploiting the effect, it is possible to tailor such materials specifically for stepwise or sequential laser frequency upconversion.

The reason for rare-earth ions to be so favored in this connection is that they display strong f - f (core f -electron) transitions associated with a redistribution of charge sufficiently close to the nucleus that the ionic environment exercises relatively little influence. As a consequence, their absorption and fluorescence spectra contain highly discrete, characteristic and well-resolved lines, with many falling in wavelength regions eminently suitable for laser excitation [50]–[52]. It should be emphasized that the materials under discussion here operate quite differently from conventional nonlinear optical crystals (where off-resonance input photons directly engage in pairwise or higher-order coupling, not involving any optical excitation of the material). By contrast, energy pooling processes are *non-parametric* and generally involve dissipative losses; indeed, associated phonon generation can expedite transfer by modifying the selection rules. The features such materials have in common with those discussed previously are largely those listed at the end of Section 2, here to be understood in an atomic context. Specifically, there must be high efficiency in photoabsorption, energy transfer and energy pooling; only the spectroscopic gradient is less relevant—as fewer transfer steps are involved.

In detail, the mechanisms that operate in rare-earth energy-pooling materials entail resonance energy transfer from two initially excited ionic fluorophore donors A and B (which may, but need not, be chemically identical), to other ion acceptors C . Two fundamentally different mechanisms can mediate the effect. In one mechanism, illustrated top left in Figure 5, energy transfer involves routing from A to C via B , at which intermediate point the additional excitation energy of B is acquired. The sum energy thereby delivered to C accomplishes the excitation of C , whose ensuing fluorescent decay produces the observed, up-converted radiation. The pooling of energy at C , which should not be thought of as a two-step process since energy need not be conserved before its completion, has been designated an *accretive* mechanism with due regard to the nature of the intermediary role of B [53]. The overall process is still regarded as proceeding in stepwise fashion because the donor excited states have a finite lifetime before energy transfer, as does the acceptor excited state prior to its fluorescent decay. A second mechanism, by means of which energy is deposited at C directly from the two donors, is termed *cooperative* [54]; it is depicted in the lower right of Figure 5. The general characteristics of both mechanisms have been determined through a detailed QED appraisal [55].

As illustrated in Figure 6, accretive pooling requires the donor decay to satisfy both single- and two-photon selection rules. In the cooperative mechanism, see

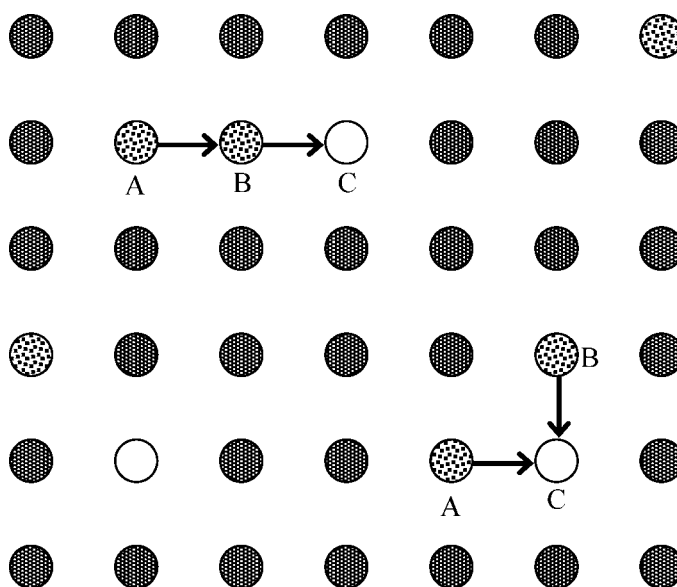


Figure 5 Accretive (upper left) and cooperative (lower right) energy pooling. Dotted circles—excited donors; open circles—acceptors; dark circles—ground state species.

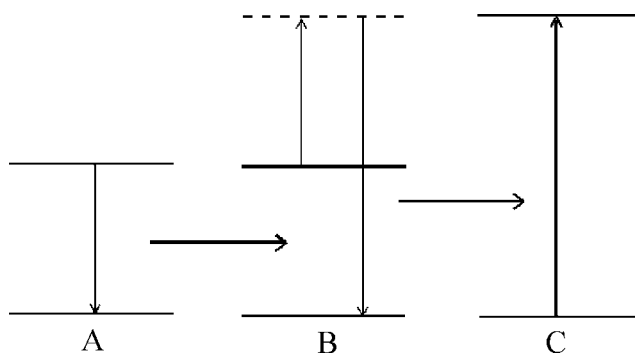


Figure 6 Energetics of excitation pooling: accretive mechanism. Horizontal lines denote electronic levels; the dotted line a virtual intermediate state. Each vertical arrow designates an electronic transition associated with single-photon selection rules.

Figure 7, single-photon transitions characterize each donor decay and excitation of the acceptor follows two-photon selection rules. To observe either form of such energy pooling, in a system accommodating donor and acceptor species appropriate to the laser input, it is necessary to furnish the necessary levels of donor excitation. The specific requirement here is for pairs of donors, *A* and *B*, to be excited through conventional single-photon absorption—so that during a certain period, they are simultaneously in excited states. In the simplest practical case, the donor species are identical and both become excited by the same input radiation. It has long ago been established that two-site single-photon absorption is not intrinsically less probable than the more familiar single-site process of two-photon absorption

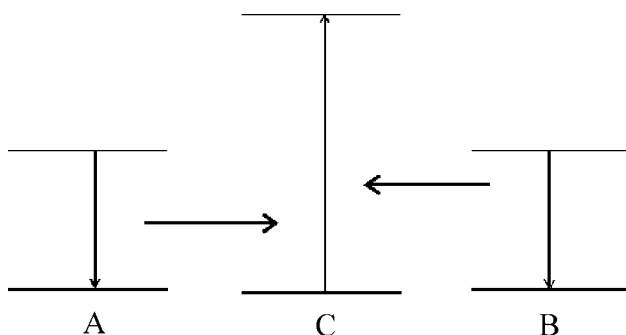


Figure 7 Energetics of excitation pooling: cooperative mechanism. Horizontal lines denote electronic levels; the dotted line a virtual intermediate state. Each vertical arrow designates an electronic transition associated with single-photon selection rules.

[56]—in general, pulsed laser irradiation can readily afford the means of establishing the necessary level of donor excitation for both cooperative and accretive energy pooling. In the optically primed system (i.e., following pump excitation), the relative number densities of excited donors and acceptors is therefore a significant determinant of mechanism, coupled with the different optical selection rules that operate for each mechanism.

In the most general case, A , B , and C may be three chemically different species (or the transitions involved in them may differ even when there is atomic equivalence). Then, in determining the efficiency of energy pooling in a doped crystal it is necessary to consider all permissible ion positions on the lattice, and also the number of ions A , B , and C , residing in their appropriate initial electronic states, per unit cell. If these are written as C_X for species X , then the appropriate concentrations, in a cubic crystal with cell length a , are expressible as $c_X = a^{-3}C_X$. Where the optical selection rules permit only one mechanism to operate, the rate of energy pooling, for each triad combination of A , B , and C ions, depends on the inverse sixth powers of two interionic distances. These are independent of each other—though calculations must discount any triad where two or three ions share identical lattice sites. Allowing for all these restrictions, and assuming random positioning of the salient ions, the lattice sums that feature in the ensemble rates result in the following energy-harvesting efficiency factors [55]:

$$\sum_{\text{lattice}} R_{AB}^{-6} R_{BC}^{-6} \rightarrow \eta C_A C_B^2 C_C a^{-12} = \eta c_A c_B^2 c_C \quad (\text{accretive}), \quad (4)$$

$$\sum_{\text{lattice}} R_{AC}^{-6} R_{BC}^{-6} \rightarrow \eta C_A C_B C_C^2 a^{-12} = \eta c_A c_B c_C^2 \quad (\text{cooperative}). \quad (5)$$

Here η is a structure-dependent numerical factor, in this instance taking the cubic lattice value, $\eta = 64.39$. The lattice weightings thus signify concentration factors for each of the participating ions in the necessary initial states. Where selection rules permit both accretive and cooperative pooling to occur (which is possible when inversion symmetry is absent), quantum interference terms have also

to be accommodated in equations for the rate of energy harvesting. For ions randomly disposed on a cubic lattice, rate contributions from such interferences are just as important as the “pure” accretive or “pure” cooperative kind. In connection with conventional (single-donor) energy transfer, calculations based on this method have been applied extensively to crystal systems [57-58], with some work also accounting for higher multipole couplings and the exchange interaction [59-60]. The corresponding extensions to multiphoton energy pooling remain the subject of active research.

Examples of twin-donor energy pooling in upconversion are rife, and not necessarily limited to systems containing more than one species of rare-earth ion. The occurrence of such a process was, for example, first identified in systems involving a concerted action among three electronically excited Pr^{3+} ions, in which two ions decay and the third is further excited through the acquisition of their combined energy [61-62]. It has been shown that an iso-ionic process also accounts for green and red emissions from Er^{3+} in fluorite-type crystals [63].

6 Energy pooling in multichromophore arrays

Much recent work on energy pooling has been directed toward nanoscale molecular emulation of the up-conversion principles that operate in rare earth materials. In particular, dendrimeric and other still more closely biomimetic light-harvesting arrays offer great scope for technological applications. In this connection, theoretical work is progressing hand in hand with current synthetic efforts, each informing the other; in both areas the difficulties are considerable, yet good progress is being made. Part of the challenge lies in dealing with the complexities associated with energy transfer in systems with intricate molecular architecture [64]. For natural systems, revelation of the light-harvesting structures has enabled theory to home in on the interplay of the relevant electronic properties, a *carte blanche* where artificial systems are concerned.

One of the first examples of energy pooling was the observation of two-photon excitation energy transfer in a trichromophoric system, as illustrated in Figure 8 [65]. Here, antenna species were excited with an optical frequency off-resonant with respect to direct one- or two-photon excitation of the central stilbene, which nonetheless exhibited energy uptake through *E/Z*-isomerization. There is also some evidence to suggest the operation of such mechanisms in photoresponsive dendrimers, particularly those based on an azobenzene core with a pair of benzyl aryl ether dendrites. Here the energy required to effect the observed *cis-trans* isomerization of the core appears to be consistent only with the pooling of excitation from two of the chromophores [66-67]. Meanwhile, other dendrimers, based on (diphenylamino)stilbene, have been shown to exhibit exceptionally high rates of two-photon absorption [68].

For multichromophore arrays, the complexity of energy pooling calculations increases considerably, especially in connection with the selection rules and orientational factors [69]; indeed, for their full analysis, multidimensional optimization

routines are required. By way of illustration, Figure 9 shows a three-dimensional plot of the variation of efficiency for cooperative transfer of energy in a twin-donor system, as a function of the angles made by the transition dipole moments of the two donors, ϕ_A , and that of the acceptor, ϕ_B , relative to the system axis. Here it is assumed that the acceptor is situated directly in-between the two donors, as in the system shown in Figure 8. However, for simplicity, it is assumed that all three transition moments are coplanar and that both donor moments are collinear [70]. Even such results apply just for one of the two contributing mechanisms; the accretive mechanism requires a similar appraisal.

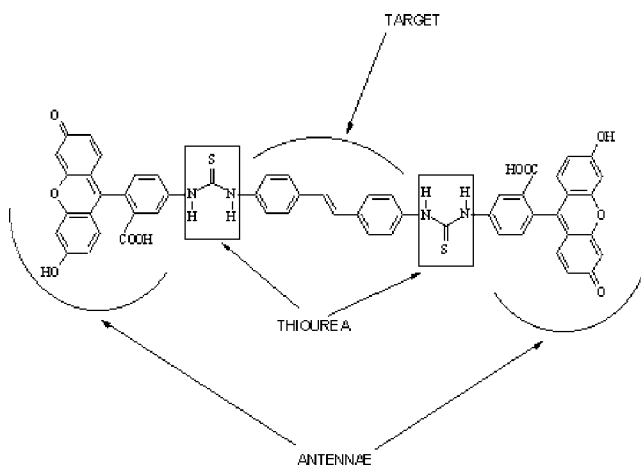


Figure 8 Twin-donor cooperative energy transfer.

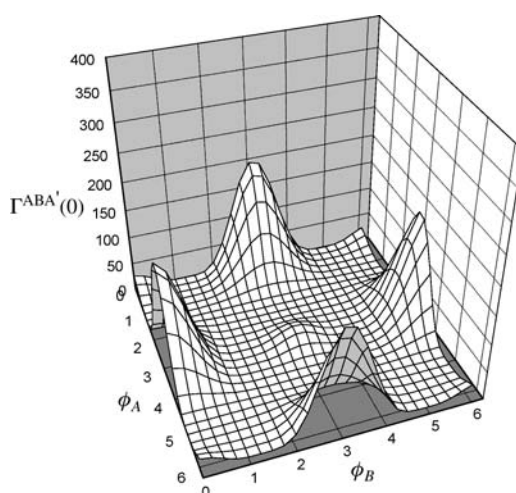


Figure 9 Three-dimensional plot showing variation of the rate of trichromophore energy pooling with angles ϕ_A and ϕ_B , in a molecular architecture favoring the cooperative mechanism.

Recent theory has highlighted the need to simplify conceptualization of multichromophore energy pooling, expedited through the application of a new interaction-pair terminology [71]. This formalism can obviate conceptual and semantic problems in the correct description of multichromophore (or in general, multicenter) interactions. In connection with triple-donor pooling, a schematic depiction as in Figure 10 accommodates the overall light-harvesting operation of a model dendrimer. It establishes the fact that four mechanisms, with varying degrees of accretive and cooperative character, may operate—unless one or more is precluded by the selection rules or by the geometry [72]. However, such illustrations fail to reveal a significant feature that arises where two or more mechanisms do come into play. It needs to be emphasized that the rate of energy harvesting generally has not only contributions associated with each route individually, but also others resulting from their *quantum interference*, which may signify either an enhancement or a reduction in the overall energy harvesting efficiency. Only detailed QED calculations reveal the extent to which the different quantum channels are interlaced. Moreover, for a multichromophore structure, the calculation of interference terms has to accommodate all possible chromophores as intermediaries. Another quantum feature concerns the time ordering of the elementary interactions. Figure 10 designedly omits any arrowheads suggesting directions or orderings for the virtual photon couplings from which each mechanistic pathway is constructed, because *all* temporally distinct orderings must be accommodated in the theory. A simple consequence of the time-energy uncertainty principle, proper accommodation of this feature is necessary in order to correctly derive the form of the electronic tensors for the chromophores involved. Through detailed calculation it emerges that these tensors are closely linked to those which characterize multiphoton absorption and scattering.

7 The future of energy pooling

Optical energy harvesting is an area of science characterized by an almost unparalleled degree of interdisciplinarity. Inspired by discoveries in photobiology, and operating on principles from materials science, spectroscopy and both quantum and nonlinear optics, this rapidly developing subject enjoys the unique facility to accommodate their insights, and to exploit their design strategies for burgeoning technological applications. In this review, attention has focused on the principles and mechanisms of biological photosynthesis, photoactive dendrimers and other

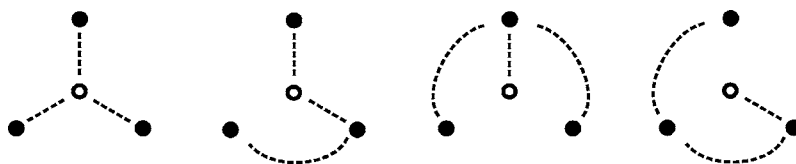


Figure 10 Depiction of energy transfer routes for a triply excited donor array.

multichromophore arrays, and also rare-earth materials. While much research effort is directed toward devising more and more closely biomimetic analogs of natural photosynthetic light-harvesting arrays, other important developments are taking place in the pursuit of materials to exploit optical nonlinearity for energy pooling under conditions of intense optical irradiation.

Much of the recent work on theory has aimed to establish the underlying principles and links between the modes of operation of such energy-harvesting systems. Key factors include the chromophore layout and geometry, the electronic structure and optical selection rules. The fundamental pair-coupling nature of energy pooling becomes manifest in the detailed form of the material parameters, and especially in the interconnection between the diverse quantum channels that can mediate multiphoton excitation trapping [73]. Current work in this area is aiming to fully elicit the significance and role of electronically delocalized excitons in multichromophore arrays, with particular regard to exclusively exciton-mediated channels for energy transfer [74–75]. A complete assessment of the interplay between all of these factors will prove crucial for the future design and optimization of new photoactive materials.

Acknowledgments

I am greatly indebted to Werner Weiglhofer, whose kind invitation to write this chapter was characteristically followed up with many helpful and insightful comments. His wisdom will be greatly missed. Thanks are due to Akhlesh Lakhtakia for much other valuable feedback—and also to Robert Jenkins, at the University of East Anglia (UEA), who in addition produced several of the diagrams. Research in the QED group at UEA is funded by the Engineering and Physical Sciences Research Council.

References

- [1] S.F. Baldwin, Renewable energy: progress and prospects, *Phys Today* **55** (4) (2002), 62–67.
- [2] Z. Zou, J. Ye, K. Sayama and H. Arakawa, Direct splitting of water under visible light with an oxide semiconductor photocatalyst, *Nature* **414** (2001), 625–627.
- [3] W.U. Huynh, J.J. Dittmer and A.P. Alivisatos, Hybrid nanorod-polymer solar cells, *Science* **295** (2002), 2425–2427.
- [4] P. Ball, Natural strategies for the molecular engineer, *Nanotechnology* **13** (2002), R15–R28.
- [5] R. van Grondelle and O.J.G. Somsen, Excitation energy transfer in photosynthesis, *Resonance energy transfer* (D.L. Andrews and A.A. Demidov, eds), Wiley, New York, NY, USA, 1999.
- [6] X. Hu and K. Schulten, How nature harvests sunlight, *Phys Today* **50** (8) (1997), 28–34.

- [7] K. Schulten, From simplicity to complexity and back: function, architecture, and mechanism of light harvesting systems in photosynthetic bacteria, *Simplicity and complexity in proteins and nucleic acids* (H. Frauenfelder, J. Deisenhofer and P.G. Wolynes, eds), Dahlem University Press, Berlin, Germany, 1999.
- [8] A. Adronov and J.M.J. Fréchet, Light-harvesting dendrimers, *Chem Commun* (2000), 1701–1710.
- [9] V.M. Agranovich and M.D. Galanin, *Electronic excitation energy transfer in condensed matter*, Elsevier/North-Holland, Amsterdam, The Netherlands, 1982.
- [10] B.W. van der Meer, J. Coker III and S.-Y.S. Chen, *Resonance energy transfer: theory and data*, VCH, New York, NY, USA, 1982.
- [11] P.R. Selvin, The renaissance of fluorescence resonance energy transfer, *Nature Struct Biol* **7** (2000), 730–734.
- [12] D.L. Andrews, A unified theory of radiative and radiationless molecular energy transfer, *Chem Phys* **135** (1989), 195–201.
- [13] D.L. Andrews and G. Juzeliūnas, Intermolecular energy transfer: retardation effects, *J Chem Phys* **96** (1992), 6606–6612.
- [14] D.L. Andrews and P. Allcock, Bimolecular photophysics, *Chem Soc Rev* **24** (1995), 259–265.
- [15] G. Juzeliūnas and D.L. Andrews, Quantum electrodynamics of resonance energy transfer, *Adv Chem Phys* **112** (2000), 357–410.
- [16] G.D. Scholes, A.H.A. Clayton and K.P. Ghiggino, On the rate of radiationless intermolecular energy-transfer, *J Chem Phys* **97** (1992), 7405–7413.
- [17] G.D. Scholes and D.L. Andrews, Damping and higher multipole effects in the quantum electrodynamical model for electronic energy transfer in the condensed phase, *J Chem Phys* **107** (1997), 5374–5384.
- [18] D.L. Dexter, A theory of sensitized luminescence in solids, *J Chem Phys* **21** (1953), 836–850.
- [19] G.D. Scholes and K.P. Ghiggino, Electronic interactions and interchromophore excitation transfer, *J Phys Chem* **98** (1994), 4580–4590.
- [20] C.-P. Hsu, G.R. Fleming, M. Head-Gordon and T. Head-Gordon, Excitation energy transfer in condensed media, *J Chem Phys* **114** (2001), 3065–3072.
- [21] K.P. Ghiggino, E.K.L. Yeow, D.J. Haines, G.D. Scholes and T.A. Smith, Mechanisms of excitation energy transport in macromolecules, *J Photochem Photobiol A: Chem* **102** (1996), 81–86.
- [22] E.K.L. Yeow, D.J. Haines, K.P. Ghiggino and M.N. Padden-Row, Electronic energy transfer in multichromophoric arrays. A sequential and superexchange dynamics study, *J Phys Chem A* **103** (1999), 6517–6524.
- [23] E.K.L. Yeow and K.P. Ghiggino, Electronic energy transfer in multichromophoric arrays. The effects of disorder on superexchange coupling and energy transfer rate, *J Phys Chem A* **104** (2000), 5825–5836.
- [24] G.J. Daniels and D.L. Andrews, The electronic influence of a third body on resonance energy transfer, *J Chem Phys* **116** (2002), 6701–6712.

- [25] L.D. Bakalis and J. Knoester, Linear absorption as a tool to measure the exciton delocalization length in molecular assemblies, *J Luminescence* **87–89** (2000), 66–70.
- [26] G.D. Scholes and G.R. Fleming, On the mechanism of light harvesting in photosynthetic purple bacteria: B800 to B850 energy transfer, *J Phys Chem B* **104** (2000), 1854–1868.
- [27] S. Gnanakaran, R. Haran, R. Kumble and R.M. Hochstrasser, Energy transfer and localization: applications to photosynthetic systems, *Resonance energy transfer* (D.L. Andrews and A.A. Demidov, eds), Wiley, New York, NY, USA, 1999.
- [28] S. Savikhin, D.R. Buck and W.S. Struve, The Fenna–Matthews–Olsen protein: a strongly coupled photosynthetic antenna, *Resonance energy transfer* (D.L. Andrews and A.A. Demidov, eds), Wiley, New York, NY, USA, 1999.
- [29] H. van Amerongen, L. Valkunas and R. van Grondelle, *Photosynthetic excitons*, World Scientific, River Edge, NJ, USA, 2000.
- [30] T. Minami, S. Tretiak, V. Chernyak and S. Mukamel, Frenkel-exciton Hamiltonian for dendrimeric nanostar, *J Luminescence* **87–89** (2000), 115–118.
- [31] A. Archut and G. Vögtle, Functional cascade molecules, *Chem Soc Rev* **27** (1998), 233–240.
- [32] S. Tretiak, V. Chernyak and S. Mukamel, Localized electronic excitations in phenylacetylene dendrimers, *J Phys Chem B* **102** (1998), 3310–3315.
- [33] R. Kopelman, M. Shortreed, Z.-Y. Shi, W. Tan, Z. Xu, J.S. Moore, A. Bar-Haim and J. Klafter, Spectroscopic evidence for excitonic localization in fractal antenna supermolecules, *Phys Rev Lett* **78** (1997), 1239–1242.
- [34] M. Lor, R. De, S. Jordens, G. De Belder, G. Schweitzer, M. Cotlet, J. Hofkens, T. Weil, A. Herrmann, K. Müllen, M. Van Der Auweraer and F.C. De Schryver, Generation-dependent energy dissipation in rigid dendrimers studied by femtosecond to nanosecond time-resolved fluorescence spectroscopy, *J Phys Chem A* **106** (2002), 2083–2090.
- [35] C. Devadoss, P. Bharathi and J.S. Moore, Energy transfer in dendritic macromolecules: molecular size effects and the role of an energy gradient, *J Am Chem Soc* **118** (1996), 9635–9644.
- [36] M.R. Shortreed, S.F. Swallen, Z.Y. Shi, W.H. Tan, Z.F. Xu, C. Devadoss, J.S. Moore and R. Kopelman, Directed energy transfer funnels in dendrimeric antenna supermolecules, *J Phys Chem B* **101** (1997), 6318–6322.
- [37] A. Bar-Haim and J. Klafter, Dendrimers as light harvesting antennae, *J Luminescence* **76&77** (1998), 197–200.
- [38] A. Bar-Haim and J. Klafter, Geometric versus energetic competition in light harvesting by dendrimers, *J Phys Chem B* **102** (1998), 1662–1664.
- [39] Z.F. Xu and J.S. Moore, Stiff dendritic macromolecules. 4. Design and synthesis of a convergent and directional molecular antenna, *Acta Polym* **45** (1994), 83–87.
- [40] U. Hahn, M. Gorka, F. Vögtle, V. Vicinelle, P. Ceroni, M. Maestri and V. Balzani, Light-harvesting dendrimers: efficient intra- and intermolecular

- energy-transfer processes in a species containing 65 chromophoric groups of four different types, *Angew Chem Int Ed* **41** (2002), 3595–3598.
- [41] F. Würthner and A. Sautter, Energy transfer in multichromophoric self-assembled molecular squares, *Org Biomol Chem* **1** (2003), 240–243.
- [42] P.G. Van Patten, A.P. Shreve, J.S. Lindsey and R.J. Donohoe, Energy-transfer modeling for the rational design of multiporphyrin light-harvesting arrays, *J Phys Chem B* **102** (1998), 4209–4216.
- [43] K. Kilså, J. Kajanus, J. Mårtensson and B. Albinsson, Mediated electronic coupling: Singlet energy transfer in porphyrin dimers enhanced by the bridging chromophore, *J Phys Chem B* **103** (1999), 7329–7339.
- [44] P. Brodard, S. Matzinger, E. Vauthey, O. Mongin, C. Papamicaël and A. Gossauer, Investigations of electronic energy transfer dynamics in multiporphyrin arrays, *J Phys Chem A* **103** (1999), 5838–5870.
- [45] D. Kuciauskas, P.A. Liddell, S. Lin, T.E. Johnson, S.J. Weghorn, J.S. Lindsey, A.L. Moore, T.A. Moore and D. Gust, An artificial photosynthetic antenna-reaction center complex, *J Am Chem Soc* **121** (1999), 8604–8614.
- [46] G. Kodis, P.A. Liddell, L. de la Garza, P.C. Clausen, J.S. Lindsey, A.L. Moore, T.A. Moore and D. Gust, Efficient energy transfer and electron transfer in an artificial photosynthetic antenna-reaction center complex, *J Phys Chem A* **106** (2002), 2036–2048.
- [47] F. Auzel, Upconversion processes in coupled ionic systems, *J Luminescence* **45** (1990), 341–345.
- [48] M. Chua and P.A. Tanner, Three-body energy transfer processes of lanthanide ions in crystals, *J Luminescence* **66&67** (1995), 203–207.
- [49] X. Zhang, J.-P. Jouart and G. Mary, Energy transfer upconversion in Ho^{3+} and Ho^{3+} , Yb^{3+} doped CdF_2 crystals, *J Phys: Condens Matter* **10** (1998), 493–500.
- [50] B.G. Wybourne, *Spectroscopic properties of rare earths*, Interscience, New York, NY, USA, 1965.
- [51] G.H. Dieke, *Spectra and energy levels of rare earth ions in crystals*, Wiley, New York, NY, USA, 1968.
- [52] V.V. Ovsyankin, *Spectroscopy of solids containing rare earth ions* (A.A. Kaplyanskii and R.M. Macfarlane, eds), Elsevier, Amsterdam, The Netherlands, 1987.
- [53] D.L. Andrews, An accretive mechanism for blue-shifted fluorescence in strongly-pumped systems: resonance energy transfer with Raman emission, *J Raman Spectrosc* **31** (2000), 791–796.
- [54] R.D. Jenkins and D.L. Andrews, Twin-donor systems for resonance energy transfer, *Chem Phys Lett* **301** (1999), 235–240.
- [55] D.L. Andrews and R.D. Jenkins, A quantum electrodynamical theory of three-center energy transfer for upconversion and downconversion in rare earth doped materials, *J Chem Phys* **114** (2001), 1089–1100.
- [56] D.L. Andrews, Anomalous absorption of the ultrafast laser supercontinuum, *Phys Rev A* **38** (1988), 5129–5139.

- [57] S.O. Vasquez, Energy transfer processes in organized media. I. A crystal model for cubic sites, *J Chem Phys* **104** (1996), 7652–7657.
- [58] S.O. Vasquez, Energy transfer processes in organized media. II. Generalization of the crystal model for dipole-dipole interactions in cubic sites, *J Chem Phys* **106** (1997), 8664–8671.
- [59] S.O. Vasquez, Crystal model for energy-transfer processes in organized media: higher-order electric multipolar interactions, *Phys Rev B* **60** (1999), 8575–8585.
- [60] C.Z. Hadad and S.O. Vasquez, Energy-transfer processes induced by exchange interactions, *Phys Rev B* **60** (1999), 8586–8594.
- [61] L.-S. Lee, S.C. Rand and A.L. Schawlow, Cooperative energy-transfer among Pr^{3+} ions in LaF_3 , *Phys Rev B* **29** (1984), 6901–6906.
- [62] A. Lezama, M. Oriá, J.R. Rios Leite and Cid B. De Araújo, Triad spectroscopy via ultraviolet upconversion in Pr^{3+} - LaF_3 , *Phys Rev B* **32** (1985), 7139–7142.
- [63] J.P. Jouart and G. Mary, Upconversion in Er^{3+} -doped fluorite-type crystals pumped by 1.5 μm tunable diode-laser, *J Luminescence* **46** (1990), 39–45.
- [64] G.D. Scholes, X.J. Jordanides and G.R. Fleming, Adapting the Förster theory of energy transfer for modeling dynamics in aggregated molecular assemblies, *J Phys Chem B* **105** (2001), 1640–1651.
- [65] M. Nickoleit, A. Uhl and J. Bendig, Non-linear simultaneous two-photon excitation energy transfer in the *wrong* direction, *Laser Chemistry* **17** (1997), 161–174.
- [66] D.M. Junge and D.V. McGrath, Photoresponsive dendrimers, *Chem Commun* (1997), 857–858.
- [67] D.L. Jiang and T. Aida, Photoisomerization in dendrimers by harvesting of low-energy photons, *Nature* **388** (1997), 454–456.
- [68] M. Drobizhev, A. Karotki, A. Rebane and C.W. Spangler, Dendrimer molecules with record large two-photon absorption cross section, *Opt Lett* **26** (2001), 1081–1083.
- [69] R.D. Jenkins and D.L. Andrews, Three-center systems for energy pooling: quantum electrodynamical theory, *J Phys Chem A* **102** (1998), 10834–10842.
- [70] R.D. Jenkins and D.L. Andrews, Orientation factors in three-centre energy pooling, *Phys Chem Chem Phys* **2** (2000), 2837–2843.
- [71] R.D. Jenkins and D.L. Andrews, Four-center energy transfer and interaction pairs: molecular quantum electrodynamics, *J Chem Phys* **116** (2002), 6713–6724.
- [72] D.L. Andrews and R.D. Jenkins, Nonlinearities in energy harvesting media, *Complex mediums II: beyond linear isotropic dielectrics* (A. Lakhtakia, W.S. Weiglhofer and I.J. Hodgkinson, eds), Proc SPIE **4467** Bellingham, WA, USA, 2001, 297–306.
- [73] R.D. Jenkins, D.L. Andrews and L.C. Dávila Romero, A new diagrammatic methodology for non-relativistic quantum electrodynamics, *J Phys B: At Mol Opt Phys* **35** (2002), 445–468.

- [74] R.D. Jenkins and D.L. Andrews, Multichromophore excitons and resonance energy transfer: molecular quantum electrodynamics, *J Chem Phys* **118** (2003), 3470–3479.
- [75] R.D. Jenkins and D.L. Andrews, Exciton energy transfer: effects of geometry and transition moment orientation in model photosystems, *Photochem Photobiol Sci* **2** (2003), 130–135.



David L. Andrews was appointed Professor of Chemical Physics at the University of East Anglia (Norwich, U.K.) in 1996 to head a theory group developing photonic applications of quantum electrodynamics. Following earlier work on multiphoton processes in molecules, the focus is now on nonlinear optics and energy transfer. Andrews is the author of the unified theory of resonance energy transfer; his group was also the first to identify its two-photon analogue and, in another area, the molecular properties supporting five- and six-wave mixing processes. The author of more than 150 journal articles, his name is also

on eight books, including *Lasers in Chemistry*, his widely adopted textbook now in its third edition. A Fellow of the Royal Society of Chemistry (RSC) and of the Institute of Physics, he has international collaborations with several groups. He has twice held visiting Fellowships at the University of Western Ontario in the Centre for Chemical Physics and also at the Department of Physics and Astronomy at the University of Canterbury (Christchurch, New Zealand). Previously Chair of the RSC Molecular Spectroscopy Group, he is now Chair of the SPIE Nanotechnology Technical Group. His current research interests include optical vortexes and multiphoton energy harvesting.

Magnetoelectric Effects in Insulating Magnetic Materials

Hans Schmid

Abstract

The study of magnetoelectric effects originated with the conjecture of Pierre Curie in 1894 that materials may exist that can be polarized electrically by means of a magnetic field and magnetized by means of an electric field. The different kinds of such magnetoelectric effects known today can be classified on a thermodynamic and symmetry basis. This chapter surveys the thermodynamically reversible *induced effects*, i.e., the linear and bilinear magnetoelectric effects; the piezomagnetoelectric and thermodynamically irreversible *spontaneous effects*, i.e., electric field, magnetic field and stress-induced switching or reorientation of domains with spontaneous polarization, spontaneous magnetization, spontaneous toroidal moment and spontaneous deformation. Spontaneous toroidal moments of *ferrotoroidics* are now recognized as measurable contributions to the linear magnetoelectric effect. Electro-, magneto- and piezotoroidic effects have been postulated. Examples of insulating materials displaying different kinds of magnetoelectric effect are discussed. Useful scientific applications of the effects are the determination of magnetic symmetry, representing a powerful complementary tool for magnetic structure determination by neutron diffraction; the study of magnetic phase transitions, magnetic phase diagrams, toroidal moments and toroidal effects; *poling* of antiferromagnetic domains; magnetoelectric spin chirality control; linear magneto-optic effect and magnetoelectric spectroscopy at optical frequencies. Technologically promising are composites, for example, to fabricate nonreciprocal magnetic photonic crystals with magnetoelectric symmetry.

1 Introduction

In 1894, Pierre Curie conjectured that materials may exist that can be polarized electrically by means of a magnetic field and magnetized by means of an electric field [1]. Although he was already aware of the axial vector symmetry of the magnetic field, he did and could not yet say more since the microscopic origins of magnetism were not yet understood. We know today that such materials do exist, but it took more than half a century until the necessary knowledge for demonstrating the phenomenon was acquired step by step. The corresponding effects are now called *magnetoelectric*, a term originally introduced by Peter Debye in 1926 when trying in vain to find a magnetic or electric-field-induced *Richteffekt* (orientating effect) of dipolar paramagnetic molecules [2]. In the 1920s and 1930s, many unsuccessful experimental attempts at finding such magnetoelectric effects were made, well chronicled and discussed in detail by O'Dell [3]. Further progress had to await Compton's idea of the quantized spinning electron in 1921 [4] and the Uhlenbeck-Goudsmit statement of 1926 [5] that the electron must have both spin and magnetic moment. A very important milestone is Wigner's introduction in 1932 of the *time-reversal* symmetry operation R , implying that velocities, electrical current density $\mathbf{j}(r)$, spin density $\mathbf{S}(r)$ and magnetic field $\mathbf{H}$ change sign under application of R ($R\mathbf{j} = -\mathbf{j}$, $R\mathbf{S} = -\mathbf{S}$, $R\mathbf{H} = -\mathbf{H}$), while charge density ρ and electric field $\mathbf{E}$ remain unchanged by R ($R\rho = \rho$, $R\mathbf{E} = \mathbf{E}$) [6-7]. These findings were essential for Landau when developing his theory of phase transitions in 1937 [8], and allowed him to see that the physical properties of a crystal depend both on the symmetry of charge density ρ and current density $\mathbf{j}$ of moving charges, hence allowing *two types of crystals*: (i) those crystals for which $\mathbf{j} = \mathbf{0}$ (when the crystal symmetry coincides with that of ρ), and (ii) those crystals for which $\mathbf{j} \neq \mathbf{0}$. This implies that for magnetic crystals we have $R\mathbf{j} = -\mathbf{j} \neq \mathbf{0}$ and $R\mathbf{S} = -\mathbf{S} \neq \mathbf{0}$ [7]. However, it was not until 1957 that Landau and Lifshitz stated [9] that both *piezomagnetism* and what is now called the *linear magnetoelectric effect* should exist in principle for certain crystal classes (point groups) of magnetically ordered crystals. Concerning the quadratic effects of this type, they remarked that such effects should exist in principle, even in isotropic materials, but that "they present no interest because of being extremely weak." It is shown in Sections 3.8 and 3.9 of this chapter, however, that such effects are well measurable and have become important when studying magnetic symmetry and magnetic phase transitions. Following Landau and Lifshitz, Dzyaloshinskii demonstrated theoretically in 1959 [10] that the magnetic symmetry of the antiferromagnetic phase of Cr_2O_3 would be appropriate to show a linear magnetoelectric effect with both a magnetic-field-induced polarization and an electric-field-induced magnetization. The predicted electric-field-induced magnetization was measured in 1960 by Astrov [11-12] on Cr_2O_3 (now denoted ME_E effect) and the converse effect, the magnetic-field-induced polarization (now denoted ME_H effect), was first demonstrated on Cr_2O_3 by Folen, Rado and Stalder in 1961 [13].

This chapter surveys the linear and bilinear (quadratic) magnetoelectric effects, which are, in principle, thermodynamically reversible effects; it also explores the possibilities of thermodynamically irreversible domain-switching phenomena between spontaneous ferroelectric polarization, spontaneous ferromagnetic or ferrimagnetic magnetization and spontaneous deformation. Such a coupling was discovered in nickel-iodine boracite $\text{Ni}_3\text{B}_7\text{O}_{13}\text{I}$ [14], in which material the 180° switching of the ferroelectric polarization causes a 90° rotation of the spontaneous magnetization, and a rotation of the magnetic field by 90° causes a 180° switching of the spontaneous electric polarization.

2 Thermodynamic potential

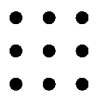
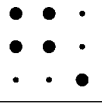
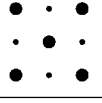
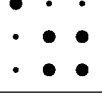
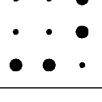
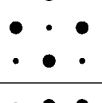
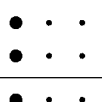
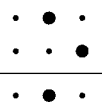
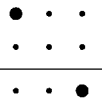
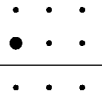
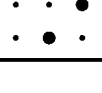
For obtaining an overview of the possible magnetoelectric effects, it is practical to develop the stored free enthalpy function as a limited series [15] and to test thereafter for all 122 magnetic crystallographic point groups (Shubnikov point groups) (Table II of Ref. 15; Table 1 of Ref. 16) which of the terms and which of their tensor coefficients remain invariant under the symmetry operations of the group. For developing the density of stored energy function (free enthalpy) in SI units, we follow Rivera [17], a reference work for detailed definitions, units and measurements. This function, $g(\mathbf{E}, \mathbf{H}; T)$ [$\text{V A s/m}^3 = \text{J/m}^3$], at constant temperature T and with the variables electric field $\mathbf{E}$ [V/m] and magnetic field $\mathbf{H}$ [A/m] (with indexes running from 1 to 3) can be developed as follows:

$$g(\mathbf{E}, \mathbf{H}; T) = \dots + {}^S P_i E_i + {}^S J_i H_i + \frac{1}{2} \varepsilon_0 \varepsilon_{ik} E_i E_k + \frac{1}{2} \mu_0 \mu_{ik} H_i H_k + \alpha_{ik} E_i H_k + \frac{1}{2} \beta_{ijk} E_i H_j H_k + \frac{1}{2} \gamma_{ijk} H_i E_j E_k + \dots, \quad (1)$$

where ε_0 = free-space permittivity $= 1/(c^2 \mu_0)$ [A s/(V m)], ε_{ik} = relative permittivity, μ_0 = free-space permeability $= 4\pi \times 10^{-7}$ [V s/(Am)], μ_{ik} = relative permeability, c = free-space light velocity $\approx 3 \times 10^8$ [m/s], ${}^S \mathbf{P}$ = spontaneous polarization [A s/m^2], ${}^S \mathbf{J}$ = spontaneous magnetization [V s/m^2], α [s/m] = tensor of linear ME effect, “EH” effect, α_{ik} is nonsymmetric in ik , β [s/A] = tensor of bilinear magnetoelectric “EHH” effect, β_{ijk} is symmetric in jk , γ [s/V] = tensor of bilinear magnetoelectric “HEE” effect, γ_{ijk} is symmetric in jk .

In (1), the permittivity term $(1/2)\varepsilon_0 \varepsilon_{ik} E_i E_k$ and the permeability term $(1/2)\mu_0 \mu_{ik} H_i H_k$ are allowed in all crystal symmetry groups, the spontaneous polarization term ${}^S P_i E_i$ in 31 Shubnikov point groups (ferroelectrics and non-ferroelectric pyroelectrics), the spontaneous magnetization term ${}^S J_i H_i$ in 31 partially different Shubnikov groups, and the linear magnetoelectric effect term $\alpha_{ik} E_i H_k$ in 58 Shubnikov groups (19 ferromagnetic and 39 antiferromagnetic groups; for the tensor form of α_{ik} see Table 1). The bilinear magnetoelectric effect term $(1/2)\beta_{ijk} E_i H_j H_k$ is permitted in 66 Shubnikov groups also allowing the piezoelectric effect (structurally non-centrosymmetric diamagnetic, paramag-

Table 1 Tensor forms of the linear magnetoelectric effect.

2	1	$\bar{1}'$				
0	(2)	(m')	(2/m')	1 st setting		
3	2	m'	2/m'	2 nd setting		
0	(2)	(m')	(2/m')	3 rd setting		
0	(2')	(m)	(2'/m)	1 st setting		
3	2'	m	2'/m	2 nd setting		
0	(2')	(m)	(2'/m)	3 rd setting		
3	222	(2m'm')	(m'2m')	m'm'2	m'm'm'	
3	2'2'2	mm2	(2'mm')	(m2'm')	mmm'	
1	(2'22')	(m2m)	mm'2'	(2'm'm)	(mm'm)	
0	(22'2')	(2mm)	(m'2'm)	(m'm2')	(m'mm)	

15

netic, and spin-ordered commensurate and incommensurate materials). The bilinear magnetoelectric effect term $(1/2)\gamma_{ijk}H_iE_jE_k$ is permitted in 66 Shubnikov groups allowing also the piezomagnetic effect, possible in all 31 point groups permitting (weak) ferromagnetism or (weak) ferrimagnetism, and 35 ones allowing antiferromagnetism [15]. The tensor form of β_{ijk} is identical to that of the piezoelectric effect tensor, and the tensor form of γ_{ijk} is identical to that of the piezo-

Table 1 (Continued).

8	$4 \quad \bar{4} \quad 4/m' \quad 3 \quad \bar{3}' \quad 6 \quad \bar{6}' \quad 6/m'$											
3	$4' \quad \bar{4} \quad 4'/m'$											
11	$422 \quad 4m'm' \quad [\bar{4}'2m' \quad \bar{4}'m'2] \quad 4/m'm'm' \quad 32 \quad 3m' \quad \bar{3}'m'$ $622 \quad 6m'm' \quad [\bar{6}'m'2 \quad \bar{6}'2m'] \quad 6/m'm'm'$											
5	<table><tr><td>$4'22'$</td><td>$4'm'm'$</td><td>$\bar{4}2m$</td><td>$\bar{4}m'2'$</td><td>$4'/m'm'm'$</td></tr><tr><td>$4'2'2'$</td><td>$4'mm'$</td><td>$\bar{4}m2$</td><td>$\bar{4}2'm'$</td><td>$4'/m'mm'$</td></tr></table>	$4'22'$	$4'm'm'$	$\bar{4}2m$	$\bar{4}m'2'$	$4'/m'm'm'$	$4'2'2'$	$4'mm'$	$\bar{4}m2$	$\bar{4}2'm'$	$4'/m'mm'$	
$4'22'$	$4'm'm'$	$\bar{4}2m$	$\bar{4}m'2'$	$4'/m'm'm'$								
$4'2'2'$	$4'mm'$	$\bar{4}m2$	$\bar{4}2'm'$	$4'/m'mm'$								
11	$42'2' \quad 4mm \quad [\bar{4}'2'm \quad \bar{4}'m2'] \quad 4/m'mm \quad 32' \quad 3m \quad \bar{3}'m'$ $62'2' \quad 6mm \quad [\bar{6}'m2' \quad \bar{6}'2'm] \quad 6/m'mm$											
5	$23 \quad m\bar{3}' \quad 432 \quad \bar{4}'3m' \quad m\bar{3}'m'$											

43 + 15 = 58

- Key to notation:
- zero component
 - non-zero component
 - equal components
 - components with equal modulus but opposite sign
- Bold numerals: number of magnetic point groups.
Group symbols in round brackets have not to be counted.
Group symbols in square brackets belong to the same group.

magnetic effect tensor. Both tensors have been determined by Ascher [18]. Tables in a more convenient form are given by Grimmer [19]. Ensembles of Shubnikov point groups permitting the invariants E, H, EH, EHH and HEE can be found in Table II of Ref. 15 and Table 1.6.8.4 of Ref. 20.

3 Linear and bilinear magnetoelectric effects

When applying a magnetic field to a magnetoelectric crystal and collecting the induced charges, we use the notation ME_H effect; and when applying an electric field and measuring the induced magnetization, we speak of ME_E effect.

3.1 ME_H effects

Upon differentiation of the negative of (1) with respect to E_k , we obtain the components of the total polarization P_k [A s/m²]:

$$P_k(\mathbf{E}, \mathbf{H}; T) = -\frac{\partial g}{\partial E_k} = \dots + {}^sP_k + \varepsilon_0 \varepsilon_{ik} E_i + \alpha_{ki} H_i + \frac{1}{2} \beta_{kij} H_i H_j + \gamma_{ijk} H_i E_j. \quad (2)$$

From this result we obtain the basic equation in SI units (wherein $\mathbf{P}$ and $\mathbf{D}$ have the same units) for the magnetic-field-induced polarization of the ME_H effect at $\mathbf{E} = \mathbf{0}$:

$$\frac{Q}{S_k} = D_k = P_k = \alpha_{ki} H_i + \frac{1}{2} \beta_{kij} H_i H_j, \quad (3)$$

where Q [C] = [A s]) is the bound charge induced by the magnetic field and S_k [m²] is the surface of one of the crystal's electrodes. Equation (3) is the basic equation for the quasi-static and dynamic ME_H measuring methods.

3.2 ME_E effects

By differentiating $-g$ with respect to H_k , we obtain the components of the total magnetization J_k [V s/m²]:

$$J_k(\mathbf{E}, \mathbf{H}; T) = -\frac{\partial g}{\partial H_k} = \dots + {}^sJ_k + \mu_0 \mu_{ik} H_i + \alpha_{ik} E_i + \beta_{ijk} E_i H_j + \frac{1}{2} \gamma_{kij} E_i E_j. \quad (4)$$

For $\mathbf{H} = \mathbf{0}$, the total electric-field-induced magnetization becomes

$$J_k = +\alpha_{ik} E_i + \frac{1}{2} \gamma_{kij} E_i E_j. \quad (5)$$

Equation (5) is the basic equation for measuring the quasi-static and dynamic ME_E effects using SI units [17].

3.3 Tensor form of the linear magnetoelectric effect

In Table 1 (which is a slightly modified version of Rivera's Figure 2 [17]) the 18 forms of the linear magnetoelectric effect tensor α_{ik} of the 58 Shubnikov point groups that allow the effect are presented. To obtain the total number of 58 groups,

the ones between brackets must not be counted, since representing only the possible alternate orientations of the group without brackets. Groups in square brackets have to be counted only once; the choice depends on the orientation of the smallest unit cell. These different details may appear as a redundancy, but have proven to be useful because erroneous attributions of coefficients are frequently found in the literature, due to arbitrary use of differently chosen “standard notations,” which, in fact do not exist [21], as well to erroneous tensor orientation in certain published tables (noted in Ref. 17). In the case of the monoclinic point groups usually the “Second setting” is used in crystallography as the standard form, rarely the “First setting,” with the monoclinic principal axis being b ($\parallel y$) and c ($\parallel z$), respectively, whereas the “Third setting” with the principal axis being a ($\parallel x$) is never used in crystallography. In the original Table 1 [17], [22], only the tensor form of the “First” and “Second” settings is given. However, in a recent neutron scattering and magnetoelectric study of the magnetic structure of LiCoPO_4 [23], it turned out that in order to attribute the correct monoclinic magnetic point group it was necessary to examine the compatibility of the magnetoelectric tensor form of all three types of setting, the “Third setting” having been left to be the only acceptable one. That is why an extended and corrected version of the original Table 1 [17], [22] has been presented in this chapter for the convenience of the experimentalist.

3.4 Tensor form of the bilinear (quadratic) magnetoelectric effects

As stated previously, the tensors β and γ (“ α_{ijk} ” in Ascher’s notation [18]), have the same form as the piezoelectric and piezomagnetic tensors, respectively, given by Grimmer in Figure 1 of Ref. 19, *in transposed matrix representation* the various forms for the 66 Shubnikov groups allowing piezoelectricity and for the 66 groups allowing piezomagnetism (necessitating magnetic long-range order). In Grimmer’s Figure 1 the groups allowing piezoelectricity and the tensor β are found in the central columns labeled t , and the groups permitting piezomagnetism and tensor γ are given in the columns labeled s (for more detail see item 2.5 of Ref. 17).

3.5 Measuring units

There exist three systems of units for the coefficients of the linear magnetoelectric effect, which have been used by different authors: the *Gaussian system in conventional form*, the *Gaussian system in rationalized form*, and the SI. Because the first two have not always been clearly explained in publications, various errors in cited data have found their way into the literature because of use or nonuse of the coefficient 4π . Therefore, we reproduce in Table 2 the conversion of units of Ref. 17, Figure 2. The relation between the three kinds of units is the following:

$$4\pi\alpha'_{ij} = \alpha'_{ij} = 3 \times 10^8 \alpha_{ij}, \quad (6)$$

Table 2 Conversions between (nonrationalized) Gaussian and SI units for some often-used quantities and for the coefficients of the linear (α) and bilinear (β, γ) magnetoelectric effects, defined in (1), (2) and (4) of this chapter (reproduced from Figure 1 of Ref. 17).

Conversion of units		
Gaussian units	SI units	
Electricity		
$E' = 1$ [statvolt/cm]	$E = 3 \times 10^4$	[V/m]
$D' = 1$ [statcoulomb/cm ²]	$D = 1/(12\pi 10^5)$	[C/m ²]
$Q' = 1$ [statcoulomb]	$Q = 1/(3 \times 10^9)$	[C]
$C' = 1$ [cm]	$C = 1/(9 \times 10^{11})$	[F]
Magnetism		
$H' = 1$ [Oe]	$H = 1000/4\pi \approx 79.6$	[A/m]
$B' = 1$ [Gauss]	$B = 1/10^4$	[V s/m ² = Tesla]
Density of Energy		
$w' = 1$ [g/cm s ²] = erg/cm ³ = Oe statvolt/cm]	$w = 0.1$	[kg/(ms ²) = J/m ³ = V a s/m ³]
Magnetoelectric effects		
$\alpha' = 1$ [–]	$\alpha = 1/(3 \times 10^8)$	[s/m]
$\beta' = 1$ [= Oe ^{–1}]	$\beta = 4\pi/(3 \times 10^{11})$	[s/A]
$\gamma' = 1$ [cm/statvolts]	$\gamma = 1/(9 \times 10^{12})$	[s/V]

where α_{ij}^r is expressed in *rationalized Gaussian units*, α_{ij}' in *Gaussian (conventional, nonrationalized) units*, and α_{ij} in *SI units*. The conversion between the coefficients of β_{kij} and γ_{kij} (in SI units) and β_{kij}' and γ_{kij}' is also given in Table 2.

3.6 Examples of materials with a linear magnetoelectric effect

In Table 3 we present a nonexhaustive list of materials [20] with their maximum value α_{max}^r for the coefficient of the linear magnetoelectric effect given in *rationalized Gaussian units*. It is noteworthy, however, that data is mostly published in conventional (cgs) Gaussian and/or in SI units. The highest values for the coefficients are found among rare-earth compounds, such as DyAlO₃, TbAlO₃, DyPO₄ and in particular TbPO₄, for which a maximum value $|\alpha_{\text{obs}}^r|_{\text{max}} = 1.1 \times 10^{-2}$ at $T = 1.92$ K was measured [26], which is the highest value of a magnetoelectric coefficient ever reported. However, since some doubt subsisted about a factor 4π and hence about the true order of magnitude for α of TbPO₄, (whether expressed in rationalized or nonrationalized Gaussian units), a re-examination of the coefficient was undertaken recently [27]. These experiments confirmed that cgs rationalized Gaussian units were used by Rado *et al.*, and yielded a value of $\alpha_{\text{max}} = 1.75 \times 10^{-2} = 735$ ps/m at 1.50 K, even surpassing that reported originally by Rado *et al.* [26] by a factor of 1.6.

Among the 3*d* transition element compounds, the cobalt boracites Co₃B₇O₁₃Br and Co₃B₇O₁₃I reach the highest values of α , followed closely by lithium cobalt

Table 3 Magnetoelectric coefficients of some materials with a linear magnetoelectric effect (reproduced from Table 1.6.8.3 of Ref. 20).

Compound	T_N or T_C (K)	Magnetic point group	α'_{obs} (max.) (rationalized Gaussian units)	References*
Fe ₂ TeO ₆	219	4/m'm'm'	3×10^{-5}	[7]-[9], [69]
DyAlO ₃	3.5	m'm'm'	2×10^{-3}	[11]-[13]
GdAlO ₃	4.0	m'm'm'	1×10^{-4}	[14]
TbAlO ₃	4.0	m'm'm'	1×10^{-3}	[12], [15]-[17]
TbCoO ₃	3.3	mmm'	3×10^{-5}	[12], [16], [18]
Cr ₂ O ₃	318	$\bar{3}$ 'm'	1×10^{-4}	[45]-[49], [69-70], W162
Nb ₂ Mn ₄ O ₉	110	$\bar{3}$ 'm'	2×10^{-6}	[52-53]
Nb ₂ Co ₄ O ₉	27	$\bar{3}$ 'm'	2×10^{-5}	[52-53]
Ta ₂ Mn ₄ O ₉	104	$\bar{3}$ 'm'	1×10^{-5}	[53]
Ta ₂ Co ₄ O ₉	21	$\bar{3}$ 'm'	1×10^{-4}	[53]
LiMnPO ₄	35	m'm'm'	2×10^{-5}	[55-56], [58], [60]
LiFePO ₄	50	mmm'	1×10^{-4}	[57-58]
LiCoPO ₄	22	mmm'	7×10^{-4}	[54-55], R161
LiNiPO ₄	23	mmm'	4×10^{-5}	[54-55], [61]
GdVO ₄	2.4	4'/m'm'm	3×10^{-4}	[69]
TbPO ₄	2.2	4'/m'm'm	1×10^{-2}	[65-66] (see Section 3.6)
DyPO ₄	3.4	4'/m'm'm	1×10^{-3}	[67-68]
HoPO ₄	1.4	4'/m'm'm	2×10^{-4}	[71]
Mn ₃ B ₇ O ₁₃ I	26	m'm2'	2×10^{-6}	C204
Co ₃ B ₇ O ₁₃ Cl	12	m	3×10^{-4}	S204
Co ₃ B ₇ O ₁₃ Br	17	m'm2'	2×10^{-3}	88C1
Co ₃ B ₇ O ₁₃ I	38	m'm2'	1×10^{-3}	90C3
Ni ₃ B ₇ O ₁₃ I	61.5	m'	2×10^{-4}	[73-74], [76]-[78], 90C2
Ni ₃ B ₇ O ₁₃ Cl	9	m'm2'	3×10^{-4}	74R2, 91R1
Cu ₃ B ₇ O ₁₃ Cl	8.4	m'm'2	3×10^{-6}	88R1
FeGaO ₃	305	m'm2'	4×10^{-4}	[83]-[85]
TbOOH	10.0	2/m'	4×10^{-4}	[112]
DyOOH	7.2	2/m'	1×10^{-4}	[91], [112]
ErOOH	4.1	2'/m	5×10^{-4}	[92], [112]
Gd ₂ CuO ₄	6.5	mmm'	1×10^{-4}	W161
MnNb ₂ O ₆	4.4	mmm'	3×10^{-6}	[100-101]
MnGeO ₃	16	mmm'	2×10^{-6}	[97]-[99]
CoGeO ₃	31	mmm'	1×10^{-4}	[69]
CrTiNdO ₅	13	mmm'	1×10^{-5}	[69], [88]

*Explanation of the signs: For numbers see Ref. [24]; for 88C1, 90C3, 88R1, 90C2, 74R2, 91R1 see Ref. 25 and for W162, R161, C204, S204, and W 161 see Refs. 108 (Part II, p. 141), 108 (Part I, p. 147), 109 (p. 125), 109 (p. 57) and 108 (Part I, p. 133), respectively.

phosphate LiCoPO₄ [28-29]. The coefficient α_{yx} (at 4.2 K) of that latter compound is, however, still 24 times smaller than that of the revised maximum value of TbPO₄. For developing new magnetoelectric materials, Co²⁺ ions are also promising because of their high spin orbit coupling and high anisotropy, but structures with a higher density of magnetoelectrically active ions would also be necessary if higher Néel and Curie temperatures are aimed for. Composite arrays with overall

magnetoelectric symmetry seem to be more promising for technical applications, e.g., spectral nonreciprocity (see Section 5.1).

3.7 Spontaneous ferroelectric versus magnetic-field-induced polarization

In order to obtain a better sense of what these magnetoelectric coefficients mean, a comparison may be made between the magnitude of the more familiar spontaneous polarization of ferroelectrics and the magnitude of the magnetic field necessary to induce a magnetoelectrically generated polarization of comparable order of magnitude. As an example we take the relatively high coefficient $\alpha_{\text{obs}}^r = 2 \times 10^{-3}$ of $\text{Co}_3\text{B}_7\text{O}_{13}\text{Br}$ at 4.2 K (Table 3). If we want to reach, for example, the magnetoelectrically induced polarization $P_k = 7.6 \times 10^{-2} \mu\text{C}/\text{cm}^2$, i.e., the value of the very small spontaneous polarization of the weakly ferromagnetic/ferroelectric boracite $\text{Ni}_3\text{B}_7\text{O}_{13}\text{I}$ at 4.5 K [30], then a magnetic field of 38 tesla (T) would be necessary. It is noted that at such a high field strength, the magnetoelectric phase (m'/m^2) of $\text{Co}_3\text{B}_7\text{O}_{13}\text{Br}$ would probably no longer be stable. In order to reach magnetoelectrically the magnitude of spontaneous polarization of the well-known ferroelectric BaTiO_3 , i.e., $25 \mu\text{C}/\text{cm}^2$, completely unrealistic high magnetic fields would be required. Thus, the linear magnetoelectric effect is a rather weak effect. This is one of the reasons why no technical applications have so far become known. However, in spite of the smallness of the magnetoelectric coefficients, they can be detected with great sensitivity, in particular by dynamic techniques with very low modulating fields [17], [28]. By reorienting the spontaneous polarization of a ferroelectric by means of a magnetic field (see Example III in Section 4), much larger charges may be induced in principle.

3.8 The bilinear magnetoelectric *EHH* effect

The *EHH* effect was originally named *paramagnetoelectric effect* (“PME” effect) by Hou and Bloembergen [31], who measured it on the structurally non-centrosymmetric paramagnetic, tetragonal crystal $\text{NiSO}_4 \cdot 6\text{H}_2\text{O}$ of point group 4221'. Later it was measured on several magnetically ordered ferroelectrics like boracites (superposed on the linear effect) [32], BaMnF_4 [33–34] and the perovskite BiFeO_3 [34–35]. The latter compound is antiferromagnetically incommensurate in the ferroelectric phase, in which the linear ME effect is cancelled due to the incommensurate structure, while the *EHH* effect survives, simulating a paramagnetic non-centrosymmetric phase. By rotating in a magnetic field a ferroelectric/ferroelastic single domain in its three orthogonal principle planes, all allowed components of the tensor β_{ijk} have been determined for BiFeO_3 [35]. The microscopic theory of the *EHH* effect has so far only been discussed for $\text{NiSO}_4 \cdot 6\text{H}_2\text{O}$, where the electric field change in the crystal field splitting (D) is made responsible for the effect [31].

3.9 The bilinear magnetoelectric *HEE* effect

The *HEE* effect, which requires magnetically ordered structures with one of the 66 point groups allowing the piezomagnetic effect (all 31 groups of the ferro(i)magnets and 35 groups of antiferromagnetics allow the effect), has been measured on ferrimagnetic garnets [36], but no other work on other compounds has been reported since then. The *HEE* term is also responsible for the linear magneto-optic effect (see Section 5.1).

3.10 ME_H and ME_E effect measurements

For measuring the *EH* and *EHH* effects, both dynamic and quasi-static measurement techniques have been applied in the past. The quasi-static ME_H method has been developed to great perfection [17], [28] and can be considered as the most straightforward one. It can be used up to very high static magnetic fields [29]. The method is very accurate because the charges induced by the magnetic field can be measured directly without necessity of calibration, whereas the ME_E methods need calibration. Moreover, if transparent gold electrodes are used, control in polarized light of the ferroelastic (ferroelectric) domain state (due to birefringence contrast) and the ferromagnetic domains (due to spontaneous Faraday rotation, if noteworthy in the visible) is of invaluable help. Pulsed field methods have also been developed for measuring the ME_H effect at very high fields [37]. For measuring the *EHH* effect the quasi-static ME_H method is also useful, except in case of very small tensor components β_{ijk} . In that case the dynamic method, developed by Rivera [17], excels due to an amplification effect that is proportional to the strength of a magnetic DC bias field. The ME_E method, initiated by Astrov [12], may find a revival by using a SQUID magnetometer, as has been demonstrated by Kita [38].

3.11 Optical magnetoelectric measurements

Regarding optical measurements of the magnetoelectrically generated Faraday rotation (first measured by O'Dell [39]) or Kerr effect, measurements along the optical axis, i.e., in the absence of birefringence or bireflectance, are relatively easy with classical high-sensitivity polarimeters, as was shown for Cr_2O_3 [40]. However, if birefringence is superposed on such a weak Faraday rotation or a weak Kerr effect on bireflectance, the use of the Sagnac interferometer has been proposed for magnetoelectric materials [41]. This technique has the advantage of being sensitive only to effects due to broken time reversal symmetry, such as Faraday rotation and magneto-optic Kerr effect and not to reciprocal properties, such as reciprocal birefringence and reciprocal bireflectance. The so-called HAUP (High Accuracy Universal Polarimeter) [42]-[44] and a polarimetric method based on tilting a crystal [45-46] also separate rotation from birefringence and may be used in principle, but so far no measurements on magnetoelectrics have been reported with these three methods.

3.12 The piezomagnetoelectric effect $EH\sigma$

The piezomagnetoelectric effect corresponds to the term $\pi_{ijkl} E_i H_j \sigma_{kl}$ of the density of stored free enthalpy (not incorporated in (1)), for which the fourth-rank tensor π_{ijkl} ($i, j, k, l = 1, 2, 3$)—in matrix form, $\pi_{\mu\nu}$ ($\mu = 1$ to 9 and $\nu = 1$ to 6)—has been determined for all magnetic Shubnikov point groups admitting the effect [47]. These groups are exactly those in Table II of Ref. 15 which in the column “Type of ordering” have no D in the subcolumn “Magnetic” and no O in the subcolumn “Electric” (see the footnote on page 270 of Ref. 47). The effect has different facets. It allows in principle (i) the induction of a polarization $P_i = \dots + \pi_{ijkl} H_j \sigma_{kl}$ by applying a magnetic field H_j and stress σ_{kl} simultaneously, (ii) the induction of a magnetization $M_j = \dots + \pi_{ijkl} E_i \sigma_{kl}$ by applying stress σ_{kl} and an electric field E_i simultaneously, and (iii) the induction of a deformation $\varepsilon_{kl} = \dots + \pi_{ijkl} E_i H_j$ by applying an electric field E_i and a magnetic field H_j simultaneously. The converse effects are also possible. By applying a magnetic bias field one creates a piezoelectric, by applying an electric bias field a piezomagnetic, and by applying stress a linear magnetoelectric material. The latter route was attempted, although without success, for demonstrating the effect on antiferromagnetic Cr_2O_3 (point group $\bar{3}'m'$) [48]. For finding the predicted effect on antiferromagnetic LiCoPO_4 (point group mmm'), a magnetic field bias and a piezoelectric oscillator configuration were used [49]. However, also for this material, the PME effect could not be established. It was supposed that the mechanical stiffness of LiCoPO_4 was too high. Magnetically ordered metal-organic compounds are possibly better candidates. Recently, however, the point group of LiCoPO_4 was found to be monoclinic $2'$, rather than orthorhombic mmm' [23]. Thus, a re-examination of the experiments should be considered, since the potential presence of time-reversed domains of the lower symmetry phase with their opposite sign of the induced polarization (or magnetization) may have led to cancellation of the total magnetoelectric signal.

4 Spontaneous magnetoelectric effects and related phenomena

It is common to allocate the qualifier *spontaneous* to quantities such as the spontaneous magnetization ^SM_i (axial vector), the spontaneous polarization ^SP_i (polar vector), the spontaneous toroidal moment ^ST_i (t-odd polar vector = *axio-polar* [50] vector) [51], or the symmetric second-rank tensor spontaneous deformation $^S\varepsilon_{ij}$. One or more of these quantities appear, usually upon cool-down, “spontaneously” at a first- or second-order phase transition. The question as to whether one, two or all three of the quantities ^SM_i , ^SP_i and ^ST_i are allowed to occur in a particular crystal depends upon its Shubnikov point group. There are 31 groups permitting ^SP_i , 31 partially different groups permitting ^SM_i [52], and 31 partially different ones determined originally for the spontaneous current density and the velocity of a moving crystal [53] which transform like ^ST_i . Two among the three vectors at a

time, i.e., ${}^SM_i$ and ${}^SP_i$, ${}^SP_i$ and ${}^ST_i$ or ${}^ST_i$ and ${}^SM_i$, are allowed in 13 point groups, whereas all three vectors are permitted in only 9 point groups [15-50]. This *magic trinity* of symmetry was originally demonstrated by Ascher [50-53].

An interesting question arises in the case of a crystal having both ${}^SM_i$ and ${}^SP_i$, i.e., whether an external magnetic field can switch ${}^SP_i$ by acting via ${}^SM_i$, and *vice versa*, whether an external electric field can switch ${}^SM_i$ by acting on ${}^SP_i$. At first this would seem to be impossible since the spontaneous magnetization remains invariant upon sign reversal of the electric field and spontaneous polarization, and the spontaneous polarization remains invariant under sign reversal of the magnetic field and spontaneous magnetization. One can show, however, *that interactions are possible if the phase possessing both ${}^SM_i$ and ${}^SP_i$ is also ferroelastic*. This is a necessary, but not sufficient condition.

The concept of *ferroelasticity* has been introduced by Aizu [54] in analogy with ferromagnetism and ferroelectricity. It occurs in phases with different (degenerate) orientation states (domains) of the *spontaneous deformation* ${}^\varepsilon_{ij}$, which usually forms below a phase transition. At variance, the so-called *co-elastic* [55] also form a spontaneous deformation below a phase transition (e.g., hexagonal $\leftrightarrow$ hexagonal); but there is no degeneracy of its orientation, with the consequence that no ferroelastic domains do occur. Whereas the possibility of the occurrence of the vectors ${}^SM_i$, ${}^SP_i$ and ${}^ST_i$ in a crystal is unequivocally determined by the Shubnikov point group alone, the possibility of occurrence of ferroelasticity in a phase necessitates the knowledge of two point groups—that of the high-temperature phase (called *prototype* or *aristotype*) and that of the low-temperature *ferroic* phase.

For exploring all potential switching and nonswitching cases, one can use two tools: (i) the thermodynamic classification [15] into *primary*, *secondary* and *tertiary ferroics* [56]-[58] giving the different terms of the free energy allowed by symmetry and the corresponding driving forces for domain switching and reorientation (see Table 4), and (ii) Aizu's classification of the prototype/ferroic phase point group pairs, called *species* [54], giving information on full, partial or no coupling between the different types of domain during switching. In the sense of Aizu, “full(y)” means that the externally applied magnetic field, electric field or stress can control (reorient or switch) all of the symmetry-allowed ferromagnetic, ferroelectric, or ferroelastic domain states, respectively. “Partial(ly)” means that the external constraint is not able to switch or reorient all possible domains individually.

There exist 773 species of groups, tabulated by Aizu [54]. The possibility of coupling or noncoupling between the spontaneous quantities ${}^SM_i$, ${}^SP_i$ and ${}^ST_i$ and ${}^\varepsilon_{ij}$ is determined by the type of species. In order to obtain an overview of all switching possibilities, the 773 species (formed by all 122 Shubnikov point groups) have been grouped into 30 ensembles [59], each one being characterized by the same type of domain switching and reorientation possibilities (Tables I and II of Ref. 59)). All possible combinations between *fully* and *partially ferroelectric*, *nonferroelectric*, *fully* and *partially ferroelastic*, *nonferroelastic (coelastic)*, *fully* and *partially ferromagnetic*, *antiferromagnetic* and *paramagnetic/diamagnetic* have been considered and illustrated with examples (Table II of Ref. 59). Three such examples are given here:

Table 4 Ferroic "driving forces" of domain switching and reorientation due to differences in domain states. [Reprinted figure with permission from I. Kornev et al, *Phys. Rev. B* **62** (2000). Copyright (2000) by the American Physical Society, <http://link.aps.org/abstract/PRB/v62/p12247>.]

Type of ferroic	"Driving force" $\Delta g \propto$	States differ in:	
Primary ferroics			
Ferromagnetic	$\Delta^S M_i H_i$	spontaneous magnetization	$^S M_i$
Ferroelectric	$\Delta^S P_i E_i$	spontaneous polarization	$^S P_i$
Ferrotoroidics	$\Delta^S T_i S_i$	spontan. toroidal moment	$^S T_i$
Ferroelastic	$\Delta^S \varepsilon_{ij} \sigma_{ij}$	spontaneous deformation	$^S \varepsilon_{ij}$
Secondary ferroics****			
Ferrobimagnetic	$\Delta \chi_{ij} H_i H_j$	magnetic susceptibility	χ_{ij}
Ferrobielectric	$\Delta \kappa_{ij} E_i E_j$	electric susceptibility	κ_{ij}
Ferrobielastic	$\Delta s_{ijkl} \sigma_{ij} \sigma_{kl}$	elastic compliance	s_{ijkl}
Ferroelastoelectric	$\Delta d_{ijk} E_i \sigma_{jk}$	piezoelectric coefficient	d_{ijk}
Ferromagnetoelastic	$\Delta q_{ijk} H_i \sigma_{jk}$	piezomagnetic coefficient	q_{ijk}
Ferromagnetolectric***	$\Delta \alpha_{ij} E_i H_j$	magnetolectric* coeff.	α_{ij}
Tertiary ferroics			
Ferrotriectricity	$\Delta \kappa_{ijk} E_i E_j E_k$	nonlinear electric suscept.	κ_{ijk}
Ferrotrimagnetism	$\Delta \chi_{ijk} H_i H_j H_k$	nonlinear magnetic suscept.	χ_{ijk}
Ferroelastobielectricity	$\Delta \gamma_{iljk} \sigma_{jk} E_i E_l$	electrostriction coefficient	γ_{iljk}
Ferroelastobimagnetism	$\Delta \lambda_{iljk} \sigma_{jk} H_i H_l$	magnetostriction coefficient	λ_{iljk}
Ferromagnetobielectricity	$\Delta \gamma_{ijk} H_i E_j E_k$	magnetobielectric** coeff.	γ_{ijk}
Ferroelectrobimagnetism	$\Delta \beta_{ijk} E_i H_j H_k$	electrobimagnetic** coeff.	β_{ijk}
Ferromagnetoelastoelectricity	$\Delta \pi_{ijkl} H_i E_j \sigma_{kl}$	piezomagnetolectric coeff.	π_{ijkl}
Ferromagnetobielasticity	$\Delta \psi_{iklmn} H_i \sigma_{kl} \sigma_{mn}$	magnetobielastic coeff.	ψ_{iklmn}

*Linear magnetolectric effect.

**Bilinear magnetolectric effects.

***N.B.: In the nomenclature of Newnham, the term "ferromagnetolectric" is somewhat infelicitous because it is also sometimes used in literature for a ferroelectric being simultaneously ferro(i)magnetic in the same phase. If the toroidic domains would have been discovered before the ferromagnetic ones, we might have today on equal right a long list of primary, secondary and tertiary "Toroido"-s of "Toroics" instead of "Ferro"-s!

****Since the electrotoroidic and magnetotoroidic terms can be expressed by the magnetobielectric and electrobimagnetic ones, resp., due to $S_i = (\mathbf{E} \times \mathbf{H})_i$, they have not been specially mentioned in the list of secondary ferroics.

Example I. The most interesting ensemble is *fully ferromagnetic, fully ferroelectric and fully ferroelastic* (Number 1 of Table I of Ref. 59), which offers the maximum coupling *in principle*: the magnetic field can produce all allowed ferromagnetic domains, all allowed ferroelectric domains, and all allowed ferroelastic domains. An electric field can control all ferroelectric and all ferroelastic domain states because the ferroelectric and ferroelastic domains are identical. It can also reorient the ferromagnetic spontaneous magnetization direction by angles other than 180° , but it cannot reverse the spontaneous magnetization by 180° inside a ferroelastic domain. The same holds true for applied stress. All other kinds of ensembles have inferior coupling capabilities. As an illustration of a species of Ensemble Number 1, we take the species $43m1'Fm'_y2'_z$, which is realized for many homologous compositions of boracites (Table II of Ref. 59). Here, $^S\mathbf{P}$ is aligned along the orthorhombic $2'_z$ axis, $^S\mathbf{M}$ perpendicular to the m_y plane, and $^S\mathbf{T}$ perpendicular to

the m'_x plane. Hence, the directions of the three vectors and the orientation of $^S\varepsilon_{ij}$ are rigidly coupled along the three crystallographic axes. Therefore, an electric field switching of $^S\mathbf{P}$ by 180° rotates $^S\mathbf{M}$, $^S\mathbf{T}$ and $^S\varepsilon_{ij}$ by 90° in the plane perpendicular to the $2'_z$ axis. *Vice versa*, the rotation of a sufficiently strong magnetic field by 90° from the y (SM) to the x (ST) direction provokes a 180° degree switching of $^S\mathbf{P}$ and a 90° rotation of $^S\mathbf{M}$, $^S\mathbf{T}$ and of the spontaneous deformation tensor $^S\varepsilon_{ij}$ around $2'$. It has to be emphasized that symmetry only describes the possibilities of switching and reorientation, but that in practice the coercive fields may be so high that these possibilities cannot really be exploited. Thus, while the switching of boracites with species $\bar{4}3m1'Fm'm2'$ induced by the electric field is straightforward in principle, the magnetic-field-induced 180° switching of $^S\mathbf{P}$ and the reorientation of $^S\mathbf{M}$, $^S\mathbf{T}$ and $^S\varepsilon_{ij}$ is not possible at low temperatures due to too high ferroelectric/ferroelastic coercive fields. Only in nickel iodine boracite ($Ni_3B_7O_{13}I$), in which coexisting ferroelectricity and (weak) ferromagnetism were discovered, and wherein the ferroelectric/ferroelastic and ferromagnetic Curie temperatures coincide at 61 K, both the electric-field- and the magnetic-field-induced switching have been demonstrated [14]. The initially admitted species $\bar{4}3m1'Fm'_xm_y2'_z$ [14], [60] was later determined to be $\bar{4}3m1'Fm'_x$ [30]. The orthorhombic symmetry $m'_xm_y2'_z$ was in fact mimicked by lamellar domains of the monoclinic phase m'_x . Species $\bar{4}3m1'Fm'_x$ also belongs to the interesting ensemble of species Number 1 [59].

Example II. Let us also cite species $mmm1'F2$, permitting $^S\varepsilon_{ij}$, as well as $^S\mathbf{P}$, $^S\mathbf{M}$ and $^S\mathbf{T}$ along the monoclinic polar 2 axis, but all properties are “partial,” hence excluding any coupling between the spontaneous quantities [59].

Example III. For terbium molybdate, $Tb_2(MoO_4)_3$, a paramagnetic fully ferroelectric/fully ferroelastic compound (species $\bar{4}2m1'Fm_xm_y2_z1'$, ensemble 28), it was shown [61] that a magnetic field of 10 T alone, applied along the orthorhombic b axis at 78 K, aligns the a axis along the field and reverses the spontaneous polarization along the 2_z axis by 180° . This behavior is in agreement with symmetry and can be understood as being caused by a ferroelastic “driving force” $\Delta g \propto \Delta ^S\varepsilon_{ij}\sigma_{ij}$ (see Table 4) with the stress σ_{ij} being generated internally via an interaction involving magnetostriction and elastic compliance [59].

5 Selected applications

A great number of potential device applications of magnetoelectric effects, mostly based on the nonreciprocal (optical) properties, have already been imagined and critically discussed in detail as early as 1974 [62] (for a more recent discussion of nonreciprocity in magnetoelectric crystals see Refs. 63 and 64). None of them have found direct technological use so far, essentially because of the relative smallness of the magnetoelectric coefficients. However, there exist extremely useful scientific applications of the magnetoelectric effects in solid state physics and crystallography, and a selection of these is now provided.

5.1 Optical effects

The interaction of light with magnetoelectric crystals has been thoroughly discussed by Pisarev [40]. New nonreciprocal phenomena, allowed due to the particular symmetry of linear magnetoelectric materials (i.e., the combined space-time reversal operation $\bar{1}' = \bar{1}1'$), have become possible. One phenomenon is the *electric-field-induced nonreciprocal (Faraday) rotation*, first demonstrated on anti-ferromagnetic Cr_2O_3 [39] and more recently in detail on the same compound [40]. Other phenomena which have been observed in Cr_2O_3 are the *gyrotropic birefringence* (leading to a nonreciprocal rotation of the optical indicatrix), and the nonreciprocal reflection of light. These phenomena change sign for time-reversed magnetic domains [40].

A related nonreciprocal phenomenon, allowed in all point groups permitting the linear magnetoelectric effect, is referred to as *spectral nonreciprocity* [65]. It is characterized by

$$\omega(\mathbf{k}) \neq \omega(-\mathbf{k}), \quad (7)$$

where ω represents the angular frequency and $\mathbf{k}$ the wave vector. This effect was predicted by Fuchs as early as in 1965 [66] and was discussed extensively by O'Dell [3]. Because of the smallness of the effect in ordinary magnetoelectric crystals, no related technical application in optics and microwave technology has so far been possible. Recently, however, the engineering of complex periodic arrays, made up of at least one magnetically ordered component—not necessarily allowing the magnetoelectric effect—and one dielectrically anisotropic component, has been proposed. The overall magnetic point group symmetry of the array should allow the linear magnetoelectric effect. Such arrays have been considered theoretically with a view to creating *nonreciprocal magnetic photonic crystals* with a *strong* spectral nonreciprocity, using generic readily available nonmagnetoelectric components such as ferrites and anisotropic dielectrics, satisfying generic constitutive relations

$$\mathbf{D} = \hat{\epsilon}\mathbf{E}, \quad \mathbf{B} = \hat{\mu}\mathbf{H}, \quad (8)$$

where $\mathbf{D}$, $\mathbf{E}$, $\mathbf{B}$ and $\mathbf{H}$ denote the alternating components of the electromagnetic field [65]. This concept appears highly attractive, since photonic crystals are certainly bound to become important in information processing technology [67-68].

Another optical phenomenon, which is associated with the bilinear *HEE* invariant of the thermodynamic potential and useful, for example, for magnetic point group determination, is the *linear magneto-optic (LMO) effect* [69-70]. It is allowed for all 66 piezomagnetic Shubnikov point groups. The tensor form of the LMO effect is identical with that of the piezomagnetic effect (for tensor tables in transposed matrix representation see the columns marked “s” in Figure 1 of Ref. 19) and with that of the *Mockels effect*. The latter was first proposed by Ascher [71] as the linear magneto-optic effect (generated by the *HHH* term), which is

the real analog to the linear electro-optic effect (i.e., the *Pockels effect*) generated by the *EEE* term.

5.2 Poling (spin reversal) of antiferromagnetic domains

Antiferromagnetic domain *poling* is allowed for the 39 antiferromagnetic Shubnikov point groups (among a total of 59; see Table II of Ref. 15), which allow the magnetoelectric $\alpha_{ik} E_i H_k$ term and give herewith rise to the driving force $\Delta g \propto \Delta \alpha_{ij} E_i H_j$ for magnetoelectric domain switching and annealing (see Table 4). This effect was first demonstrated on antiferromagnetic Cr_2O_3 [72-73]. For the determination of the linear magnetoelectric effect, it is obligatory to *pole* the antiferromagnetic domains prior to the measurements, because an equal distribution of mutually time-reversed domains cancels the measuring signal of the linear magnetoelectric effect. Certain higher order terms shown in Table 4 can in principle also be used for poling antiferromagnetic domains [59].

5.3 Magnetic point group determination

By determining experimentally the tensor components of the linear magnetoelectric effect, the magnetic point group of magnetically ordered materials can in principle be determined, particularly if the nuclear structure is known from x-ray scattering and if single crystals, which must be insulating in the magnetically ordered phase, are available.

Single crystals or ferroelastic domains of 1 to 3 mm in size are often sufficient for quantitative magnetoelectric measurements; but even if only ceramic samples can be prepared, one can find out whether only diagonal, off-diagonal, or diagonal and off-diagonal coefficients are occurring [74]. In order to find the correct magnetic point group magnetoelectrically, it is recommended to rotate the magnetic field in the three principal planes of the coordinate system (see Ref. 35, for example); otherwise, an unrecognized twinned state in the magnetic-field-induced phase may simulate a wrong symmetry due to superposition of signals from different tensor components. If that happens, as was the case for Cr_2O_3 [75], the recently developed visualization of (anti)ferromagnetic domains—based on nonlinear optical spectroscopy, i.e., magnetic second harmonic generation contrast between domains differing in spin direction—may be complementary or even more appropriate for point group determination. This technique requires, however, the material to have transparency regions matching the laser wavelength used [76]. This requirement is fulfilled for Cr_2O_3 [77], for example.

With knowledge of the space group of the nuclear structure (which is now usually easily determined by x-ray diffraction) and the magnetic point group derived from magnetoelectric experiments, the number of potential magnetic space groups is drastically reduced. This allows magnetic structure refinements in a few expected magnetic space groups, even when only polycrystalline material is available for diffraction. It must be remarked that, in publications on magnetic structure determinations by neutron scattering, the authors often cannot advance as far as

an unequivocal attribution of the magnetic space group, particularly if only polycrystalline samples of poly-domain multiferroics [78] are at their disposal. In incommensurate magnetic structures, the linear magnetoelectric effect is found to be cancelled and the material may mimic the symmetry of the paramagnetic phase [33]-[35]. Such information can also be useful. Thus, the magnetoelectric effect and its toroidal *extension* (see Section 5.8) can be very precious complementary tools for magnetic structure determinations.

5.4 Determination of magnetic field *versus* temperature phase diagrams

The magnetoelectric effect has also been found to be very valuable for magnetic phase diagram studies if single crystals are available. Here one can go to very high magnetic fields both by pulsed field techniques [37] as well as by high continuous fields [29], [75], [79-80] (up to 20–30 T), for which neutron diffraction equipment is so far not available. The two techniques can be complementary.

5.5 Determination of Néel temperatures and critical exponents

By using the dynamic magnetoelectric measurement technique [17], [28], the Néel temperature (in favorable cases, also the critical exponent) can be determined with great accuracy. An example is provided by the antiferromagnetic phosphate LiCoPO_4 [28].

5.6 Magnetoelectric butterfly hysteresis loops

The so-called magnetoelectric butterfly hysteresis loops are known to be the signature of a (weak) ferro(i)magnetic moment of a material displaying a linear ME effect. Such loops are based on a discontinuous change of sign of the ME signal, coinciding with the coercive field of the 180° switching of the ferromagnetic domains, when measuring the magnetic-field-induced polarization *versus* applied magnetic field. The phenomenon was first observed on the weakly ferromagnetic/ferroelectric/ferroelastic boracite $\text{Ni}_3\text{B}_7\text{O}_{13}\text{I}$ [14] and later also on other homologous boracite compositions [81]-[83].

The 3d-metal(II) lithium phosphates LiM(II)PO_4 were considered for a long time as compensated colinear orthorhombic antiferromagnets (point group mmm' and $\text{mm}'\text{m}$). However, the observation of butterfly loops in LiCoPO_4 [28-29], [79] and LiNiPO_4 [80] stimulated the search for a spontaneous magnetization. Indeed, SQUID measurements revealed an ultraweak one in LiCoPO_4 [84], consistent with recent neutron diffraction data that also pointed towards a lowering of symmetry (to monoclinic, point group $2'$) [23]. Detection of an ultraweak magnetization on LiNiPO_4 followed shortly thereafter [85]. *Thus, magnetoelectric butterfly loops are precious monitors of (weak) ferro(i)magnetism and true magnetic symmetry.*

5.7 Disclosure of problems relating to symmetry property relationships

Whereas the determination of the tensor components of the magnetoelectric effect may yield the magnetic point group and help to finalize the magnetic space group in straightforward cases, more complicated cases may occur wherein the magnetoelectric behavior calls for studying the material in more detail. An example is the phosphate KNiPO_4 [86], for which neutron diffraction and the magnetoelectric effect yield the antiferromagnetic point group $mm2$; whereas the measurement of a spontaneous magnetization, both on powders and single crystals, indicates the point group $m'm'2'$. The magnetoelectric tensor components required for this group are, however, incompatible with the ones observed. These inconsistencies provide inviting prospects for further study for elucidating the correct symmetry.

5.8 Magnetoelectric control of a screw spin structure

Below 20 K the spinel ZnCr_2Se_4 is a helimagnet with the screw spin vector along a pseudo-cubic $\langle 100 \rangle$ direction. By cooling in simultaneous magnetic and electric fields, the rotatory direction of the chirality has been controlled [87]. All aspects of this phenomenon are not yet clear. A similar problem has been discussed theoretically for hexagonal antiferromagnetic insulators with a frustrated spin arrangement of the CsMnBr_3 type. An electric field induced breaking of chiral degeneracy is expected to occur in this compound [88].

5.9 Determination of toroidal contributions to the magnetoelectric signal

As mentioned previously, there exist 31 Shubnikov point groups, 13 ferro(i)magnetic and 18 antiferromagnetic ones, thereby permitting a spontaneous toroidal moment $^S T_i$, which is a polar vector changing sign both under space and time reversal [50-51], [53]. A simple geometric representation of a toroidal moment is a solenoid formed into a torus with an even number of windings. The toroidal dipole moment is given by [91-92] (in the cgs system)

$$\mathbf{t} = \int \mathbf{T} dV = \frac{1}{10c} \int [(\mathbf{j} \cdot \mathbf{r})\mathbf{r} - 2r^2\mathbf{j}] d^3r, \quad (9)$$

where $\mathbf{j}$, $\mathbf{r}$ and c are current density, radius vector and free-space light velocity, respectively. The toroidal moment density $\mathbf{T}$ is usually only referred to as the *toroidal moment*, which is in fact $\mathbf{t} = \int \mathbf{T} dV$, analogous to the electric and magnetic dipole moments $\mathbf{p} = \int \mathbf{P} dV$ and $\mathbf{m} = \int \mathbf{M} dV$, with the density of polarization $\mathbf{P}$ and magnetization $\mathbf{M}$, respectively. In a similar way, $\mathbf{M}$ is sometimes called the *magnetic moment* [91]. A toroidal moment can be caused by both orbital ordering and spin ordering. If one is only interested in the spin part of the spontaneous toroidal

moment ${}^S\mathbf{T}$, it has been defined [51], [93-94] by the sum over the spins of all spin-bearing particles in the unit cell, with the cell's center as origin:

$${}^S\mathbf{T} = \frac{1}{2}\mu_B \sum_a \mathbf{r}_a \times \mathbf{S}_a. \quad (10)$$

Therein, $\mathbf{S}_a$ stands for the spin moment and $\mathbf{r}_a$ for the radius vector of the magnetic cation-ion “ a ” in the unit cell. Such a moment ${}^S\mathbf{T}$ has been calculated using nonmagnetic and magnetic structural data for the piezoelectric/piezomagnetic ferromagnet $\text{Ga}_{2-x}\text{Fe}_x\text{O}_3$ [93-94].

It has been shown [51] that the toroidal moment density $\mathbf{T}$ can play the role of an order parameter, with curl $\mathbf{H}$ or electric current density $\mathbf{j}$ (including displacement current $(1/c)\partial\mathbf{D}/\partial t$) serving as the thermodynamically conjugated field for the parameter $\mathbf{T}$, in the same way as the electric field $\mathbf{E}$ and the magnetic field $\mathbf{H}$ play that role for the spontaneous polarization and magnetization, respectively. The use of $\mathbf{T}$ as a phase transition order parameter was initiated by Sannikov [95-96]. He also introduced the term *ferrotoroidic* [97] for materials with a spontaneous toroidal moment, in analogy with *ferroelectric* and *ferromagnetic*. See Ref. 90 for a discussion of the alternate term *ferrotoroidic*, that appears to be more in line with Aizu's nomenclature [54]. By taking the spontaneous toroidal moment as phase transition order parameter, Sannikov was able to explain by means of phenomenological theory the experimentally observed anomalous sharp peaks of the magnetoelectric susceptibility *versus* temperature of ferroelectric/ferromagnetic/ferrotoroidic orthorhombic boracites (point group $m'm'2'$) as the signature of contributions from the spontaneous toroidal moment [96]-[99]. It will be a challenge to verify these contributions predicted by theory, by calculating them from magnetic and nuclear structural data, as was done for the solid solution $(\text{Ga}^{3+}/\text{Fe}^{3+})$ crystal $\text{Ga}_{2-x}\text{Fe}_x\text{O}_3$ [93-94]. It would seem, however, that the study of simpler model substances with well-defined formulas and magnetic structures would be rewarding. Since the toroidal moments can have both spin and orbital contributions, related magnetic structure determinations should also try to separate these contributions. In this context, the joint use of neutron diffraction, magnetoelectric measurements and magnetic x-ray scattering—the latter permitting separation of spin and orbital moment [100]—appears to be promising.

There exists a contribution to the free enthalpy, $\sim T_i(\mathbf{E} \times \mathbf{H})_i$, in which the *physical meaning of the order parameter* toroidal moment $\mathbf{T}$ has been identified (up to a constant) as the antisymmetric component of the magnetoelectric tensor α_{ik} [101]. In the general case, the three components of the vector $\mathbf{T}$ are proportional to the three components of the antisymmetric part of the magnetoelectric tensor, while the source for the vector $\mathbf{T}$ is the vector $\mathbf{S}$ with components $S_i \sim (\mathbf{E} \times \mathbf{H})_i$, in the same way as the electric and magnetic field vectors are source vectors for polarization and magnetization, respectively. The vector $\mathbf{S}$ transforms like the spontaneous toroidal moment ${}^S\mathbf{T}$ and necessitates off-diagonal components of the magnetoelectric tensor. It is consistent that the 31 Shubnikov point groups permitting a spontaneous toroidal moment do have off-diagonal magnetoelectric coefficients; a few

of them have diagonal ones too (Table 1). It should then be permitted to add the driving force $\sim \Delta {}^S T_i S_i$ for domain switching to those in the list of *primary ferroics* (Table 4). The switching and poling of purely antiferromagnetic ferrotoroidic domains, necessarily magnetoelectric, can be achieved by means of magnetoelectric poling and annealing as described for the first time for the antiferromagnet Cr_2O_3 [72-73], but in the toroidal case this necessitates crossed electric and magnetic fields at right angle to ${}^S \mathbf{T}$.

Usually, however, symmetric and antisymmetric components of the *secondary ferroic* magnetoelectric coefficient α_{ij} (Tables 2 and 4) will occur in materials permitting ${}^S T_i$ and will have to be separated by extracting the anti-symmetric part. This part of the off-diagonal magnetoelectric coefficients has a contribution proportional to ${}^S T_i$. From the asymmetry of these off-diagonal components of the magnetoelectric tensor, evidence for the presence of a spontaneous toroidal moment can be obtained. This has been done for the first time for the polar piezoelectric ferromagnetic $\text{Ga}_{2-x}\text{Fe}_x\text{O}_3$ (point group $m2'm'$) [93-94]. Such evidence for a toroidal moment was also obtained for a high magnetic field-induced spin-flop phase of antiferromagnetic Cr_2O_3 [102-103] with monoclinic point group $2'/m$ [77] and for an antiferromagnetic ferroelectric phase of the perovskite BiFeO_3 with induced magnetic point group $3m$, obtained after destruction of the field-free, magnetically incommensurate antiferromagnetic phase [37] by means of high magnetic fields.

Beside the spontaneous toroidal moments allowed in the 31 ferrotoroidic point groups, toroidal moments can in principle also be induced (reversibly in the thermodynamical sense) by an electric field, by a magnetic field or by stress. This corresponds to the *electrotoroidic*, *magnetotoroidic* and *piezotoroidic* effects, respectively, which have recently been postulated [90]. The first two are analogous to the linear magnetoelectric effect and the third is analogous to the piezoelectric and piezomagnetic effects.

In nuclear and particle physics a toroidal moment—more precisely its spin part—was predicted theoretically by Zel'dovich [104] in 1957 and for which the name “anapole” was proposed by Kompanets [104]. It is remarkable that the first experimental evidence of an anapole was obtained only 50 years later (in 1997) by the measurement of the anapole of the cesium ${}^{133}\text{Cs}$ nucleus [105] and that nearly simultaneously the first experimental evidence for the existence of toroidal moments in crystals ($\text{Ga}_{2-x}\text{Fe}_x\text{O}_3$ [93-94], Cr_2O_3 [102-103], boracites [95]-[99]) was derived from contributions to the linear magnetoelectric effect (for a historical overview see Ref. 90).

6 Conclusions

This chapter has tried to give a short introduction into the field of magnetoelectric effects aimed at the newcomer in the field, putting emphasis on experimental aspects and leaving theory aside. For a review of the theory of microscopic mechanisms responsible for the magnetoelectric effect, involving symmetric exchange, dipolar interactions, antisymmetric (Dzyaloshinsky) exchange, single ion anisotropy and Zeeman energy, the interested reader should consult Ref. 106, as

well as the key Refs. 107-110 both for phenomenological and microscopic aspects. Whereas the symmetry-conditioned properties and phenomenological aspects of magnetoelectrics are rather well understood, there is a great paucity of theoretical studies of the microscopic mechanisms in specific compounds.

It was remarked that the magnetoelectric effects have already gained great importance in solid state physics and in the crystallography of magnetic structures; but so far no devices have been made, as the interaction coefficients are too small. Most measurements of the linear and bilinear magnetoelectric effects have so far been performed in the low-frequency or quasi-static range, but interesting studies in the optical frequency range are emerging and will probably soon acquire the interest of researchers.

The study of spontaneous toroidal moments, which are of magnetoelectric nature, is a new field with steeply rising interest. The demonstrated possibility of their measurement by magnetoelectric techniques, and their calculation on the basis of accurate magnetic and nuclear structural data, jointly represent a precious complementary tool for magnetic structure determination by neutron scattering and magnetic x-ray scattering of antiferromagnetic and (weakly) ferromagnetic magnetoelectric crystals.

In order to overcome the weakness of the intrinsic monophase magnetoelectric coefficients by creating an *artificial* two-phase magnetoelectric effect, bilayers and multilayers of composites, consisting of a piezoelectric and of a magnetostrictive component, are being studied [111-112]. Another technologically very promising approach, to overcome the weak magnetoelectric coefficients of already known crystals, is the recently theoretically discussed possibility of creating nonreciprocal magnetic photonic crystals (i.e., a two-phase system) with a *strong* spectral nonreciprocity, based on periodic arrays with magnetoelectric symmetry and consisting of a component that is magnetically ordered and another component that is dielectrically anisotropic [65].

Acknowledgments

The author is grateful to Prof. Hans Grimmer, Prof. André Authier and the International Union of Crystallography for permitting reproduction of Table 3 prior to publication. Thanks are due to Dr. Jean-Pierre Rivera for critical advice, help in modifying Table 1, permission of reproducing Tables 1 and 2 and citing his data on TbPO_4 prior to publication, and to Gordon and Breach Publishers for permitting reproduction of figures.

References

- [1] P. Curie, Sur la symétrie dans les phénomènes physiques, symétrie d'un champ électrique et d'un champ magnétique, *J de Physique*, 3^e série, **III** (1894), 393–415; reprinted in: *Oeuvres de Pierre Curie*, Gauthier-Villars, Paris, France, 1908, 118–141.

- [2] P. Debye, Bemerkung zu einigen neuen Versuchen über einen magneto-elektrischen Richteffekt, *Z Phys* **35** (1926), 300–301.
- [3] T.H. O'Dell, *The electrodynamics of magneto-electric media*, North-Holland, Amsterdam, The Netherlands, 1970.
- [4] A.H. Compton, The magnetic electron, *J Franklin Inst* **192** (1921), 145–155.
- [5] G.E. Uhlenbeck and S. Goudsmit, Spinning electrons and the structure of spectra, *Nature* **117** (1926), 264–265.
- [6] E.P. Wigner, Über die Operation der Zeitumkehr in der Quantenmechanik, *Gött Nachr Math Phys* (1932), 546–559.
- [7] P. Tolédano, Magnetoelectric symmetry and the Landau theory of phase transitions, *Ferroelectrics* **161** (1994), 257–273.
- [8] L.D. Landau, Zur Theorie der Phasenumwandlungen. II, *Phys Zs Soviet* **11** (1937), 545–555.
- [9] L.D. Landau and E.M. Lifshitz, *Electrodynamics of continuous media*, Addison-Wesley, Reading, MA, USA, 1960.
- [10] I.E. Dzyaloshinskii, On the magneto-electrical effect in antiferromagnetics, *Zh Exp Teor Fiz* **33** (1959), 881–882 [*Sov Phys JETP* **10** (1960), 628–629].
- [11] D.N. Astrov, The magnetoelectric effect in antiferromagnetics, *Zh Exp Teor Fiz* **38** (1960), 984–985 [*Sov Phys JETP* **11** (1960), 708–709].
- [12] D.N. Astrov, Magnetoelectric effect in chromium oxide, *Zh Exp Teor Fiz* **40** (1961), 1035–1041 [*Sov Phys JETP* **13** (1961), 729–733].
- [13] V.J. Folen, G.T. Rado and E.W. Stalder, Anisotropy of the magneto-electric effect in Cr_2O_3 , *Phys Rev Lett* **6** (1961), 607–608.
- [14] E. Ascher, H. Rieder, H. Schmid and H. Stössel, Some properties of ferroelectric-ferromagnetic nickel-iodine boracite $\text{Ni}_3\text{B}_7\text{O}_{13}\text{I}$, *J Appl Phys* **37** (1966), 1404–1405.
- [15] H. Schmid, On a magnetoelectric classifications of materials, *Int J Magn* **4** (1973), 337–361.
- [16] D.B. Litvin, Point group symmetries, *Introduction to complex mediums for optics and electromagnetics* (W.S. Weiglhofer and A. Lakhtakia, eds), SPIE Press, Bellingham, WA, USA, 2003.
- [17] J.-P. Rivera, On definitions, units, measurements, tensor form of the linear magnetoelectric effect and on a new dynamic method applied to Cr-Cl boracite, *Ferroelectrics* **161** (1994), 165–180.
- [18] E. Ascher, Higher-order magneto-electric effects, *Phil Mag* **17** (1968), 149–157.
- [19] H. Grimmer, The forms of tensors describing magnetic, electric, and toroidal properties, *Ferroelectrics* **161** (1994), 181–189.
- [20] A.S. Borovik-Romanov and H. Grimmer, Magnetic properties, *International tables for crystallography, Volume D: Physical properties of crystals* (A. Authier, ed) (in press).
- [21] D.B. Litvin, communication to H. Schmid on 24 October 2002.
- [22] H. Schmid, Magnetoelectric effects in insulating magnetic materials, *Complex mediums* (A. Lakhtakia, W.S. Weiglhofer and R.F. Messier, eds), Proc SPIE **4097**, Bellingham, WA, USA, 2000, 12–24.

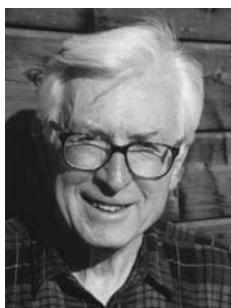
- [23] D. Vaknin, J.L. Zarestky, L.L. Miller, J.-P. Rivera and H. Schmid, Weakly coupled antiferromagnetic planes in single-crystal LiCoPO_4 , *Phys Rev B* **65** (2002), 224414.
- [24] D.E. Cox, Spin ordering in magnetoelectrics, *Int J Magn* **6** (1974), 67–75.
- [25] E. Burzo, Magnetic and related properties of crystalline boron-oxy-halides $\text{M}_3\text{B}_7\text{O}_{13}\text{X}$, *Landolt-Börnstein, New Series* (O. Madelung, ed) Group III, *Solid State Physics*, **27**, Subvol. h, *Boron containing oxides* (H.P.J. Wijn, ed), Springer, Berlin, Germany, 1993, 128–204.
- [26] G.T. Rado, J.M. Ferrari and W.G. Maisch, Magnetoelectric susceptibility of magnetically annealed TbPO_4 , *Phys Rev B* **29** (1984), 4041–4048.
- [27] J.-P. Rivera, communication to H. Schmid, 25 October 2002.
- [28] J.-P. Rivera, The linear magnetoelectric effect in LiCoPO_4 revisited, *Ferroelectrics* **161** (1994), 147–164.
- [29] H. Wiegmann, *Magnetoelectric effects in strong magnetic fields*, Ph.D. thesis 1994, Konstanzer Dissertationen, Hartung-Gorre Verlag, Konstanz, Germany, 1995.
- [30] J.-P. Rivera and H. Schmid, Electrical and optical measurements on nickel iodine boracite, *Ferroelectrics* **36** (1981), 447–450.
- [31] S.L. Hou and N. Bloembergen, Paramagnetoelectric effect in crystals, *Int J Magn* **5** (1974), 327–336.
- [32] J.-P. Rivera and H. Schmid, Linear and quadratic magnetoelectric (ME) effect in Ni-Cl boracite, *J Appl Phys* **70** (1991), 6410–6412.
- [33] Ph. Sciau, M. Clin, J.-P. Rivera and H. Schmid, Magnetoelectric measurements on BaMnF_4 , *Ferroelectrics* **105** (1990), 201–206.
- [34] J.F. Scott and D.R. Tilley, Magneto-electric anomalies in BaMnF_4 , *Ferroelectrics* **161** (1994), 235–243.
- [35] C. Tabares-Muñoz, J.-P. Rivera, A. Bezinges, A. Monnier and H. Schmid, Measurement of the quadratic magnetoelectric effect on single crystalline BiFeO_3 , *Jap J Appl Phys* **24**, Suppl. **24-2** (1985), 1051–1053.
- [36] M. Mercier, Magnetoelectric behaviour in garnets, *Int J Magn* **6** (1974), 77–88.
- [37] Yu.F. Popov, A.K. Zvezdin, G.P. Vorob'ev and A.K. Kadomtseva, Linear magnetoelectric effect and phase transitions in bismuth ferrite, BiFeO_3 , *Pis'ma Zh Eksp Teor Fiz* **57** (1993), 65–68 [*JETP Lett* **57** (1993), 69–73].
- [38] E. Kita, DC magnetoelectric effect measurements by a SQUID magnetometer, *Ferroelectrics* **162** (1994), 397–400.
- [39] T.H. O'Dell and E.A.D. White, Electric field induced Faraday rotation in chromium oxide, *Phil Mag* **22** (1970), 649–653.
- [40] R.V. Pisarev, Crystal optics of magnetoelectrics, *Ferroelectrics* **162** (1994), 191–209.
- [41] J.S. Dodge, A. Kakitunik and M.M. Fejer, Magneto-optic measurements with a Sagnac interferometer, *Ferroelectrics* **162** (1994), 387–395.
- [42] J. Kobayashi and Y. Uesu, A new optical method and apparatus “HAUP” for measuring simultaneously optical activity and birefringence of crystals. I. Principles and construction, *J Appl Cryst* **16** (1983), 204–211.

- [43] J. Kobayashi, Application of optical activity to studies of phase transitions, *Phase Transitions B* **36** (1991), 95–128.
- [44] T. Asahi and J. Kobayashi, Polarimeter for anisotropic optically active materials, *Introduction to complex mediums for optics and electromagnetics* (W.S. Weiglhofer and A. Lakhtakia, eds), SPIE Press, Bellingham, WA, USA, 2003.
- [45] W. Kaminsky and A.M. Glazer, Measurement of optical rotation in crystals, *Ferroelectrics* **183** (1996), 133–141.
- [46] W. Kaminsky, Topographies of chiral and associated optical properties in FeBO_3 using a novel polarimeter, the “tilter,” *Ferroelectrics* **204** (1996), 233–246.
- [47] H. Grimmer, The piezomagnetoelectric effect, *Acta Cryst A* **48** (1992), 266–271.
- [48] G.T. Rado, Statistical theory of magnetoelectric effects, *Phys Rev* **128** (1962), 2546–2556.
- [49] J.-P. Rivera and H. Schmid, Search for the piezomagnetoelectric effect, *Ferroelectrics* **161** (1994), 91–97.
- [50] E. Ascher, Kinetoelectric and kinetomagnetic effects in crystals, *Int J Magn* **5** (1974), 287–295.
- [51] A.A. Gorbatsevich and Yu.V. Kopaev, Toroidal order in crystals, *Ferroelectrics* **161** (1994) 321–334.
- [52] L.A. Shuvalov and N.V. Belov, The symmetry of crystals in which ferromagnetic and ferroelectric properties appear simultaneously, *Kristallografiya* **7** (1962), 192–197 [*Sov Phys Crystallogr* **7** (1962), 150–151].
- [53] E. Ascher, Some properties of spontaneous currents, *Helv Phys Acta* **39** (1966), 40–48.
- [54] K. Aizu, Possible species of ferromagnetic, ferroelectric, and ferroelastic crystals, *Phys Rev B* **2** (1970), 754–772.
- [55] E.K.H. Salje, *Phase transitions in ferroelastic and co-elastic crystals*, Cambridge University Press, Cambridge, UK, 1990.
- [56] R.E. Newnham, Domains in minerals, *Am Mineral* **59** (1974), 906–918.
- [57] R.E. Newnham, *Structure-property relations, Crystal chemistry of non-metallic materials*, Vol. 2 (R. Roy, ed), Springer, Berlin, Germany, 1975.
- [58] R.E. Newnham and L.E. Cross, Secondary ferroics and domain-devided piezoelectrics, *Ferroelectrics* **10** (1976), 269–276.
- [59] H. Schmid, On the possibility of ferromagnetic, antiferromagnetic, ferroelectric, and ferroelastic domain reorientations in magnetic and electric fields, *Ferroelectrics* **221** (1999), 9–17.
- [60] H. Schmid, Sector growth and twinning in nickel boracites grown by transport reactions, *Rost Kristallof* **7** (1967), 32–65 [*Crystal Growth* **7** (1969), 25–52].
- [61] B.K. Ponomarev, S.A. Ivanov, Yu. E. Popov, V.D. Negri and B.S. Red’kin, Magnetoelectric properties of some rare earth molybdates, *Ferroelectrics* **161** (1994), 43–48.

- [62] E. Van Wood and A.E. Austin, Possible applications of magnetoelectric materials, *Int J Magn* **5** (1974), 303–315.
- [63] C.M. Krowne, Nonreciprocity in magnetoelectric crystals, *IEEE Trans Magn* **31** (1995), 2209–2214.
- [64] C.M. Krowne, Comments and corrections to “Nonreciprocity in magnetoelectric crystals,” *IEEE Trans Magn* **31** (1995), 4312–4313.
- [65] F. Figotin and I. Vitebsky, Nonreciprocal magnetic photonic crystals, *Phys Rev E* **63** (2001), 066609.
- [66] R. Fuchs, Wave propagation in a magnetoelectric medium, *Phil Mag* **11** (1965), 647–649.
- [67] D. Felbacq and F. Zolla, Photonic crystals, *Introduction to complex mediums for optics and electromagnetics* (W.S. Weiglhofer and A. Lakhtakia, eds), SPIE Press, Bellingham, WA, USA, 2003.
- [68] J.D. Joannopoulos, R.D. Meade and J.N. Winn, *Photonic crystals—Molding the flow of light*, Princeton University Press, Princeton, NJ, USA, 1995.
- [69] N.F. Kharchenko, The linear magneto-optic effect as a manifestation of a higher order magnetoelectric effect, *Ferroelectrics* **162** (1994), 173–189.
- [70] V.V. Eremenko, N.F. Kharchenko, Yu.G. Litvinenko and V.M. Naumenko, *Magneto-optics and spectroscopy*, Springer, New York, NY, USA, 1992, Ch. 2.
- [71] E. Ascher, Magnetic anisotropy: A reformulation and its consequences, *Helv Phys Acta* **39** (1966), 466–476.
- [72] T.J. Martin, Antiferromagnetic domain switching in Cr_2O_3 , *Phys Lett* **17** (1965), 83–85.
- [73] T.J. Martin and J.C. Anderson, Antiferromagnetic domain switching in Cr_2O_3 , *IEEE Trans Magn* **2** (1966), 446–449.
- [74] R.M. Hornreich, The magnetoelectric effect in polycrystalline powders—annealing at general angles, *Int J Magn* **6** (1974), 47–49.
- [75] H. Wiegmann, A.G. Jansen, P. Wyder, J.-P. Rivera and H. Schmid, Magnetoelectric effect of Cr_2O_3 in strong magnetic fields, *Ferroelectrics* **162** (1994), 141–146.
- [76] M. Fiebig, D. Fröhlich, St. Leute, Th. Lottermoser, V.V. Pavlov and R.V. Pisarev, Determination of magnetic symmetry by optical second-harmonic generation, *J Magn Mater* **226–230** (2001), 961–962.
- [77] R.V. Pisarev, M. Fiebig and D. Fröhlich, Nonlinear optical spectroscopy of magnetoelectric and piezomagnetic crystals, *Ferroelectrics* **204** (1997), 1–21.
- [78] H. Schmid, Multi-ferroic magnetoelectrics, *Ferroelectrics* **162** (1994), 317–338.
- [79] I. Kornev, J.-P. Rivera, S. Gentil, A.G.M. Jansen, M. Bichurin, H. Schmid and P. Wyder, Magnetoelectric properties of LiCoPO_4 . Microscopic theory, *Physica B* **371** (1999), 304–308.
- [80] I. Kornev, J.-P. Rivera, S. Gentil, A.G.M. Jansen, M. Bichurin, H. Schmid and P. Wyder, Magnetoelectric properties of LiCoPO_4 and LiNiPO_4 , *Phys Rev B* **62** (2000), 12247–12253.

- [81] J.-P. Rivera, H. Schmid, J.M. Moret and H. Bill, Measurement of the magnetoelectric effect in Ni-Cl boracite, *Int J Magn* **6** (1974), 211–220.
- [82] J.-P. Rivera and H. Schmid, Evidence for the magnetoelectric effect in nickel-bromine boracite, $\text{Ni}_3\text{B}_7\text{O}_{13}\text{Br}$, *Ferroelectrics* **55** (1984), 295–298.
- [83] M. Senthil Kumar, J.-P. Rivera, Z.G. Ye, S.D. Gentil and H. Schmid, Magnetoelectric effect in Co-Cl boracite, *Ferroelectrics* **204** (1997), 57–71.
- [84] N.F. Kharchenko, Yu.N. Kharchenko, R. Szymczak, M. Baran and H. Schmid, Weak ferromagnetism in the antiferromagnetic magnetoelectric crystal LiCoPO_4 , *Fiz Nizk Temp* **27** (2001), 1208–1213 [*Low Temp Phys* **27** (2001), 895–898].
- [85] N.F. Kharchenko, communication to H. Schmid, 2002.
- [86] M. Lujan, J.P. Rivera, S. Kizhaev, H. Schmid, G. Triscone, J. Muller, Z.-G. Ye, B. Mettout and R. Bozerar, Magnetic measurements and magnetoelectric effect of pyroelectric KNiPO_4 single crystals, *Ferroelectrics* **161** (1994), 77–89.
- [87] K. Siratori, Magneto-electric effect and solid state physics, *Ferroelectrics* **161** (1994), 29–41.
- [88] M.L. Plumer, H. Kawamura and A. Caillé, Symmetry aspects of electric-field induced breaking of chiral degeneracy in the stacked triangular antiferromagnet, *Ferroelectrics* **161** (1994), 221–227.
- [89] V.L. Ginzburg, A.A. Gorbatsevich, Yu.V. Kopayev and B.A. Volkov, On the problem of superdiamagnetism, *Solid State Commun* **50** (1984), 339–343.
- [90] H. Schmid, On ferrotoroidics and elastotoroidic, magnetotoroidic and piezotoroidic effects, *Ferroelectrics* **252** (2001), 41–50.
- [91] V.L. Ginzburg, Phase transitions in ferroelectrics (some historical notes), *Usp Fiz Nauk* **171** (2001), 1091–1097 [*Physics Usp* **44** (2001), 1037–1043].
- [92] V.L. Ginzburg, *Applications of electrodynamics in theoretical physics and astrophysics*, Gordon and Breach, New York, NY, USA, 1989, 151.
- [93] Yu.F. Popov, A.M. Kadomtseva, G.P. Vorob'ev, V.A. Timofeeva, M.M. Tegeranchi and A.K. Zvezdin, Linear magnetostriction and magneto-electric effect in piezoelectric $\text{Ga}_{2-x}\text{Fe}_x\text{O}_3$, *Ferroelectrics* **204** (1997), 269–277.
- [94] Yu.F. Popov, A.M. Kadomtseva, G.P. Vorob'ev, V.A. Timofeeva, D.M. Ustinin, A.K. Zvezdin and M.M. Tegeranchi, Magnetoelectric effect and toroidal ordering in $\text{Ga}_{2-x}\text{Fe}_x\text{O}_3$, *Zh Eksp Teor Fiz* **114** (1998), 263–272 [*JETP* **87** (1998), 146–151].
- [95] D.G. Sannikov and I.S. Zheludev, Possible phase transition generating a spontaneous toroidal moment in nickel boracites, *Fiz Tverd Tela* (Leningrad) **27** (1985), 1369–1372 [*Sov Phys Solid State* **27** (1985) 826–828].
- [96] D.G. Sannikov, Phenomenological theory of the magnetoelectric effect in some boracites, *Zh Eksp Teor Fiz* **111** (1997), 536–546 [*JETP* **84** (1997), 293–299].
- [97] D.G. Sannikov, Ferrotoroidic phase transition in boracites, *Ferroelectrics* **219** (1998), 177–181.

- [98] D.G. Sannikov, A sequence of two ferrotoroidal phase transitions in nickel-bromine boracite $\text{Ni}_3\text{B}_7\text{O}_{13}\text{Br}$, *Pis'ma Zh Eksp Teor Fiz* **73** (2001), 447–450 [*JETP Lett* **73** (2001), 401–403].
- [99] D.G. Sannikov, A phenomenological theory of possible sequences of ferrotoroidal phase transitions in boracites, *Zh Eksp Teor Fiz* **120** (2001), 661–667 [*JETP* **93** (2001), 579–585].
- [100] D. Gibbs, X-ray magnetic scattering, *Synchr Rad News* **14** (5) (2001), 4–10.
- [101] A.A. Gorbatshevich, Yu.V. Kopaev and V.V. Tugushev, Anomalous nonlinear effects at phase transitions to ferroelectric and magnetoelectric states, *Zh Eksp Teor Fiz* **85** (1983), 1107–1121 [*Sov Phys JETP* **58** (1983) 643–651].
- [102] Yu.F. Popov, A.M. Kadomtseva, D.V. Belov, G.P. Vorob'ev and A.K. Zvezdin, Magnetic field-induced toroidal moment in the magnetoelectric Cr_2O_3 , *Pis'ma Zh Eksp Teor Fiz* **69** (1999), 302–306 [*JETP Lett* **69** (1999), 330–335].
- [103] S.S. Krotov, A.M. Kadomtseva, Yu.F. Popov, A.K. Zvezdin, G.P. Vorob'ev and D.V. Belov, Magnetoelectric interactions and induced toroidal ordering in Cr_2O_3 , *J Magn Magn Mater* **226-230** (2001), 963–964.
- [104] Ya. B. Zel'dovich, Electromagnetic interaction with parity violation, *Zh Eksp Teor Fiz* **33** (1957), 1531–1533 [*Sov Phys JETP* **6** (1958), 1184–1186].
- [105] C.S. Wood, S.C. Bennet, D. Cho, B.P. Masterson, J.L. Roberts, C.E. Tanner and C.E. Wieman, Measurement of parity nonconservation and an anapole moment in cesium, *Science* **275** (1997), 1759–1763.
- [106] G.A. Gehring, On the microscopic theory of the magnetoelectric effect, *Ferroelectrics* **161** (1994), 275–285.
- [107] A.J. Freeman and H. Schmid (eds), *Magnetoelectric interaction phenomena in crystals*, Gordon and Breach, London, UK, 1975.
- [108] H. Schmid, A. Janner, H. Grimmer, J.-P. Rivera and Z.-G. Ye (eds), *Proceedings of the 2nd international conference on magnetoelectric interaction phenomena in crystals (MEIPIC-2)*, Ascona, Switzerland, *Ferroelectrics* **161** (1994), 1–348; *Ferroelectrics* **162** (1994), [349–748] 1–400.
- [109] M. Bichurin (ed), *Proceedings of the 3rd international conference on magnetoelectric interaction phenomena in crystals (MEIPIC-3)*, Novgorod, Russia, *Ferroelectrics* **204** (1997), 1–355.
- [110] M. Bichurin (ed), *Proceedings of the 4th international conference on magnetoelectric interaction phenomena in crystals (MEIPIC-4)*, Veliky Novgorod, Russia, *Ferroelectrics* **279** (2002), 1–166; *Ferroelectrics* **280** (2002), [167–384] 1–218.
- [111] G. Srinivasan, E.T. Rasmussen, B.J. Levin and R. Hayes, Magnetoelectric effects in bilayers and multilayers of magnetostrictive and piezoelectric perovskite oxides, *Phys Rev B* **65** (2002), 134402.
- [112] G. Srinivasan, E.T. Rasmussen, B.J. Levin and R. Hayes, Erratum: Magnetoelectric effects in bilayers and multilayers of magnetostrictive and piezoelectric perovskite oxides, *Phys Rev B* **66** (2002), 029901.



Hans Schmid is Professor Emeritus of Solid State Chemistry and Materials Science at the University of Geneva, Switzerland. He received his academic training at the Technical University in Graz, Austria, where he obtained a Dr. techn. degree in physical chemistry in 1955. From 1957 to 1977 he worked as a research scientist, senior scientist and head of the Crystal Physics Section of Battelle Geneva Research Center in the fields of ferroelectrics, ferromagnetics, ferroelastics and crystal optics. In 1964 he synthesized single crystals of halogenoborates (boracites), in which he discovered for the first time, in collaboration with Edgar Ascher, the simultaneous occurrence and mutual coupling of ferroelectricity, ferromagnetism and ferroelasticity in the same phase. In 1977 he was appointed Professor at the Department of Inorganic, Analytical and Applied Chemistry of the University of Geneva. There his research was dedicated to the deepening of understanding of the crystallography, physics and chemistry of magnetically ordered ferroelectrics and ferroelastics with magnetoelectric interactions. He serves on the editorial board of the international journal *Ferroelectrics* as well as on the International Advisory Committee of the ISFD (International Symposia on the Domain Structure of Ferroic and Related Materials).

Magneto-optics: A Critical Review

Allan D. Boardman and Ming Xie

Abstract

A basic review of the critical features of linear magneto-optics is given, as a precursor to an extensive discussion of complex nonlinear magneto-optic waveguides. An elegant global envelope approach is used for an investigation of spatial solitons influenced by nonreciprocal behavior, controlled by the presence of magnetic materials. The kind of complex planar waveguides discussed have great potential in nonlinear optics, because they are destined to support and enhance a major change in the technology of programmable chip-level devices. After addressing the fundamental magneto-optic properties and the kind of materials needed, the envelope theory exploits the Voigt effect, instead of the more obvious Faraday effect. This choice leads to desirable design features. Spatial-soliton light beams are deployed because they can be stably generated within a planar waveguide. The attractive features of the type of magneto-optic waveguides investigated are highlighted with a few numerical simulations that promise useful formats for the fully integrated isolator functionality that modern laser systems demand.

1 Introduction

Magneto-optics describes the propagation of light in a dielectric medium exposed to an applied quasi-static, or dc, magnetic field, $\mathbf{H}_{app} = \mathbf{H}_0$. It is a phenomenon that deeply involves the polarization state of light, which is a measure of its vector property [1]-[5]. The applied magnetic field introduces *forced gyrotropy* into the propagation characteristics [2]. There are many ways to exploit this behavior of light for applications, and among them is the introduction of magneto-optic layers into a planar waveguide structure.

Before this possibility is discussed, however, it is useful to engage in a general discussion of magneto-optic *forced gyrotropy* and the kind of materials that can be used in a few basic applications. To begin with, an electromagnetic wave propagating in a magneto-optic material is associated with a displacement vector [3]

$$\mathbf{D} = \varepsilon_0 [\underline{\underline{\varepsilon}} \cdot \mathbf{E} + i \mathbf{g} \times \mathbf{E}]. \quad (1)$$

Here, ε_0 is the permittivity of free space, and $\underline{\underline{\varepsilon}}$ is the frequency-dependent relative permittivity that exists in the absence of $\mathbf{H}_{app}$, while the *gyration vector* $\mathbf{g} \propto \mathbf{H}_{app}$ is introduced. From the structure of (1), it is clear that $\mathbf{g} \times \mathbf{E}$ is normal to $\mathbf{E}$, so that the overall dielectric property tensor of a magneto-optic material must have off-diagonal elements. Also, $\mathbf{g}$ does not depend upon which way the light wave is traveling. This means that it continues to act even after the wave has been forced to change its direction by a reflection. Hence, the physical property described by (1) is nonreciprocal.

Although *gyrotropy* is the word adopted in Western literature to describe a situation involving nonreciprocity, it is important to note that the word *gyrotropy* has also been adopted in Soviet literature to describe chirality [6]. In fact, there is nothing wrong in this because *gyrotropy* comes from the Greek word *gyros* meaning circle, so *gyrotropy* is a generic description of any rotation of the vibration plane of light.

To be specific, if a plane, linearly polarized, light wave passes once through a transparent block of material that is in a magneto-optic state due to an applied magnetic field along the propagation direction, then its vibration ellipse will be rotated [1-2]. If this wave is reflected back through the material then, as implied by (1) and the previous discussion, the vibration ellipse will continue to rotate in the *same direction*. Faraday discovered this phenomenon that is now called the *Faraday effect*. It describes the propagation of light *parallel* to, or *antiparallel* to, an applied magnetic field. Another magneto-optic effect is called the Voigt effect, or the Cotton-Mouton effect, also named after their co-discoverers who brought it to our attention at the beginning of the last century. This is when Voigt observed magnetic double refraction, in 1902, as white light passed through sodium vapor that was exposed to a strong magnetic field, applied *perpendicular* to the propagation direction. In 1907, Cotton and Mouton also found double refraction, for the same configuration. This time, however, the effect was a thousand times greater, because they passed white light through nitrobenzene. It is helpful to note at this stage

of the discussion that the kind of double refraction given by the Cotton-Mouton magneto-optic effect is the analog of the Kerr electrooptic effect [4]. Furthermore, since either of the names Voigt or Cotton-Mouton is now used without loss of clarity, Voigt is the name of choice in this chapter [1].

For a bulk medium, the signature of the Faraday effect is the rotation of the electric field vector, and this can be resolved into two counter-rotating circularly polarized waves. Each wave is characterized by a refractive index that is proportional to the applied magnetic field. A reversal of the field interchanges the effective refractive indexes and it is this property that makes the Faraday effect *nonreciprocal*. On the other hand, for a bulk isotropic medium the Voigt effect is *reciprocal*, because the birefringence is proportional to the square of the applied magnetic field. At first look, this property seems to be a major setback for the use of the Voigt effect in the design of integrated magneto-optic structures. Fortunately this is not the case, since the Voigt effect can also be forced to be nonreciprocal, but in an *asymmetric waveguide* [7-8]. The details are discussed later in this chapter.

Recent advances in material sciences now make many integrated formats possible in principle, so the real issue is how to make sure that nonreciprocity is not lost simply because of inevitable experimental fabrication tolerances [9]. The solution to this problem involves deep control of the complexity of the waveguide. For example, an application destined for the high-volume laser market has to depend heavily upon waveguide design or some other idea such as postfabrication processing, to overcome weaknesses created by fabrication tolerances [10]. It is possible to achieve this, as shown by the global trends towards integrated magneto-optic devices. They are poised to become pivotal in many applications [9]. The integration of magneto-optic devices with semiconducting substrates embraces active devices, such as lasers, detectors and amplifiers [11]-[13]. In addition, nonlinearity adds a very desirable power parameter, and this direction has been clearly mapped out in terms of using an asymmetric magneto-optic waveguide constructed from third-order optically *nonlinear* material and *linear* magneto-optic layers [14]-[19].

Magneto-optic materials are easily created, but in the absence of an applied magnetic field, they consist of magnetized subvolumes (domains) randomly oriented with respect to each other. The net result is little or no magnetization, so that an applied magnetic field is always used to achieve an overall magnetization [20]. Also, of all the materials available, it is films of the dielectric insulator yttrium iron garnet (YIG) that have attracted a lot of attention over the years [9]. The form of the relative permittivity tensor [17] is built into (1). The fact that $\mathbf{g} \times \mathbf{E}$ is perpendicular to $\mathbf{E}$ means that an external magnetic field creates off-diagonal elements in the relative permittivity tensor, proportional to the induced magnetization. These elements are defined here as $\pm i n_m^2 Q$, where n_m is the normal linear refractive index of the material that appears in the diagonal elements of the permittivity tensor, and Q is a magneto-optic parameter proportional to the applied magnetic field. In general, Q is a spatial distribution that is a relatively complicated function of the magnetization of the material. Another consideration is that Q often vanishes in the absence of an applied magnetic field, leaving no permanent magnetization. The modeling

outlined in this chapter, therefore, assumes that the net magnetization can return to zero upon switching off the applied magnetic field. This may not always be the case, in practice, but it may happen naturally [20]. Added to this consideration is a concern that the desired length of device may be compromised by attenuation. This is not such a problem, however, because integrated-optical circuit, magneto-optic devices are designed to use rather weakly absorbing materials, such as layers of $(\text{LuNiBi})_3(\text{FeAl})_5\text{O}_{12}$ deposited upon $\text{Gd}_3\text{Ga}_5\text{O}_{12}$ substrates [21]. Also, the substitution by cerium of the rare-earth garnet yields large magneto-optic coefficients by way of compensation [9]. It is also accepted that the widespread use of bismuth and aluminium doping reduces attenuation to even more acceptable levels. All of these comments serve to highlight evidence that magneto-optic devices are very promising candidates for switching and routing [21]–[26].

The point of using magneto-optics is to introduce extra degrees of freedom into guided-wave processes, over and above those that appear through crystal symmetry [27]. Indeed, exploiting magneto-optic properties will produce very impressive integrated units [11]–[13] when compared to those based purely upon GaAs or LiNbO_3 technologies. To achieve this goal, the key issues are the optimized selection of magneto-optic materials and the design of robust magneto-optic waveguides. After this is achieved, there is then the option of adding nonlinear effects for even greater flexibility.

As stated earlier, low-loss propagation can now be routinely expected and the classic material to use is an epitaxially grown YIG film. Looking ahead, however, to the type of applications to be expected, it is desirable to explore the feasibility of using YIG semiconductor structures, so that amplification can be added during propagation. Experimentally, it is advantageous that YIG is optically transparent over the 1–5 μm range of wavelengths and also that its magnetization is saturable by small, easily generated magnetic fields. The downside is that YIG cannot be deposited epitaxially [28] onto a material like GaAs or its derivatives, which is exactly the kind of material that could be selected to provide amplification through carrier drift mechanisms. Even so, buffer layers or diluted magnetic semiconductors [16] can be used instead, without loss of performance, and it has been pointed out recently that YIG can actually be “attached” to GaAs [29].

The really important applications for magneto-optic devices include pivotal activity in laser control [13] through isolators; and a lot of effort has gone into production and design [10] compatible with low-loss nonreciprocal circuits. Both isolators and circulators are required as part of optical communication systems, with circulators particularly able to facilitate routing capabilities and two-way communication. In this context, a lot of transverse-electric–transverse magnetic (TE-TM) conversion work has appeared in the linear domain, but phase matching remains a difficult problem to overcome. Also, manufacturing tolerances can reduce device performance to unacceptable levels [10]. Another comment about TE-TM conversion is that it relies too much upon the deployment of the traditional Faraday effect; so a different way forward is to use nonreciprocal modes in the kind of asymmetric planar waveguide discussed earlier [10], [21]. In such an arrangement, the magnetic field is not only perpendicular to the propagation direction but is also parallel

to the surface of a planar plane. An analysis of its dispersion properties reveals that the guided linear TE modes are not affected, to first order, by an applied magnetic field [14]–[19]. On the other hand, TM guided modes have amplitudes that possess a finite derivative with respect to the propagation direction, proportional to the applied magnetic field. It is this conclusion that has led to a preference for the Voigt effect over the Faraday effect, because there is no need then to address phase-matching considerations, as any coherence exponential factors involving the birefringence are entirely absent. The *asymmetric* guide, therefore, is the “designer’s choice” for applications like isolators.

All conventional work on isolators and circulators has, up to now, ignored the question of optical nonlinearity. This is rather surprising at first sight, since the magneto-optic community is currently intensely interested in nonlinear magneto-optics. The explanation lies in the focus of the magnetism community upon the measurement of the Kerr rotation of the vibration ellipse of light *reflected* from ferromagnetic surfaces. Their attention is directed to second-harmonic generation because it was discovered recently [30] that a giant nonlinear Kerr rotation can be observed in the second-harmonic wave created by reflection off low-dimensional structures (thin films), using a beam from a Ti:sapphire laser. This second-harmonic generation is due to routine symmetry-breaking at the surface. The exciting part is that the nonlinear Kerr rotation is an order of magnitude larger than anything that can be achieved in the linear domain. This enhancement of the Kerr rotation is aided by magnetically induced elements of the nonlinear susceptibility tensor [30]. The activity surrounding the Kerr rotation effects is in sharp contrast, therefore, to the guided wave work to be discussed in this chapter.

2 Linear magneto-optics of bulk material

2.1 Fundamentals of planewave behavior

Magneto-optics is a form of forced gyrotropy and its signature is the appearance of off-diagonal elements in the relative permittivity tensor. To establish this material property in a very general way is beyond the limits of this chapter, because the physical arguments will differ depending upon whether an insulator or conductor is being discussed. Nevertheless, the generic form of the relative permittivity tensor can be inferred from the behavior of a free electron, with charge q and velocity $\mathbf{v}$, in an applied magnetic field $\mathbf{H}_{app} = \mathbf{H}_0$. It experiences a force $q\mathbf{v} \times \mathbf{H}_{app}$, which leads to the conclusion that in a ferromagnetic metal the relative permittivity tensor has off-diagonal elements. The generic conclusion is that *all* materials exposed to a magnetic field have a relative permittivity tensor with off-diagonal elements proportional to the magnitude of the applied magnetic field. The magnitude of the proportionality constant determines the availability of the magneto-optic effect for applications. The generic form is adopted and expressed through the magneto-optic parameter Q .

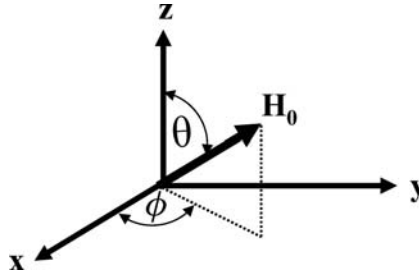


Figure 1 Orientation of the external magnetic field $\mathbf{H}_{app} = \mathbf{H}_0$.

The coordinate system used is displayed in Figure 1, which shows the dc magnetic field applied in an arbitrary direction, giving [14]-[19]

$$\underline{\underline{\varepsilon}} = n_m^2 \begin{pmatrix} 1 & -iQ \cos \theta & iQ \sin \theta \sin \phi \\ iQ \cos \theta & 1 & -iQ \sin \theta \cos \phi \\ -iQ \sin \theta \sin \phi & iQ \sin \theta \cos \phi & 1 \end{pmatrix}. \quad (2)$$

For the magneto-optic insulators to be discussed here, the relative permeability is unity, so that all the electromagnetic properties are governed by the relative permittivity. Some popular forms of $\underline{\underline{\varepsilon}}$ are as follows:

$$\textbf{Polar: } \left\{ \theta = \frac{\pi}{2}, \phi = \frac{\pi}{2} \right\}, \underline{\underline{\varepsilon}} = \begin{pmatrix} n_m^2 & 0 & -iQn_m^2 \\ 0 & n_m^2 & 0 \\ iQn_m^2 & 0 & n_m^2 \end{pmatrix}. \quad (3)$$

This is useful when the magnetization is perpendicular to the boundaries of the material and propagation of electromagnetic waves parallel to the boundaries is anticipated.

$$\textbf{Longitudinal: } \{ \theta = 0, \phi = 0 \}, \underline{\underline{\varepsilon}} = \begin{pmatrix} n_m^2 & -iQn_m^2 & 0 \\ iQn_m^2 & n_m^2 & 0 \\ 0 & 0 & n_m^2 \end{pmatrix}. \quad (4)$$

In this case, propagation takes place along an applied magnetic field that is also parallel to the boundaries.

$$\textbf{Transverse: } \left\{ \theta = \frac{\pi}{2}, \phi = 0 \right\}, \underline{\underline{\varepsilon}} = \begin{pmatrix} n_m^2 & 0 & 0 \\ 0 & n_m^2 & -iQn_m^2 \\ 0 & iQn_m^2 & n_m^2 \end{pmatrix}. \quad (5)$$

This form is appropriate when electromagnetic wave propagation is not only parallel to the boundaries but is also perpendicular to a magnetization parallel to the boundaries.

Note that in all of these tensors it is a good approximation, based upon our knowledge of real materials, to set all the diagonal elements equal to each other. Strictly speaking, they are not all exactly equal, but not taking the practical step of setting them equal to each other creates unnecessarily complicated results.

The solution of the Maxwell equations needs the behavior of matter when exposed to influences like an external magnetic field to be fed in through the constitutive equations that relate the electric field $\mathbf{E}$ and magnetic flux density $\mathbf{B}$ to the displacement vector $\mathbf{D}$ and magnetic field $\mathbf{H}$. For magneto-optic materials these are now established as

$$\mathbf{D} = \varepsilon_0(\underline{\underline{\varepsilon}} \cdot \mathbf{E}), \mathbf{B} = \mu_0 \mathbf{H}, \quad (6)$$

where μ_0 is the permeability of free space.

Consider now a plane wave propagating along the z -direction, defined as

$$\mathbf{E}(x, y, z) = \mathbf{E}_0 \exp(i\omega t - i\beta z), \mathbf{H}(x, y, z) = \mathbf{h}_0 \exp(i\omega t - i\beta z), \quad (7)$$

where ω is angular frequency, c_0 is the velocity of light in free space, β is a wavenumber, and $\mathbf{E}_0$ and $\mathbf{h}_0$ are amplitude vectors. For a bulk medium with no boundaries, a straightforward application of the Maxwell equations shows that although the plane wave has a transverse magnetic field, the same may not be true for the electric field. A manipulation of the component equations leads to the following characteristic matrix equation:

$$\begin{pmatrix} 1 - \frac{c_0^2 \beta^2}{\omega^2 n_m^2} & -iQ \cos \theta & iQ \sin \theta \sin \phi \\ iQ \cos \theta & 1 - \frac{c_0^2 \beta^2}{\omega^2 n_m^2} & -iQ \sin \theta \cos \phi \\ -iQ \sin \theta \sin \phi & iQ \sin \theta \cos \phi & 1 \end{pmatrix} \begin{pmatrix} E_{0x} \\ E_{0y} \\ E_{0z} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}. \quad (8)$$

The expansion of the determinant of the matrix in (8) gives a dispersion equation in the form

$$\frac{c_0^2 \beta^2}{\omega^2 n_m^2} = 1 - \frac{Q^2 \sin^2 \theta \pm \sqrt{Q^4 \sin^4 \theta + 4Q^2 \cos^2 \theta}}{2}. \quad (9)$$

There are two mathematical solutions for β and, physically, these represent waves with polarizations that do not change during propagation. The wavenumber β does not depend upon ϕ because the bulk medium has rotational symmetry about the z axis. It does not therefore matter how $\mathbf{H}_0$ is applied, just as long as it is in the xz plane, for which $\phi = 0$. It is very convenient that setting $\phi = 0$ does not lose any information, because it enables two famous geometries to be drawn out without much complication. Respectively, these geometries are called the Faraday configuration ($\theta = 0$) and the Voigt (or Cotton-Mouton) configuration ($\theta = \pi/2$).

2.2 The Faraday configuration

The adoption of the Faraday configuration means the investigation of wave propagation parallel to the applied dc magnetic field $\mathbf{H}_0$. The dispersion equation (9) then yields

$$\frac{c_0^2}{\omega^2} \frac{\beta_{\pm}^2}{n_m^2} = 1 \pm Q, \quad E_{0z} = 0; \quad (10)$$

and, furthermore,

$$\frac{c_0^2}{\omega^2} \frac{\beta_+^2}{n_m^2} = 1 + Q, \quad E_{0y} = i E_{0x}; \quad (11)$$

$$\frac{c_0^2}{\omega^2} \frac{\beta_-^2}{n_m^2} = 1 - Q, \quad E_{0y} = -i E_{0x}. \quad (12)$$

The ratio of the transverse field components is $\pm i$, so that the plane wave is circularly polarized. Let us define

$$\mathbf{E}_+ = \frac{|E_0|}{\sqrt{2}} (\hat{\mathbf{x}} + i\hat{\mathbf{y}}) \exp(i\omega t - i\beta_+ z), \quad (13)$$

$$\mathbf{E}_- = \frac{|E_0|}{\sqrt{2}} (\hat{\mathbf{x}} - i\hat{\mathbf{y}}) \exp(i\omega t - i\beta_- z), \quad (14)$$

where $|E_0| = \sqrt{E_{0x}^2 + E_{0y}^2}$. It should also be emphasized that the waves are transverse, because *both* E_{0z} and h_{0z} are zero.

It was realized very early on that the electric field of a linearly polarized wave is actually the sum of two circularly polarized waves; i.e.,

$$\mathbf{E} = A_+ \mathbf{E}_+ + A_- \mathbf{E}_-, \quad (15)$$

in which A_+ and A_- are complex amplitudes. The physical consequences of this decomposition can best be appreciated through a simple yet powerful example. Suppose that a plane wave, which is linearly polarized along the x axis, is launched at $z = 0$ with amplitude $|E_0|$; i.e.,

$$\mathbf{E}(z = 0) = |E_0| \hat{\mathbf{x}} \exp(i\omega t) = A_+ \mathbf{E}_+(z = 0) + A_- \mathbf{E}_-(z = 0). \quad (16)$$

The component equations are therefore

$$\hat{\mathbf{x}}: \quad A_+ \frac{|E_0|}{\sqrt{2}} + A_- \frac{|E_0|}{\sqrt{2}} = |E_0|, \quad (17)$$

$$\hat{\mathbf{y}}: \quad i A_+ \frac{|E_0|}{\sqrt{2}} - i A_- \frac{|E_0|}{\sqrt{2}} = 0. \quad (18)$$

Equations (17) and (18) show that $A_+ = A_- = 1/\sqrt{2}$. Upon propagation to $z > 0$ the electric field evolves to

$$\begin{aligned}
 \mathbf{E}(z > 0) &= \frac{|E_0|}{2} (\hat{\mathbf{x}} + i\hat{\mathbf{y}}) \exp(i\omega t - i\beta_+ z) + \frac{|E_0|}{2} (\hat{\mathbf{x}} - i\hat{\mathbf{y}}) \exp(i\omega t - i\beta_- z) \\
 &= \frac{|E_0|}{2} \exp(i\omega t - i\bar{\beta}z) \left[(\hat{\mathbf{x}} + i\hat{\mathbf{y}}) \exp\left(-i\frac{\beta_+ - \beta_-}{2}z\right) \right. \\
 &\quad \left. + (\hat{\mathbf{x}} - i\hat{\mathbf{y}}) \exp\left(-i\frac{\beta_- - \beta_+}{2}z\right) \right] \\
 &= |E_0| \exp(i\omega t - i\bar{\beta}z) \left[\hat{\mathbf{x}} \cos\left(\frac{\beta_+ - \beta_-}{2}z\right) + \hat{\mathbf{y}} \sin\left(\frac{\beta_+ - \beta_-}{2}z\right) \right],
 \end{aligned} \tag{19}$$

where $\bar{\beta} = (\beta_+ + \beta_-)/2$. This is still a linearly polarized wave, but now the electric field vector has *rotated* through an angle $\theta_F = ((\beta_+ - \beta_-)/2)z$. This is precisely what is meant by *Faraday rotation*. The Faraday rotation per unit length is $d\theta_F/dz = (\beta_+ - \beta_-)/2 \approx (180^\circ/\lambda_0)(nQ)$, where λ_0 is the free-space wavelength.

A commonly used [9] magneto-optic insulating material is $(\text{GdBiCa})_3\text{Fe}_5\text{O}_{12}$ that, when magnetized to saturation, has a magneto-optic parameter $Q \approx 7.5 \times 10^{-4}$. A typical application will be in the infrared regime, with $\lambda_0 = 1.152 \mu\text{m}$ and a magneto-optic material with refractive index $n_m \approx 2.4$, so that $d\theta_F/dz \approx 2800^\circ/\text{cm}$. This means that the required length to generate a 180° rotation is $L_F \approx 0.64 \text{ mm}$. Loss during propagation can be included easily in the analysis by assuming that the refractive index n_m is complex; e.g., $n_m = n_r - in_i$. A decay length can then be defined as the propagation distance $L_d = \lambda_0/(2\pi n_i)$, over which the amplitude of the wave is reduced to $1/e$ of its original value. For a film made from $(\text{GdBiCa})_3\text{Fe}_5\text{O}_{12}$, $n_i \approx 0.19 \times 10^{-4}$, so that $L_d \approx 9.65 \text{ mm}$ at $\lambda_0 = 1.152 \mu\text{m}$. It is very good news that $L_d \gg L_F$ because it permits such typical magneto-optic films to be treated as if they are optically transparent.

2.3 The Voigt configuration

In this configuration, the plane wave propagates *perpendicularly* to a magnetic field applied to a dielectric material. With reference to Figure 1, if $\mathbf{H}_0$ is applied parallel to the x axis, then $\theta = \pi/2$ and the dispersion equation (9) has the following solutions:

$\mathbf{E}_{\parallel}$ wave:

$$\left(\frac{c_0^2}{\omega^2}\right) \left(\frac{\beta_{\parallel}^2}{n_m^2}\right) = 1, \quad [E_{0y} = 0, E_{0z} = 0], \quad \mathbf{E}_{\parallel} = E_{0x} \hat{\mathbf{x}} \exp(i\omega t - i\beta_{\parallel} z), \tag{20}$$

$\mathbf{E}_\perp$ wave:

$$\left(\frac{c_0^2}{\omega^2}\right)\left(\frac{\beta_\perp^2}{n_m^2}\right) = 1 - Q^2, \quad \left[E_{0x} = 0, \frac{E_{0z}}{E_{0y}} = -iQ\right],$$

$$\mathbf{E}_\perp = E_{0y}(\hat{\mathbf{y}} - iQ\hat{\mathbf{z}})\exp(i\omega t - i\beta_\perp z). \quad (21)$$

Whereas $\mathbf{E}_\parallel$ is a transverse wave linearly polarized parallel to $\mathbf{H}_0$, $\mathbf{E}_\perp$ is elliptically polarized in a plane perpendicular to $\mathbf{H}_0$. In all applications of magneto-optic properties of materials, it is important to use the Q -dependence of the dispersion. This desire focuses attention upon $\mathbf{E}_\perp$ but the dispersion of $\mathbf{E}_\perp$ wave depends upon Q^2 ; and a typical value of Q could be as small as $\approx 10^{-4}$ so that the $\mathbf{E}_\perp$ wave is almost linearly polarized. Q can be made larger through doping [9] of materials like YIG, but it is still of interest to look at the magnetically induced birefringence $\beta_\perp - \beta_\parallel$. This is associated with a beat length L_b , defined as the propagation distance over which $\mathbf{E}_\perp$ experiences a 2π phase shift, with respect to $\mathbf{E}_\parallel$; i.e.,

$$L_b = 2\pi \frac{1}{|\beta_\perp - \beta_\parallel|} \approx \frac{2\lambda_0}{n_m Q^2}. \quad (22)$$

The beat length $L_b \approx 1.71$ m for $(\text{GdBiCa})_3\text{Fe}_5\text{O}_{12}$, which means that the magnetically induced birefringence is really much smaller than the Faraday effect in a typically magneto-optic medium.

If it is not possible to get away from a Voigt effect proportional to Q^2 , then it would not be of much interest in waveguiding applications: its effect would be reciprocal because reversing Q does not change the sign of Q^2 . Fortunately, the Voigt configuration is rather interesting in an asymmetric waveguide since order Q effects then appear, making Voigt competitive with Faraday for nonreciprocal applications.

The foregoing paragraphs describe some rather important planewave magneto-optic effects. The conclusion about the Voigt effect is that $\beta_\parallel$ is independent of Q , while $\beta_\perp$ is proportional to Q^2 . Voigt waves in the bulk will display reciprocity, therefore, since reversal of the magnetization does not change the dispersion. If nonreciprocity is required, then an *asymmetric waveguide structure* must be used. This fact can be simply exposed using a perturbative wave envelope theory, as shown in Section 3. It can of course be arrived at using an exact waveguide analysis. It takes much longer to arrive at the nonreciprocal behavior that way, however, and adds nothing to the issue addressed in this chapter. An example of the kind of asymmetric guide discussed here consists of a magneto-optic substrate supporting a nonmagnetic nonlinear film bounded by an air cladding. For such a guide $O(Q)$ effects are present and the propagation of TM waves becomes nonreciprocal. This property is pivotal to many potential applications. On the other hand, the Faraday configuration turns out to be undesirable for applications such as isolators, because of phase-matching problems [7]-[9], [21]-[26].

3 Envelopes in a waveguide

A paradigmatic asymmetric planar waveguide is sketched in Figure 2. It is designed to guide an optical beam along the z axis. The beam direction is perpendicular to a magnetization, created by a dc magnetic field applied along the x axis. In detail, the waveguide has an upper air boundary and consists of an insulating dielectric layer of thickness d attached to a thick magneto-optic substrate. The layer becomes optically nonlinear, and at sufficiently high optical beam powers, a spatial soliton is formed. Broadly speaking, this occurs when diffraction along the x direction is completely compensated by the self-focusing created by the nonlinearity. This argument assumes that the nonlinearity will be typical of the Kerr-like power-induced refractive index increase found in many materials. Naturally there are many kinds of nonlinearity, and the details given here will have to be changed to accommodate them. The principles, however, will remain the same.

Any nonlinear balancing process is along the x axis only. It is a one-dimensional process; therefore, any solitons produced this way will be stable [14]-[19]. In essence, there is a freezing of diffraction along the y axis by the waveguiding process.

Any analysis of planar guides inevitably focuses upon modes that are classified as TM (transverse magnetic) and TE (transverse electric) guided waves. The generic form of a TM guided wave is delineated by

$$\mathbf{E} = T(x, z) [\xi_y(y)\hat{\mathbf{y}} + \xi_z(y)\hat{\mathbf{z}}] \exp\left(-i\frac{\omega}{c_0}\alpha z\right) \exp(i\omega t). \quad (23)$$

The shape vector $\xi(y)$ describes the modal field in the waveguide. The $T(x, z)$ envelope is a slowly varying amplitude and the fast spatial variation is $\exp(-i(\omega/c_0)\alpha z)$, where $\beta = (\omega/c_0)\alpha$.

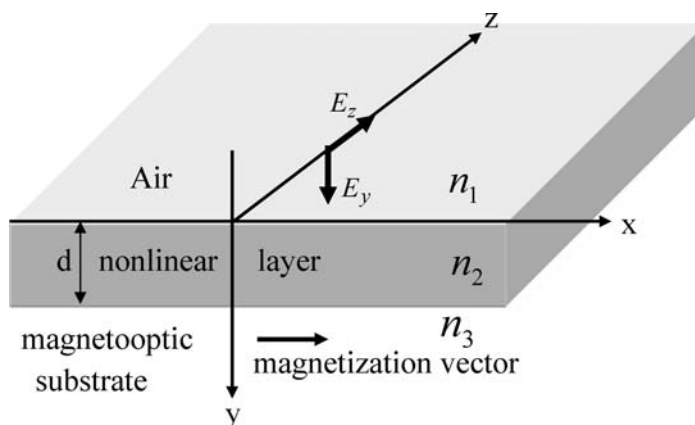


Figure 2 Sketch of a planar waveguide structure, unbounded in the $\pm x$ directions.

A general wave guided in a planar structure is a combination of TE and TM modes and carries an electric field $\mathbf{E}_1 + \mathbf{E}_2$, where

$$\text{TE: } \mathbf{E}_1 = T_1(z) \xi_x \hat{\mathbf{x}} \exp\left(-\frac{\omega}{c_0} \alpha_1 z\right) \exp(i\omega t), \quad (24)$$

$$\text{TM: } \mathbf{E}_2 = T_2(z) \left[\xi_y \hat{\mathbf{y}} \exp\left(-i \frac{\omega}{c_0} \alpha_2 z\right) + \xi_z \hat{\mathbf{z}} \exp\left(-i \frac{\omega}{c_0} \alpha_2 z\right) \right] \exp(i\omega t). \quad (25)$$

Here, α_1 and α_2 are the appropriate effective refractive indexes. Free diffraction of beams is allowed in the (x, z) plane, which will be deployed to create spatial solitons as the beam power rises. Before that is investigated, however, it is necessary to determine the individual contributions to the way in which the slowly varying amplitudes evolve from spatial modulation due to (i) diffraction, (ii) magneto-optics and (iii) nonlinearity. In other words, the evolution equations for the envelope functions T_1 and T_2 must be derived. The various contributions can be found by treating them independently and then adding them together as linear perturbations.

The investigation opens with the wave equation for a linear, isotropic, nonconducting waveguide:

$$\nabla(\nabla \cdot \mathbf{E}) - \nabla^2 \mathbf{E} = \left(\frac{\omega}{c_0}\right)^2 \varepsilon_{is} \mathbf{E}. \quad (26)$$

In this case, $\nabla \cdot \mathbf{E} = 0$, ε_{is} is the relative permeability, and the polarizations due to any magneto-optic and nonlinear influence are set to zero. Given the forms (24) and (25), we obtain

$$\nabla^2 [\mathbf{E}_1 + \mathbf{E}_2] + \left(\frac{\omega}{c_0}\right)^2 \varepsilon_{is} [\mathbf{E}_1 + \mathbf{E}_2] = 0. \quad (27)$$

The next steps are to multiply (27) by the complex conjugate $\mathbf{E}_1^*$ and then integrate over $y \in (-\infty, \infty)$. During this process, the term $\partial^2 T_1 / \partial z^2$ is neglected because T_1 varies very slowly with z . The result is that

$$\begin{aligned} & -\left(\frac{\omega^2}{c_0^2}\right) |T_1|^2 \int \varepsilon_{is} |\xi_x|^2 dy - 2i \frac{\omega}{c_0} \alpha_1 T_1^* \frac{\partial T_1}{\partial z} \int |\xi_x|^2 dy \\ & + \left(\frac{\omega^2}{c_0^2}\right) |T_1|^2 \int \varepsilon_{is} |\xi_x|^2 dy = 0. \end{aligned} \quad (28)$$

The succinct conclusion is that

$$\frac{\partial T_1}{\partial z} = 0. \quad (29)$$

This answer is not surprising because there is no reason to believe that T_1 can do anything other than remain constant, within the adopted set of assumptions. The same conclusion is reached for T_2 , so the final result within a slowly varying approximation is that

$$\frac{\partial T_1}{\partial z} = 0, \quad \frac{\partial T_2}{\partial z} = 0, \quad (30)$$

in the purely *linear, non-magneto-optic* domain.

In practice, a beam of light will be introduced into the waveguide that will be trapped in the y direction because of the guiding process; but it will be free to diffract along the x direction. This is modeled by the inclusion of diffraction as a spatial modulation perturbation with respect to x . The expected diffraction in the (x, z) plane is accounted for then by setting $T_1 = T_1(x, z)$ and $T_2 = T_2(x, z)$. This step leads to

$$\begin{aligned} & -\left(\frac{\omega^2}{c_0^2}\right) |T_1|^2 \int \varepsilon_{is} |\xi_x|^2 dy - 2i \frac{\omega}{c_0} \alpha_1 T_1^* \frac{\partial T_1}{\partial z} \int |\xi_x|^2 dy \\ & + \left(\frac{\omega^2}{c_0^2}\right) |T_1|^2 \int \varepsilon_{is} |\xi_x|^2 dy + T_1^* \frac{\partial^2 T_1}{\partial x^2} \int |\xi_x|^2 dy = 0, \end{aligned} \quad (31)$$

and the overall conclusions that

$$2i \frac{\omega}{c_0} \alpha_1 \frac{\partial T_1}{\partial z} = \frac{\partial^2 T_1}{\partial x^2}, \quad 2i \frac{\omega}{c_0} \alpha_2 \frac{\partial T_2}{\partial z} = \frac{\partial^2 T_2}{\partial x^2}; \quad (32)$$

$$\frac{d}{dz} \int_{-\infty}^{\infty} |T_1|^2 dx = 0, \quad \frac{d}{dz} \int_{-\infty}^{\infty} |T_2|^2 dx = 0. \quad (33)$$

These equations very elegantly show how to account for diffraction of light beams. They also show that, at this stage, the total power is conserved and that TE and TM modes are *not coupled* to each other by diffraction. Note that the latter is expressed through the second derivatives with respect to x .

The magneto-optic behavior emerges from the envelope equation after its influence upon the evolution of T_1 and T_2 is extracted from an analysis of the wave equation

$$\nabla(\nabla \cdot \mathbf{E}) - \nabla^2 \mathbf{E} = \frac{\omega^2}{c_0^2} \varepsilon_{is} \mathbf{E} + \frac{\omega^2}{c_0^2} \mathbf{P}_m = \frac{\omega^2}{c_0^2} \underline{\underline{\varepsilon}} \cdot \mathbf{E}, \quad (34)$$

where $\mathbf{P}_m$ is the material polarization contribution arising from its magneto-optic property. It is assumed that all the off-diagonal elements that represent the magneto-optic influence are, in principle, nonzero at this stage of the argument. Also, in practice, the magneto-optic effect can be introduced in any part of

the waveguide because an integration is going to be undertaken over the whole waveguide structure.

Since there is now forced gyrotropy in the system, $\nabla \cdot \mathbf{E} \neq 0$, even though $\nabla \cdot \mathbf{D} = 0$. The algebra is more complicated now but the basic procedure is still the same. The final result for the variation of T_1 is

$$2i \frac{\omega}{c_0} \frac{\partial T_1}{\partial z} = i \frac{\omega^2}{c_0^2} T_2 \left[\frac{\int (\varepsilon_{xy} \xi_y \xi_x^* + \varepsilon_{xz} \xi_z \xi_x^*) dy}{\alpha_1 \int |\xi_x|^2 dy} \right] \exp \left(-i \frac{\omega}{c_0} \gamma z \right), \quad (35)$$

where $\gamma = \alpha_2 - \alpha_1$ is a refractive index mismatch, and ε_{xy} and ε_{yz} are possible off-diagonal elements of the permittivity tensor, due to the magneto-optic properties. Clearly, the integrations over the waveguide structure introduce some mean values, which are

$$\bar{\varepsilon}_{xy} = \frac{\int \varepsilon_{xy} \xi_y \xi_x^* dy}{\alpha_1 \int |\xi_x|^2 dy}, \quad \bar{\varepsilon}_{xz} = -\frac{\int \varepsilon_{xz} (i \xi_z) \xi_x^* dy}{\alpha_1 \int |\xi_x|^2 dy}, \quad (36)$$

where $i \xi_z$ is actually a real-valued quantity because ξ_z is imaginary.

For the TM amplitude, we get similarly

$$2i \frac{\omega}{c_0} \frac{\partial T_2}{\partial z} = -i \frac{\omega^2}{c_0^2} T_1 (\bar{\varepsilon}_{xy} - i \bar{\varepsilon}_{xz}) \exp \left(i \frac{\omega}{c_0} \gamma z \right) + 2 \frac{\omega^2}{c_0^2} \bar{\varepsilon}_{yz} T_2, \quad (37)$$

in which the new averaged quantity is

$$\bar{\varepsilon}_{yz} = \frac{c_0}{\omega} \frac{\int \varepsilon_{yz} \left(\xi_y \frac{\partial \xi_y}{\partial y} \right) dy}{\beta_2^2 \int (|\xi_y|^2 + |\xi_z|^2) dy}. \quad (38)$$

The magnitude and style of the magneto-optic impact in a waveguide structure depends not only upon how many elements of the permittivity tensor will be nonzero but also how they are positioned within that tensor. All of this is driven by the magnetic field orientation with respect to the propagation direction, and the number of surfaces. In this chapter, only $\bar{\varepsilon}_{yz}$ remains in the envelope equation, and any problems due to phase matching are avoided. This is the so-called transverse case that exploits TM modes and therefore the Voigt effect. It has become the configuration of choice for isolator designs in current times.

The type of magneto-optic investigation to be presented here begins with the *linear* magneto-optic conclusions

$$\frac{\partial T_1}{\partial z} = 0, \quad i \frac{\partial T_2}{\partial z} = \frac{\omega}{c_0} \bar{\varepsilon}_{yz} T_2. \quad (39)$$

Note that $\partial T_2 / \partial z = 0$ if the waveguide is symmetric, as can be seen from an inspection of $\bar{\varepsilon}_{yz}$ [14]–[19]. Also, for this *transverse orientation* TE waves are uncoupled from the TM waves. This is a very useful and natural simplification. $\bar{\varepsilon}_{yz}$ is

evaluated over the whole waveguide structure and its sign reverses when the propagation direction is reversed (by a mirror reflection, for example), so the guiding is now truly *nonreciprocal*. In principle, the integrations in the definition of $\bar{\epsilon}_{yz}$ need to be examined carefully to make sure that $\bar{\epsilon}_{yz}$ is not identically zero for any reason.

Within the framework of this type of theory, it is also quite straightforward to include dissipation. One of the problems with undoped magneto-optic materials is their potential for dissipation. If this is not addressed, it could be a fatal obstacle to the deployment of magneto-optic materials in useful applications. One way to proceed is to use rare-earth garnets with cerium or bismuth substitution. The damping of guided waves is then decreased quite dramatically and the useable length of device becomes attractive.

The use of a transverse magnetic field orientation removes undue dependence upon design tolerances. This means that stable magneto-optic waveguide designs can be produced that are hardly influenced at all by the inevitable slight variations in design dimensions from one device to another. On this basis, our proposal is achievable and propagation lengths as small as 5 mm are realistic, with attenuation down to 0.2 and even $<0.1 \text{ cm}^{-1}$. It is also safely assumed that any epitaxial optical birefringence, which is stress and/or growth induced, will have only a small influence upon the nonreciprocal propagation to be exploited here. For the model investigated here, the magneto-optic material and the nonlinear material are in some kind of contact, which does not necessarily have to be epitaxial.

The only way TE and TM waves can couple to each other in this transverse magneto-optic configuration is through optical nonlinearity. For isotropic materials, with unsaturated third-order nonlinearity characterized by the fourth-rank electric susceptibility tensor components χ_{xxyy} and χ_{xyyx} , the x -component of the nonlinear polarization is

$$P_x^{NL} = \frac{3}{4} \left[2\chi_{xxyy} \left(|E_x|^2 + |E_y|^2 + |E_z|^2 \right) E_x + \chi_{xyyx} \left(E_x^2 + E_y^2 + E_z^2 \right) E_x^* \right]. \quad (40)$$

Within the spirit of additive perturbative influences, the effect of the nonlinear polarization can be considered first in isolation, to evaluate its effect on the evolution of T_1 and T_2 . Omitting all other influences, we see that the third-order nonlinearity [31] causes the evolution equations to become [32]

$$2i \frac{\omega}{c_0} \frac{\partial T_1}{\partial z} = \frac{\omega^2}{c_0^2} \bar{\chi}_e T_1 |T_1|^2 + \frac{\omega^2}{c_0^2} \bar{\chi}_c T_1 |T_2|^2 + \frac{\omega^2}{c_0^2} \bar{\chi}_F T_1^* T_2^2 e^{-2i \frac{\omega}{c_0} \gamma z}, \quad (41)$$

$$2i \frac{\omega}{c_0} \frac{\partial T_2}{\partial z} = \frac{\omega^2}{c_0^2} \bar{\chi}_m T_2 |T_2|^2 + \frac{\omega^2}{c_0^2} \bar{\chi}_c T_2 |T_1|^2 + \frac{\omega^2}{c_0^2} \bar{\chi}_F T_2^* T_1^2 e^{2i \frac{\omega}{c_0} \gamma z}. \quad (42)$$

In these equations, $\bar{\chi}_e$ and $\bar{\chi}_m$ are the *self-phase modulation* coefficients for the TE and TM waves, respectively; $\bar{\chi}_c$ is the *cross-phase modulation* coefficient between

the TE and TM wave; and $\bar{\chi}_F$ is the *four-wave mixing* coefficient, quantifying energy exchange between the polarizations. The definitions of these coefficients are

$$\bar{\chi}_m = \frac{3}{4} \frac{\int \chi_{xxxx} (\xi_y^4 + \xi_z^4 + 2(1 - \delta)\xi_y^2\xi_z^2 + 2\delta\xi_y^2\xi_z^2) dy}{\alpha_2 \int \xi_x^2 dy} \approx \bar{\chi}_e, \quad (43)$$

$$\frac{\bar{\chi}_c}{\bar{\chi}_e} \approx 1 - \delta, \quad \frac{\bar{\chi}_F}{\bar{\chi}_e} \approx \delta \quad \delta = \frac{\chi_{xyyy}}{\chi_{xxxx}}, \quad (44)$$

where ξ_x , ξ_y , and ξ_z now satisfy the condition $|\xi_x|^2 \simeq |\xi_y|^2 + |\xi_z|^2$.

The width of an optical beam is defined with reference to its cross-sectional shape. For a Gaussian beam, it is quite useful to define its “size” as its width at a radius where the intensity is $1/e$ of the peak value. The straight diameter is an obvious measure for a uniform beam. Whatever the shape of the beam, it is useful to introduce a quantity D_0 as a “measure” of its “size.” It is then useful to define something loosely called a “diffraction” length, $L_D = 2\alpha_1 D_0^2(\omega/c_0)$; it is roughly the distance over which diffraction operates to try and double the width of the beam. Given these ideas, then the following useful transformations and parameters can be introduced:

$$x = D_0 x', \quad z = L_D z', \quad (45a)$$

$$T_1 = \frac{c_0}{\omega D_0} \frac{\sqrt{2}}{\sqrt{\alpha_1 \bar{\chi}_e}} \psi_1, \quad T_2 = \frac{c_0}{\omega D_0} \frac{\sqrt{2}}{\sqrt{\alpha_1 \bar{\chi}_e}} \psi_2, \quad (45b)$$

$$\nu = 2\alpha_1 \frac{\omega^2}{c_0^2} D_0^2 \bar{\epsilon}_{yz}, \quad \gamma' = 2\alpha_1 \frac{\omega^2}{c_0^2} D_0^2 \gamma. \quad (45c)$$

The final compact form of the global envelope equations is

$$i \frac{\partial \psi_1}{\partial z} = \frac{\partial^2 \psi_1}{\partial x^2} + 2 \left(|\psi_1|^2 + |\psi_2|^2 \right) \psi_1 + 2\delta \psi_1^* \psi_2^2 e^{-2i\gamma z} - 2\delta \psi_1 |\psi_2|^2, \quad (46a)$$

$$i \frac{\partial \psi_2}{\partial z} = \frac{\partial^2 \psi_2}{\partial x^2} + \nu \psi_2 + 2 \left(|\psi_1|^2 + |\psi_2|^2 \right) \psi_2 - 2\delta \psi_2 |\psi_1|^2 + 2\delta \psi_2^* \psi_1 e^{2i\gamma z}, \quad (46b)$$

where all dashes have been dropped for notational tidiness. These equations retain a four-wave mixing term [33]: a difficulty brought in through the exponential factor. In a practical sense, however, the terms are not much of a problem because $\delta = 1/3$ for electronic nonlinearity, $\delta = 3/4$ for molecular orientation nonlinearity, and $\delta = 0$ for thermal, or electrostrictive, nonlinearity. All of these are commonly encountered forms of optical nonlinearity. No loss of generality will occur,

therefore, if we set $(1 - \delta) \simeq 1$ and assume that the four-wave mixing terms average to zero over the propagation distances that are anticipated in magneto-optic applications. The quantity $\gamma = \alpha_2 - \alpha_1$ is called an index mismatch parameter and typically has the value $\gamma \sim 0.2$ at a wavelength of $1.55 \mu\text{m}$. For these data the beat length is $L_B \approx (\pi/0.2)(\lambda_0/2\pi) \sim 3.9 \mu\text{m}$ and the diffraction length is $L_D \approx 2 \text{ mm}$. These estimates confirm that exponential factor does indeed change sign many times during the process of spatial solitary wave formation. This is why it averages to almost zero.

A practical form of the global equations [19] is, therefore, as follows:

$$i \frac{\partial \psi_1}{\partial z} = \frac{\partial^2 \psi_1}{\partial x^2} + 2 \left(|\psi_1|^2 + |\psi_2|^2 \right) \psi_1, \quad (47a)$$

$$i \frac{\partial \psi_2}{\partial z} = \frac{\partial^2 \psi_2}{\partial x^2} + 2 \left(|\psi_1|^2 + |\psi_2|^2 \right)^2 \psi_2 + \nu \psi_2. \quad (47b)$$

Except for the introduction of the term $\nu \psi_2$, (47a) and (47b) are now in the familiar vector solitary wave form. They have vector solitary wave solutions, with two components, corresponding to the TE and TM polarizations. The appearance of $\nu \psi_2$ is potentially interesting, but not if ν is simply a constant and the medium is homogeneous. In that case, the transformation $\psi_2 = u e^{i\nu z}$ removes all trace of the magneto-optic influence, bringing the equations precisely back to the standard vector form. To make the magneto-optic influence felt, and to make progress towards a useful application, a nonhomogeneous medium must be considered. An important example is a waveguide with a “step” built into it. This will enable the simulation of the behavior of vector magneto-optic solitons, or solitary waves, in more practical scenarios. This kind of step structure is similar to ridge structures deployed in linear magneto-optics.

4 Complex planar waveguide

A sufficiently complex planar waveguide structure [19] is sketched in Figure 3. It can be thought of as a perturbed system, for which the unperturbed relative permeability is ε_{is} . It is constructed from a nonlinear guiding layer with a linear refractive index n_2 , which is bounded by a cladding with refractive index n_3 and by a substrate that has a refractive index n_1 ; i.e.,

$$\varepsilon_{is} = \begin{cases} n_1^2, & y < 0 \\ n_2^2, & 0 < y < d \\ n_3^2, & y > d \end{cases} \quad (48)$$

A perturbation of ε_{is} to ε'_{is} through a refractive index change, quite independently of any magneto-optic or nonlinear influences, changes the evolution equa-

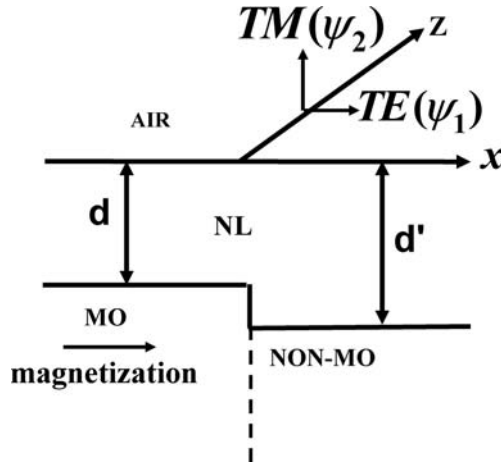


Figure 3 Discontinuous waveguide structure. MO = magneto-optic, NON-MO = nonmagnetic, NL = nonlinear.

tions for the TE and TM modes to

$$\text{TE: } 2i \frac{c_0}{\omega} \frac{\partial T_1}{\partial z} = 2 \left[\frac{\int (\varepsilon'_{is} - \varepsilon_{is}) |\xi_x|^2 dy}{2\alpha_1 \int |\xi_x|^2 dy} \right] T_1 = 2\bar{\varepsilon}_e T_1, \quad (49)$$

$$\text{TM: } 2i \frac{c_0}{\omega} \frac{\partial T_2}{\partial z} = 2 \left[\frac{\int (\varepsilon'_{is} - \varepsilon_{is}) (|\xi_y|^2 + |\xi_z|^2) dy}{\alpha_2 \int \left(\frac{\varepsilon_{is}}{\alpha_2^2} |\xi_y|^2 + |\xi_y|^2 + |\xi_z|^2 \right) dy} \right] T_2 = 2\bar{\varepsilon}_m T_2, \quad (50)$$

where ξ is the unperturbed modal field vector. Equations (44) and (45) show very clearly how waveguide spatial inhomogeneity can be built into the overall description of the propagation of nonlinear light beams along the z axis of a magneto-optic planar system. The modified envelope equations are as follows:

$$i \frac{\partial \psi_1}{\partial z} = \frac{\partial^2 \psi_1}{\partial x^2} + w_1 \psi_1 + 2 \left[|\psi_1|^2 + (1 - \delta) |\psi_2|^2 \right] \psi_1, \quad (51a)$$

$$i \frac{\partial \psi_2}{\partial z} = \frac{\partial^2 \psi_2}{\partial x^2} + (w_2 + \nu) \psi_2 + 2 \left[(1 - \delta) |\psi_1|^2 + |\psi_2|^2 \right] \psi_2. \quad (51b)$$

In these equations, δ is retained and the presence of the waveguide discontinuity is felt through what can be called the disturbance parameters

$$w_1 = 2\alpha_1 \frac{\omega^2}{c_0^2} D_0^2 \bar{\varepsilon}_e, \quad w_2 = 2\alpha_1 \frac{\omega^2}{c_0^2} D_0^2 \bar{\varepsilon}_m.$$

The waveguide has both a step and an interface; and the structure can be permitted some parameter variance in the confidence that variations will not detract from

its performance. This is because design criterions for linear nonreciprocal coupled waveguides show that it is possible to anticipate etch depths of $\sim 0.05 \mu\text{m}$, ridge widths of $\sim 1.5 \mu\text{m}$ and coupler lengths $\sim 850 \mu\text{m}$. All these possibilities can be readily controlled in the telecom and datacom wavelength range of $\sim 1.3\text{--}1.55 \mu\text{m}$ for garnet materials. Furthermore, even if a complex waveguide has postproduction tolerance problems, it can be tuned towards optimization, through approximate specification followed by post-production processing. If an index-step is introduced at $x = x_0$ and a change in the magnetization distribution is created, then the general spatial distributions shown in Figure 3 are as follows:

$$\nu = \nu(x) = \begin{cases} \nu_1, & x < x_0 \\ \nu_2, & x > x_0 \end{cases}, \quad (52a)$$

$$w_1 = w_1(x) = \begin{cases} w_{11}, & x < x_0 \\ w_{12}, & x > x_0 \end{cases}, \quad (52b)$$

$$w_2 = w_2(x) = \begin{cases} w_{21}, & x < x_0 \\ w_{22}, & x > x_0 \end{cases}. \quad (52c)$$

Figure 3 shows that there is a step in the magnetization from a magneto-optic substrate to a nonmagnetic one, which could be gadolinium gallium garnet, otherwise known as GGG. If this boundary is at $x = 0$, then $\nu_2 = 0$ for $x > 0$, and $\nu_1 \neq 0$ in the other half-space. The refractive indexes of the magneto-optic and non-magneto-optic material will be different, but this can be controlled by the effective index method described later in this section. The magneto-optic effect and the refractive index discontinuity enter the envelope equation in exactly the same way. The reasons for using the type of refractive index distribution, shown in Figure 3, however, are to ensure that $\nu = \nu(x)$ and to incorporate a possible application in which edges naturally appear. The influence of the magneto-optic parameter ν amounts to just a simple phase change, if it is only a constant.

It is important to balance the competition between $\nu(x)$ and $w_2(x)$. Although they both appear to have similar quantitative roles, the magneto-optic term has much more physical influence because it introduces nonreciprocity. Specifically, evaluating (49) for a TE mode yields the effective relative permittivity

$$\bar{\epsilon}_e = \frac{\int_{-d}^d (n_{NL}^2 - n_m^2) |\xi_x|^2 dy + \int_d^\infty (n_{GGG}^2 - n_m^2) |\xi_x|^2 dy}{2\alpha_1 \int_{-\infty}^\infty |\xi_x|^2 dy}. \quad (53)$$

Let us select some specific values for ϵ_{is} in (48). An upper air boundary is selected for which $n_3 = 1$. The nonlinear guiding layer has a low-power refractive index $n_2 = n_{NL}$. The substrate has zero magnetic field; hence, $n_2 = n_m$ to the left of the edge, and $n_2 = n_{GGG}$ to the right of the edge. An inspection of Figure 3 reveals the choice of integration ranges. A similar process leads to the effective

relative permittivity

$$\bar{\epsilon}_m = \frac{\int_{d'}^{d'} (n_{NL}^2 - n_m^2) (|\xi_y|^2 + |\xi_z|^2) dy + \int_{d'}^{\infty} (n_{GGG}^2 - n_m^2) (|\xi_y|^2 + |\xi_z|^2) dy}{\alpha_2 \int_{-\infty}^{\infty} \left(\frac{\epsilon_{is}}{\alpha_2^2} |\xi_y|^2 + |\xi_y|^2 + |\xi_z|^2 \right) dy}, \quad (54)$$

for a TM mode. Also,

$$\nu_1 = 2\alpha_2 \frac{\omega^2}{c_0^2} D_0^2 \left(i \frac{n_3^2 Q}{\alpha_1} \right) \frac{\int_{d'}^{\infty} \xi_y \xi_z dy}{\int_{-\infty}^{\infty} |\xi_x|^2 dy} = \left(2\alpha_2 \frac{\omega^2}{c_0^2} D_0^2 \right) \left(\frac{i n_m^2}{\alpha_1} \right) Q_{yz}, \quad (55)$$

where Q_{yz} is defined by (55). The quantities $\bar{\epsilon}_e$, $\bar{\epsilon}_m$ and Q_{yz} all vary with d' , the measure of the step size. There exists a value of d' for which $\bar{\epsilon}_e = \bar{\epsilon}_m = 0$ and Q_{yz} is a maximum. This competition between step size d' and the optimization of the magneto-optic effective parameter Q_{yz} is a critical feature of the system. The simulations shown briefly in Section 5 will assume that d' is set equal to the really critical value when $\bar{\epsilon}_e = \bar{\epsilon}_m = 0$. This means that although there is actually a physical junction, as expected in a real application, the whole junction behavior is magneto-optic and hence nonreciprocal.

5 Vector solitons

The coupled equations (51a) and (51b) generate values for ψ_1 and ψ_2 as the components of a *vector soliton* or a *multisoliton*. The numerical evaluation of the coupled Schrödinger equations can be performed with or without the factor $(1 - \delta)$ included in the computations. Nevertheless, $(1 - \delta)$ is usually 1 or 2/3 for any application that can be reasonably anticipated, so $\delta = 1$ is the value selected here. It is called the Manakov case and is especially interesting, mathematically speaking, because the coupled equations are integrable. The latter term means that the equations have exact soliton solutions and that there is an infinite number of conserved quantities, such as momentum and energy, associated with them. As might be expected, the numerical results presented here do not change in any significant way for the non-Manakov cases, but *solitary wave* is then a better term than *soliton*. It is common practice, however, to exchange these terms quite freely, and care must be exercised in order to take into account genuine solitary-wave properties. The latter include the possibility of being destroyed or fused together during a collision. Another useful feature of solitons is that they can be viewed as particles either moving in a potential well or approaching a potential barrier.

To appreciate the full impact of the foregoing interpretation, it is recommended that a Lagrangian analysis be performed with the aid of trial functions parameterized to take into account what is being looked for. A good trial function will permit amplitude variation with propagation distance z and will introduce a z -variation into the phase, width, initial velocity, angle of incidence to an interface, and so on.

The analysis will produce a potential function that will be used to explain what happens when a soliton (solitary wave) approaches an interface, such as the one created in Figure 3. Such a Lagrangian analysis is not set out here, however, on the grounds that it is a very lengthy and quite complicated theory for the candidate vector solitons. Nevertheless, the effective particle behavior language is still used, with terms such as barriers and potential wells.

The vector soliton is composed of two particles representing the TE and TM contents. They are bound together in the kind of potential well just referred to and are attractive to each other when their centers overlap each other. Here the potential well/barrier is at its lowest value, but its overall topology reflects the ‘step structure’ of the guide shown in Figure 3. The attraction between the TE and TM component solitons is explicable in terms of a binding energy, so there is a tendency for the TE and TM solitons to propagate together, which is where the terminology *vector soliton* comes from. Naturally, it is only truly a vector soliton if, in particle language, it resides completely within a potential well. Under certain circumstances, therefore, the composite particle representing the vector soliton is so firmly confined that it is possible for it to be bounced back by an interface, with its TE and TM components tightly bound together. In that case, the vector soliton retains its integrity and does not split on encountering the kind of barrier step shown in Figure 3. The exact simulation shown in Figure 4 confirms these explanations and shows, quite clearly, reflection without any breaking up. The figure also shows what happens when the direction of an applied magnetic field is reversed. Transmission is now permitted, which demonstrates the nonreciprocal property introduced by the magneto-optics.

If the kinetic energy of the particle is not large enough for the particle to overcome the potential step created by the waveguide discontinuity, it may not return to its previous path but reflect back along a new route. If that happens, the TE and TM components will not overlap anymore. The vector soliton is then split by the interface into separate TE and TM solitons. For example, it is possible for the TE part to be transmitted through the interface, while its TM part is reflected. This is confirmed numerically in Figure 5. The result can be understood if it is realized that the interface is induced mainly by the magneto-optic effect, so it interacts just with the TM component of the soliton. If the TE and TM components are not tightly bound together, the TE part can be transmitted but the TM part will be bounced back. Reversal of the magnetization lowers the effective height of the barrier step created at the interface; hence, full transmission of the original vector soliton occurs, as shown in Figure 5. A polarization-switching filter can be made to exploit this effect.

6 Concluding remarks

This chapter provides a review of magneto-optic linear guided waves with enough detail to convey basic principles, as well as a discussion of vector spatial solitons in a nonhomogeneous magneto-optic waveguide. The basic idea is to include magneto-optic layers, claddings or substrates into a fairly complicated pla-

nar waveguide format. This is attractive for integrated optics because the aim is to produce an integrated format that will participate in, and control, all-optical processing operations in the near future. The use of nonlinear materials in contact with magneto-optic materials is an attractive means to move forward, because it allows us to select optimum optically nonlinear substances and then attach them to the magneto-optic waveguide components. This is more desirable than trying to optimize the nonlinear *and* the magneto-optic properties in the same material. This method also avoids the debilitating absorption properties that tend to surround high- Q materials. The development relies upon the *transverse* magneto-optic configuration. This case is important because of its desirable nonreciprocal properties.

A suitable combination of linear magneto-optics and intrinsic optical nonlinearity can lead to elegant coupled, global wave equations that describe the behavior of composite (TE-TM) vector solitons. The dynamics of these solitons was described in a particle language and then demonstrated through some numerical simulations. It is emphasized that spatial solitons exploit diffraction, which is always rather strong, in the sense that it operates over distances that are only on the order of 1 mm in the propagation direction and the order of $1\ \mu\text{m}$ in the transverse direction. The deployment of spatial solitons, therefore, fits the desired “chip” format very well.

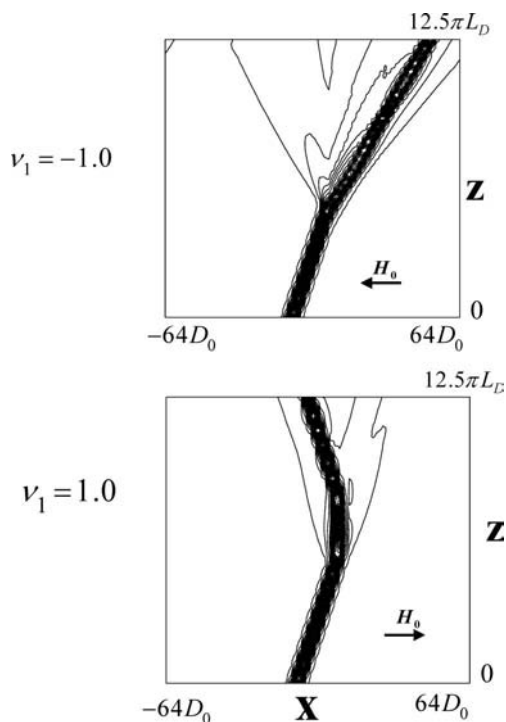


Figure 4 Nonreciprocal behavior of a vector soliton sitting in a potential well created by an interface for $\nu_1 = \pm 1.0$, $\nu_2 = 0.0$, $w_{11} = 0.0$, $w_{12} = 0.3$, $w_{21} = 0.0$, and $w_{22} = 0.4$. Typical parameter values are $H_0 = 23.87\ \text{Oe}$, $L_D = 2\alpha_2 D_0^2(w/c_0) \approx 0.585\ \text{mm}$, $\alpha_2 = 2.8862$, $\lambda_0 = 1.55\ \mu\text{m}$, and $D_0 = 5\ \mu\text{m}$.

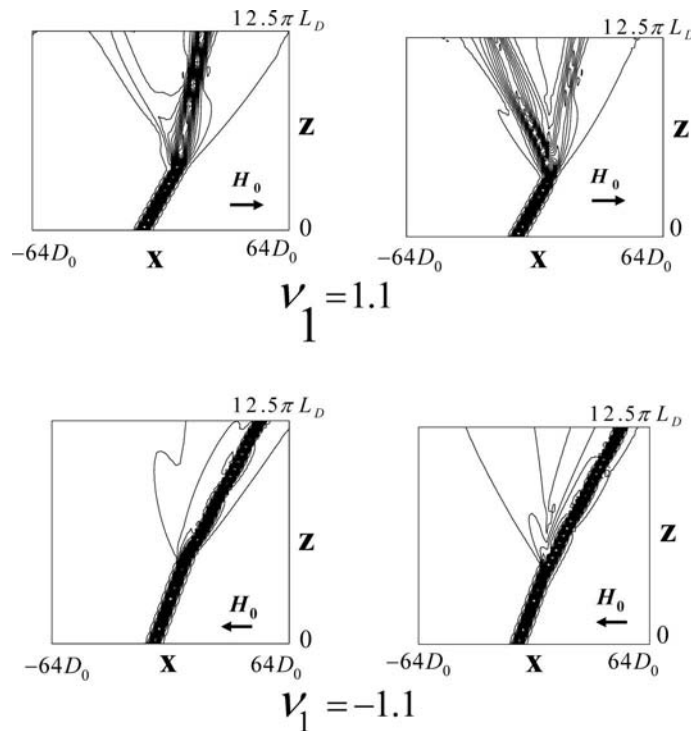


Figure 5 Simulations of vector soliton splitting. (Top) forward transmission with dc magnetic field H_0 directed along $x > 0$, $v_1 = 1.1$, $v_2 = 0.0$; (Bottom) the result of magnetic field reversal: $v_1 = -1.1$. Data: Typical values are $H_0 = 23.87$ Oe, $L_D = 2\alpha_2 D_0^2 (w/c_0) \approx 0.585$ mm, $\alpha_2 = 2.8862$, $\lambda_0 = 1.55$ μm , and $D_0 = 5$ μm .

In addition, it is perfectly feasible to scale everything down to the nanotechnology level. Control of soliton dynamics through magneto-optics is therefore a powerful idea.

Acknowledgments

The authors acknowledge support from the United Kingdom Engineering and Physics Sciences Research Council Nonlinear Magneto-optics Network.

References

- [1] E. Hecht and A. Zajac, *Optics*, Addison-Wesley, Reading, MA, USA, 1974.
- [2] J. Petykiewicz, *Wave optics*, Kluwer, Dordrecht, The Netherlands, 1992.
- [3] B.E. Saleh and M.C. Teich, *Fundamentals of photonics*, Wiley, New York, NY, USA, 1991.
- [4] F.A. Jenkins and H.E. White, *Fundamentals of optics*, McGraw-Hill, New York, NY, USA, 1976.

- [5] F. Jonsson and C. Flytzanis, Polarization state controlled multistability of a nonlinear magneto-optic cavity, *Phys Rev Lett* **82** (1997), 1426–1429.
- [6] I.V. Lindell, A.H. Sihvola, S.A. Tretyakov and A.J. Viitanen, *Electromagnetic waves in chiral and bi-isotropic media*, Artech House, Boston, MA, USA, 1994.
- [7] J.A. Mizumoto and Y. Naito, Nonreciprocal propagation characteristics of YIG thin film, *IEEE Trans Microw Theory Tech* **30** (1982), 922–925.
- [8] Y. Satomura, M. Matsuhara and N. Kumagai, Analysis of electromagnetic wave modes in anisotropic slab waveguides, *IEEE Trans Microw Theory Tech* **22** (1974), 86–92.
- [9] A.K. Zvezdin and V.A. Kotov, *Modern magneto-optics and modern magneto-optic materials*, Institute of Physics Publishing, Bristol, UK, 1997.
- [10] N. Bahlmann, M. Lohmeyer, O. Zhuromskyy, H. Dötsch and P. Hertel, Non-reciprocal coupled waveguides for integrated optical isolators and circulators for TM modes, *Opt Commun* **161** (1999), 330–337.
- [11] W. Zaets and K. Ando, Optical waveguide isolator based on nonreciprocal loss/gain of amplifier covered by ferromagnetic layer, *IEEE Photon Technol Lett* **11** (1999), 1012–1014.
- [12] J.M. Hammer, J.H. Abeles and D.J. Channin, Polycrystalline-metal-ferromagnetic optical waveguide isolator (POWI) for monolithic-integration with diode-laser devices, *IEEE Photon Technol Lett* **9** (1997), 631–633.
- [13] W. Zaets and K. Ando, Magnetically programmable bistable laser diode with a ferromagnetic layer, *IEEE Photon Technol Lett* **13** (2001), 185–187.
- [14] A.D. Boardman and M. Xie, Spatial solitons in discontinuous magneto-optic waveguides, *J Opt B: Quantum Semiclass Opt* **3** (2001), S244–S250.
- [15] A.D. Boardman and K. Xie, Vector spatial solitons influenced by magneto-optic effects in cascaded nonlinear media, *Phys Rev E* **55** (1997), 1–11.
- [16] A.D. Boardman, K. Xie and M. Xie, Applied magneto-optic soliton dynamics: TM and TE-TM-driven dynamics, *Acta Polonica* **99** (2001), 7–16.
- [17] A.D. Boardman and A.P. Sukhorukov (eds), *Soliton-driven photonics*, Kluwer, Boston, MA, USA, 2001.
- [18] S. Trillo and W. Torruellas (eds), *Spatial solitons*, Springer, Berlin, Germany, 2001.
- [19] A.D. Boardman and M. Xie, Vector spatial solitons in complex magneto-optic waveguides, *J Opt Soc Am B* **15** (2002), 3102–3111.
- [20] M.S. Sodha and N.C. Srivastava, *Microwave propagation in ferrimagnetics*, Plenum, New York, NY, USA, 1981.
- [21] T. Mizumoto, H. Chihara, N. Tokui and Y. Naito, Verification of waveguide-type of optical circulator operation, *Electron Lett* **26** (1990), 199–200.
- [22] H. Dötsch, P. Hertel, B. Lührmann, S. Sure, H.P. Winkler and M. Ye, Application of magnetic garnet films in integrated optics, *IEEE Trans Magn* **28** (1992), 2979–2984.
- [23] O. Zhuromskyy, M. Lohmeyer, N. Bahlmann, H. Dötsch, P. Hertel and A.F. Popkov, Analysis of polarization independent Mach–Zehnder-type integrated optical isolator, *J Lightwave Technol* **17** (1997), 1200–1205.

- [24] T. Jekijima, T. Fujii and K. Wakino, Optical Faraday rotator using Ce-substituted fibrous YIG single crystal grown by floating-zone method with YAG laser heating, *IEEE Trans Microw Theory Tech* **47** (1999), 2294–2297.
- [25] T. Shintaku and T. Uno, Optical waveguide isolator based upon nonreciprocal radiation, *J Appl Phys* **75** (1994), 8155–8159.
- [26] T. Mizumoto, S. Mashimo, T. Ida and Y. Naito, In-plane magnetised rare earth iron garnet for a waveguide optical isolator employing a nonreciprocal phase shift, *IEEE Trans Magn* **29** (1993), 3417–3419.
- [27] D.B. Litvin, Point group symmetries, *Introduction to complex mediums for electromagnetics and optics* (W.S. Weiglhofer and A. Lakhtakia, eds), SPIE Press, Bellingham, WA, USA, 2003.
- [28] S. Sugano and N. Kojima (eds), *Magneto-optics*, Springer, Berlin, Germany, 2000.
- [29] M. Levy, R.M. Osgood and A. Kumar, Epitaxial liftoff of thin oxide layers: yttrium iron garnets onto GaAs, *Appl Phys Lett* **71** (1997), 2617–2622.
- [30] T. Andersen and W. Hübner, Substrate effects in the magneto-optical second-harmonic generation from first principles: Fe/Cu (001), *Phys Rev B* **65** (2002), 174409.
- [31] W.A. Schroeder, D.S. McCallum, D.R. Harken, M.D. Dvorak, D.R. Anderson and A.L. Smirl, Intrinsic and induced anisotropy of nonlinear absorption and refraction in zinc blende semiconductors, *J Opt Soc Am B* **12** (1995), 401–414.
- [32] A.D. Boardman, K. Xie and A.A. Zharov, Polarization interaction of spatial solitons in optical planar waveguides, *Phys Rev A* **51** (1995), 692–705.
- [33] G.P. Agrawal, *Nonlinear fiber optics*, Academic Press, New York, NY, USA, 1995.



Allan D. Boardman is well known both in the UK and globally for his work on surface plasmons and guided wave optics, especially nonlinear waves and solitons. He is a leading theorist looking at nonlinear guided magneto-optic waves. He is Professor of Applied Physics and holds a Doctor of Science degree from the University of Durham, UK, founded by Oliver Cromwell. He has organized conferences and been a director of a number of NATO Advanced Study Institutes. He is vice-president of the United Kingdom Consortium for Optics and Photonics (UKCPO). Within the UK Institute of Physics he is the

chair of the Quantum Electronics and Optics Group and honorary secretary of the Computational Group. At the European Physical Society level, he is a board member of the Quantum Electronics and Photonics Division. His principal research activities cover linear and nonlinear surface polaritons on negative index materials, nonlinear microwave magnetics, complex microwave materials, magneto-phonic band gap phenomena, solitons in nonlinear magneto-optics, nonlinear

electromagnetic modeling, dissipative solitons, and electromagnetic modeling for millimeter-wave devices and electromagnetic compatibility problems. He is the author of more than 223 papers in major journals, the presenter of 215 conference papers, the author/editor of 7 books, a member of the editorial advisory board of the journal *Nonlinear Optics*, and editor of *Advances in Nonlinear Optics*.



Ming Xie was educated in China at Sichuan Normal University where she obtained very high grades in her first degree. Her current interests are the theory and simulation of magneto-optic phenomena, especially nonlinear guided waves. She has recently completed her Ph.D. at the University of Salford under the direction of Professor A.D. Boardman. She is the co-author of 7 papers and a book chapter. She is now the post-doctoral research assistant on a new UK joint project funded by the UK Engineering and Science Research Council. This project is a collaboration with Glasgow and Cambridge Universities

to design integrated magneto-optic components for applications in industrial applications. She is an expert on waveguide modeling.

Static and Dynamic Magnetoelasticity

Graeme Dewar

Abstract

Magnetoelasticity encompasses a wide range of phenomena, including volume and Joule magnetostriction, the Villari effect, the direct and inverse Wiedemann effects, the Matteucci effect, the ΔE effect, a magnetoelastic contribution to the apparent magnetic anisotropy, and others. These effects are conveniently codified in a magnetoelastic energy density which, together with the magnetic (including exchange) and elastic energy densities, provides a consistent thermodynamic description of magnetoelasticity. Each of these effects and the corresponding terms of the energy density are reviewed here. This energy density is described by a collection of material constants that, in principle, are derivable from theory. The physical coordinates that are kept constant in any experiment dictate the relevant combination of these material constants that is ultimately observed. Static and dynamic measurements are generally carried out with different constraints, and, not surprisingly, these experiments measure different combinations of material parameters. For example, the highly magnetostrictive smart material Terfenol-D ($\text{Dy}_{0.73}\text{Tb}_{0.27}\text{Fe}_{1.95}$) has a static magnetic anisotropy that is markedly different from the anisotropy exhibited in a dynamic measurement.

1 Introduction

The existence of magnetoelasticity, which describes the interaction between a ferromagnet's elastic state and its magnetization, has been known for more than 150 years. The purpose of this chapter is to assess the current description of magnetoelasticity, especially as it pertains to characterizing materials. The early descriptions of magnetostriction captured the phenomenons but were inadequate for characterizing materials in anything but the lowest order of the interaction. The work of H.B. Callen and E.R. Callen in the 1960s remedied this with the introduction of a new notation based on group theory.

One magnetoelastic effect is the strain associated with the direction of a magnetostrictive material's magnetization. If one magnetizes an originally unmagnetized object made of such a material, one observes a change in length of the object both parallel and perpendicular to the magnetization. This is an example of the Joule effect. Other examples of magnetoelasticity are described later. In this chapter, any interaction between the magnetization and the strains, forces or torques on an object is considered a magnetoelastic interaction. The conventional magnetoelastic energy density terms developed here describe most, but not all, observed magnetoelastic effects.

Experimenters have often quoted measurements of the various parameters of magnetoelasticity without regard to whether the conditions of the measurement were isothermal or adiabatic, whether they related to constant stress or constant strain, etc. The rationale for this was that magnetoelastic effects were small and differences between different types of measurements were smaller still. The highly magnetostrictive material Terfenol-D ($\text{Dy}_{0.73}\text{Tb}_{0.27}\text{Fe}_{1.95}$) changed this view. One of the aims of this chapter is to illustrate the importance of properly accounting for the constraints imposed on measurements.

The basics of magnetoelasticity are described here using both historical and modern notations. This chapter provides notationally consistent connecting relations appropriate to isotropic and cubic materials for these two sets of notations. A familiarity with the essentials of elasticity, for example as given by Thurston [1], is assumed of the reader. The linearized version of elastic theory with small strains is appropriate for our purposes. Du Trémolet de Lacheisserie provided a good review of magnetoelasticity in his book, which has an extensive reference list covering up to 1992 [2].

The chapter is organized as follows. The energy density terms necessary to describe the magnetoelastic interaction are presented in Section 2. That section also illustrates magnetoelastic effects with solutions for the equilibrium strains in simple cases. Section 3 delineates the major differences between measurements of magnetoelastic parameters made under static and dynamic conditions. Sections 4 and 5 describe various effects that can be adequately understood with current magnetoelastic energy density terms. Finally, in the conclusions various difficulties with current practice and with conventional expressions for magnetoelasticity are highlighted.

2 Magnetoelastic interaction

The coupled elastic and magnetic systems are conveniently described by an energy density that can be expressed formally as

$$U_{tot} = U_{elas} + U_{me} + U_{mag}, \quad (1)$$

where the terms on the right side of (1) represent the elastic, magnetoelastic and magnetic energy densities, respectively. The ultimate use of this energy density is to determine the forces and torques acting on a magnetoelastic medium. Of course, these forces and torques are simply derivatives of U_{tot} with respect to the relevant coordinates. These coordinates are usually taken to be the strains, the magnitude of the magnetization and the two angles describing the magnetization's direction relative to the elastic medium, and the three angles describing the orientation of the elastic medium. When interactions with electromagnetic fields are considered, the magnetic energy density U_{mag} includes terms that play the role of an interaction between these fields and the magnetoelastic medium. In electrically conducting mediums, the direct exchange of momentum between moving charges and the lattice represents another interaction, but it will not be considered further here [3].

It is the magnetoelastic energy density that has traditionally been used to describe the interaction between the elastic and magnetic systems and it is the one of central interest here. Before going on to it we will exhibit the elastic and magnetic energy densities to present the notation and concepts necessary to understand U_{me} .

2.1 Elastic energy

The strains that enter into U_{elas} and U_{me} have several definitions. If one follows the displacement of a material point, originally at the position described by the vector $\mathbf{R}(x, y, z)$ in a convenient, fixed Cartesian coordinate system (x, y, z) , then after a deformation that point will be at the position $\mathbf{R}'(x, y, z)$ and the displacement $\mathbf{u}(x, y, z)$ is defined by

$$\mathbf{R}'(x, y, z) = \mathbf{R}(x, y, z) + \mathbf{u}(x, y, z). \quad (2)$$

Vectors appear in boldface throughout. Two common definitions of the strain are

$$\begin{aligned} e_{11} &= \frac{\partial u_x}{\partial x}, & e_{12} = e_{21} &= \frac{\partial u_x}{\partial y} + \frac{\partial u_y}{\partial x}, \\ e_{22} &= \frac{\partial u_y}{\partial y}, & e_{23} = e_{32} &= \frac{\partial u_y}{\partial z} + \frac{\partial u_z}{\partial y}, \\ e_{33} &= \frac{\partial u_z}{\partial z}, & e_{13} = e_{31} &= \frac{\partial u_z}{\partial x} + \frac{\partial u_x}{\partial z}, \end{aligned} \quad (3)$$

and

$$\begin{aligned}\epsilon_{11} &= \frac{\partial u_x}{\partial x}, & \epsilon_{12} = \epsilon_{21} &= \frac{1}{2} \left(\frac{\partial u_x}{\partial y} + \frac{\partial u_y}{\partial x} \right), \\ \epsilon_{22} &= \frac{\partial u_y}{\partial y}, & \epsilon_{23} = \epsilon_{32} &= \frac{1}{2} \left(\frac{\partial u_y}{\partial z} + \frac{\partial u_z}{\partial y} \right), \\ \epsilon_{33} &= \frac{\partial u_z}{\partial z}, & \epsilon_{13} = \epsilon_{31} &= \frac{1}{2} \left(\frac{\partial u_z}{\partial x} + \frac{\partial u_x}{\partial z} \right).\end{aligned}\quad (4)$$

The description of strain using the e_{ij} 's, employed in Voigt's notation, is commonly used in older literature. For example, they can be found in the treatise by Love [4] and Kittel's book on solid state physics [5]. These strain components do not, however, form a second-rank tensor even though the notation is suggestive of such. It is the ϵ_{ij} 's of (4) that transform as a tensor and their use is preferred.

Magnetoelastic phenomena include rotations that traditionally have not been expressed in terms of the strains of (3) and (4). Following the notation of du Trémolet de Lacheisserie [2], the rotations can be described by the ω_{ij} 's defined by

$$\begin{aligned}\omega_{11} &= 0, & \omega_{12} = -\omega_{21} &= \frac{1}{2} \left(\frac{\partial u_x}{\partial y} - \frac{\partial u_y}{\partial x} \right), \\ \omega_{22} &= 0, & \omega_{23} = -\omega_{32} &= \frac{1}{2} \left(\frac{\partial u_y}{\partial z} - \frac{\partial u_z}{\partial y} \right), \\ \omega_{33} &= 0, & \omega_{13} = -\omega_{31} &= \frac{1}{2} \left(\frac{\partial u_z}{\partial x} - \frac{\partial u_x}{\partial z} \right).\end{aligned}\quad (5)$$

Then the Lagrangian strain tensor η_{ij} is

$$\eta_{ij} = \epsilon_{ij} + \frac{1}{2} \sum_{k=1}^3 (\epsilon_{ki} + \omega_{ki}) (\epsilon_{kj} + \omega_{kj}), \quad (6)$$

and the local rotation matrix $\mathcal{R}_{ij}$ is

$$\mathcal{R}_{ij} = \delta_{ij} + \omega_{ij} + \frac{1}{2} \sum_{k=1}^3 (\omega_{ik}\omega_{kj} - \omega_{ik}\epsilon_{kj} + \omega_{jk}\epsilon_{ki}), \quad (7)$$

where i and j take on the values 1, 2, 3; and the Kronecker delta function is $\delta_{ij} = 1$ if $i = j$, and $\delta_{ij} = 0$ otherwise. For small strains the η_{ij} 's and the ϵ_{ij} 's are essentially the same, since they only differ in second order. For general statements the η_{ij} 's are preferable, since they apply even if the strains are not small.

Group-theoretical considerations lead to certain combinations of strains being particularly useful in constructing the energy density in various types of crystals.

Here we limit ourselves to isotropic substances and cubic materials. The relevant combinations of Lagrangian strains are

Isotropic:	Cubic:	
$\eta^\alpha = \frac{\eta_{11} + \eta_{22} + \eta_{33}}{\sqrt{3}}$	$\eta^\alpha = \frac{\eta_{11} + \eta_{22} + \eta_{33}}{\sqrt{3}}$	
$\eta_1^\gamma = \frac{2\eta_{33} - \eta_{11} - \eta_{22}}{\sqrt{6}}$	$\eta_1^\gamma = \frac{2\eta_{33} - \eta_{11} - \eta_{22}}{\sqrt{6}}$	
$\eta_2^\gamma = \frac{\eta_{11} - \eta_{22}}{\sqrt{2}}$	$\eta_2^\gamma = \frac{\eta_{11} - \eta_{22}}{\sqrt{2}}$	(8)
$\eta_3^\gamma = \sqrt{2} \eta_{12}$	$\eta_1^\epsilon = \sqrt{2} \eta_{12}$	
$\eta_4^\gamma = \sqrt{2} \eta_{23}$	$\eta_2^\epsilon = \sqrt{2} \eta_{23}$	
$\eta_5^\gamma = \sqrt{2} \eta_{31}$	$\eta_3^\epsilon = \sqrt{2} \eta_{31}$	

along with similar expressions for the ϵ_{ij} 's.

The elastic energy density for cubic crystals has traditionally been expressed as

$$U_{elas} = \frac{1}{2} C_{11} (\epsilon_{11}^2 + \epsilon_{22}^2 + \epsilon_{33}^2) + C_{12} (\epsilon_{22}\epsilon_{33} + \epsilon_{33}\epsilon_{11} + \epsilon_{11}\epsilon_{22}) + 2C_{44} (\epsilon_{23}^2 + \epsilon_{31}^2 + \epsilon_{12}^2), \quad (9)$$

although the Lagrangian strains η_{ij} should replace the ϵ_{ij} 's if the strains are large. The strains are expressed in the coordinate system with axes chosen along the cube edges. In (9), C_{11} , C_{12} , and C_{44} have historically been used to describe the elastic properties of cubic materials. Isotropic materials require only two so-called material constants and several pairs of these constants have been commonly used. However, the energy density of (9) also applies to isotropic materials [6] provided $2C_{44} = C_{11} - C_{12}$. Expressions in terms of C_{11} and C_{12} for commonly used pairs of elastic constants such as Young's modulus and Poisson's ratio, the two Lamé constants, and the bulk and shear modulus are available in the literature [6].

The modern expression for the elastic energy density [2] for a cubic crystal, based on group theory, is

$$U_{elas}^{cub} = \frac{1}{2} C^\alpha (\eta^\alpha)^2 + \frac{1}{2} C^\gamma [(\eta_1^\gamma)^2 + (\eta_2^\gamma)^2] + \frac{1}{2} C^\epsilon [(\eta_1^\epsilon)^2 + (\eta_2^\epsilon)^2 + (\eta_3^\epsilon)^2]; \quad (10)$$

and for an isotropic material it is

$$U_{elas}^{iso} = \frac{1}{2} C^\alpha (\eta^\alpha)^2 + \frac{1}{2} C^\gamma \sum_{i=1}^5 (\eta_i^\gamma)^2, \quad (11)$$

where the modern elastic constants are related to the traditional ones by

$$C^\alpha = C_{11} + 2C_{12}, \quad C^\gamma = C_{11} - C_{12}, \quad C^\epsilon = 2C_{44}. \quad (12)$$

In the sequel we will also require stresses. For small strains, the stress σ_{ij} [6] is

$$\sigma_{ij} = \frac{\partial U_{tot}}{\partial \epsilon_{ij}}, \quad (13)$$

where the adiabatic (constant entropy) derivative is understood. The diagonal elements of σ_{ij} can be thought of as the force per area in the i th direction transmitted across an element of area oriented with its normal also in the i th direction. For the off-diagonal elements, the force is in the i th direction and is tangential to the element of area oriented in the j th direction. The force per volume at $\mathbf{R}$ is $\partial \sigma_{ij} / \partial x_j$, where the x_j are the coordinates of $\mathbf{R}$. Thus, the net force $\mathbf{F}$ on a body is

$$\mathbf{F} = \hat{i} \int \frac{\partial \sigma_{ij}}{\partial x_j} dV, \quad (14)$$

where $\hat{i}$ is the unit vector in the i th direction and both repeated indexes are summed.

2.2 Magnetic energy

The magnetic energy density is

$$U_{mag} = U_{ex} - \mathbf{M} \cdot \mathbf{B}_{ext} - \frac{1}{2} \mathbf{M} \cdot \mathbf{B}_{demag} + U_{an}, \quad (15)$$

where the terms on the right side of (15) are the exchange, Zeeman, demagnetization and anisotropy energy densities, respectively. The exchange energy density arises from the quantum-mechanical interactions that maintain neighboring spins aligned and is sensitive to changes in the electron number density. The Zeeman energy density represents energy associated with the magnetization's orientation with respect to an externally applied magnetic induction. The demagnetization energy density is a self-energy associated with divergences in the magnetization that occur primarily at the boundary of a ferromagnet. This term accounts for the magnetization's preference to lie along a rod's axis and not perpendicular to that axis. These terms are described in some detail elsewhere [7]. The anisotropy energy density U_{an} arises because the magnetization may favor certain crystallographic directions. An unstrained isotropic material, of course, possesses no anisotropy. The simplest kind of magnetic anisotropy one can deal with is a uniaxial anisotropy for which there is only one axis. Then, the magnetization is in its lowest energy state if it is directed along the axis (so-called easy axis) or in the plane perpendicular the axis (so-called hard axis). The historical approach to this energy density is to assert that

the uniaxial anisotropy energy can be described by a power series expansion in either $\sin \theta$ or $\cos \theta$, where θ is the angle between the direction of the magnetization and the axis. Under the transformation $\mathbf{M} \rightarrow -\mathbf{M}$ the energy must be unchanged. This eliminates odd powers of $\sin \theta$ or $\cos \theta$, and the uniaxial anisotropy energy density can be described by

$$\begin{aligned} U_{an}^{uni} &= K_{u0} + K_{u1} \sin^2 \theta + K_{u2} \sin^4 \theta + \dots \\ &= K'_{u0} + K'_{u1} \cos^2 \theta + K'_{u2} \cos^4 \theta + \dots, \end{aligned} \quad (16)$$

where the K_{ui} 's are the uniaxial anisotropy parameters. An unfortunate property of (16) is that the series is made up of nonorthogonal functions. If one determines the value of K_{u1} from experimental data by truncating the series after the second term, and then refines the experiment and extracts K_{u1} and K_{u2} , the mere inclusion of another term in the series will require that K_{u1} be modified. The values of the K_u 's in (16) that yield a best fit to data depend on both the choice of series and where one chooses to terminate that series. A much better expansion for U_{an}^{uni} would use the well-known Legendre polynomials, $P_l(\cos \theta)$, which form a complete, orthogonal set of basis functions. Unfortunately, the field has developed using a set of polynomials that are proportional to the $P_l(\cos \theta)$'s but have a different normalization. With the simple substitution of α , the direction cosine for $\cos \theta$, i.e., $\alpha = \cos \theta$, the modern expression for the uniaxial anisotropy [8] is

$$\begin{aligned} U_{an}^{uni} &= k_0 - k_2 \left(\alpha^2 - \frac{1}{3} \right) + k_4 \left(\alpha^4 - \frac{6}{7} \alpha^2 + \frac{3}{35} \right) - \dots \\ &= k_0 P_0(\alpha) - \frac{2}{3} k_2 P_2(\alpha) + \frac{5}{35} k_4 P_4(\alpha) - \dots \end{aligned} \quad (17)$$

As a practical matter there is very little need to go beyond terms involving α^4 and the series is usually truncated after the α^2 term. The k_i 's of (17) are related to the K_{ui} 's of (16) as per

$$k_2 = K_{u1} + \frac{8}{7} K_{u2}, \quad k_4 = K_{u2}. \quad (18)$$

Equation (18) holds when (16) is truncated after either the K_{u1} or K_{u2} terms, but not otherwise. The coefficients K_{u0} and k_0 are omitted from (18) since they cannot affect the torques derived from U_{an}^{uni} .

For cubic ferromagnets, one can consider the magnetic anisotropy as being formed by a combination of two sets of equivalent uniaxial-type axes. For example, we could have one set oriented along the three cube edges and the other set oriented along the four cube body diagonals [7], although this choice is not unique. There is no second-order contribution to the anisotropy energy since the direction cosines satisfy the relation $\alpha_1^2 + \alpha_2^2 + \alpha_3^2 = 1$. The historical expansion of the cubic

anisotropy energy density is

$$U_{an}^{cub} = K_0 + K_1 \left(\alpha_1^2 \alpha_2^2 + \alpha_2^2 \alpha_3^2 + \alpha_3^2 \alpha_1^2 \right) + K_2 \alpha_1^2 \alpha_2^2 \alpha_3^2. \quad (19)$$

In (19) each α_i is the cosine of the angle between the magnetization and a cube edge. The expression for U_{an}^{cub} can be continued to higher orders but it, too, is formed from nonorthogonal functions. A much better choice of basis functions is the Cubic Harmonics [9]. The modern expansion for U_{an}^{cub} , explicitly exhibiting the unnormalized polynomials of the Cubic Harmonics, is

$$U_{an}^{cub} = k_0 - \frac{1}{2}k_4 \left(\alpha_1^4 + \alpha_2^4 + \alpha_3^4 - \frac{3}{5} \right) + \frac{1}{3}k_6 \left[\alpha_1^6 + \alpha_2^6 + \alpha_3^6 - \frac{15}{11} \left(\alpha_1^4 + \alpha_2^4 + \alpha_3^4 \right) + \frac{30}{77} \right] - \dots, \quad (20)$$

where the modern anisotropy parameters are related to the classical ones by

$$k_4 = K_1 + \frac{1}{11}K_2, \quad k_6 = K_2. \quad (21)$$

2.3 Magnetoelastic energy

The expression for the magnetoelastic energy density was developed phenomenologically, since a microscopic understanding of the interaction has been available for only the last half century. One key to constructing this energy density is the observation that the specific volume, i.e., volume per gram, of magnetic materials depends on the magnitude of the magnetization but is (almost) completely independent of the direction of the magnetization. The part of the magnetoelastic energy density governing volume changes is dependent on only one coordinate (the magnetization's magnitude M) and, with modern notation, influences only one strain. Thus, one has $U_{me}^{vol}(\eta^\alpha, M, S)$ separate from the other magnetoelastic energy density terms, where S is the entropy. This energy term describes the volume magnetostriction. It leaves the rest of the magnetoelastic energy density to be determined as a function of the remaining five strains and two angles defining the direction of the magnetization. Unfortunately, the older notation for magnetostriction does not reflect this separation of volume from other coordinates. Entropy should be included as a coordinate as well. There is usually very little difference between adiabatic and isothermal material parameters describing magnetostriction, but care should be taken to state which constraint applies.

The essence of magnetostriction is that changes in the magnetization's direction cause an object to deform and, conversely, deforming a ferromagnetic object causes the equilibrium direction of the magnetization to change. The first of these effects is illustrated in Figure 1 for an isotropic ferromagnet. It is the spin-orbit interaction that causes a ferromagnet to either stretch (positive magnetostriction) or

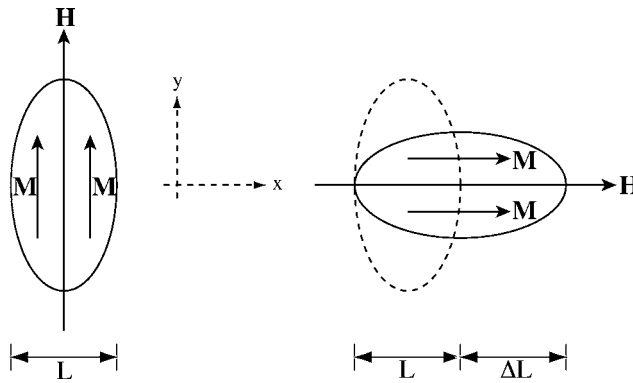


Figure 1 This is an illustration of Joule magnetostriction. Left: A ferromagnetic sample, originally spherical in the demagnetized state, is elongated when magnetized by an external magnetic field H oriented parallel to the y axis. The width of the sample transverse to the magnetic field is L . Right: The magnetic field has rotated parallel to the x axis, causing the magnetization M and the elongation to rotate as well. The width of the sample has increased by ΔL . The magnetostrictive strain is $\lambda = \Delta L/L$.

shrink (negative magnetostriction) in the direction of the magnetization. As illustrated in Figure 1, a 90° rotation of the magnetization introduces a magnetostrictive strain λ . The operation of rotating the magnetization between the two well-defined states of Figure 1 is easily accomplished in the laboratory. Unfortunately, the definition of the magnetostriction parameter λ_s involves magnetizing a material from an initial, completely random state with no net magnetization (which is often difficult to realize) to a final state with the magnetization saturated. The relation between these is $\lambda = 3\lambda_s/2$ [10].

Figure 2 illustrates the experimentally attainable case in which an imposed strain causes the magnetization to rotate. A large stress, $\sigma_{11} = F/A$, is applied to the material, resulting in an elongation in the x -direction and contraction in the y - and z -directions. Using a large externally imposed magnetic field to rotate the magnetization reveals, apart from torques associated with the demagnetization fields, a tendency for the magnetization to align either along the $\pm x$ -direction (positive magnetostriction) or perpendicular to this direction (negative magnetostriction). The imposed strain induces a uniaxial anisotropy energy that can be expanded in a power series in $\cos^2 \theta$ similar to (16). Since the strength of the anisotropy is strain-dependent, one can also expand the anisotropy energy in a power series in strain. Keeping only the lowest nontrivial term, one finds that $U_{me} \propto \epsilon_{11} \cos^2 \theta$. Of course, the energy density must also apply to strains in any direction. Thus, the simplest form of the magnetoelastic interaction energy density for isotropic ferromagnets is

$$U_{me}^{iso} = B \left[\epsilon_{11} \alpha_1^2 + \epsilon_{22} \alpha_2^2 + \epsilon_{33} \alpha_3^2 + 2(\epsilon_{23} \alpha_2 \alpha_3 + \epsilon_{31} \alpha_3 \alpha_1 + \epsilon_{12} \alpha_1 \alpha_2) \right], \quad (22)$$

where the magnetoelastic parameter B must be consistent with the situation shown in both Figures 1 and 2. For the more general case illustrated in Figure 2, the stress

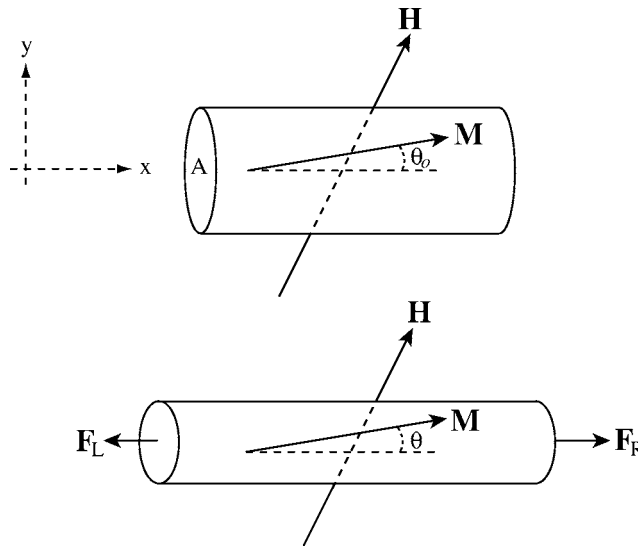


Figure 2 This is an illustration of strain-induced magnetic anisotropy. Top: A cylindrical ferromagnet of cross-sectional area A is magnetized by a magnetic field $\mathbf{H}$. Provided the cylinder is much longer than its diameter, the magnetization $\mathbf{M}$ is uniform everywhere except near the cylinder ends and attains the equilibrium direction θ_0 , which minimizes the Zeeman and shape demagnetization energies. Bottom: Equal and opposite forces $\mathbf{F}_R$ and $\mathbf{F}_L$ of magnitude F have been applied along the x axis. The stress $\sigma_{11} = F/A$ results in elongation of the rod. For a positive magnetostrictive ferromagnet, the magnetization rotates closer to the rod axis than the Zeeman and shape demagnetization energies require in order to minimize the magnetoelastic energy.

must satisfy (13); i.e.,

$$\begin{aligned}
 \sigma_{11} &= \frac{\partial U_{tot}^{iso}}{\partial \epsilon_{11}} \neq 0, & \sigma_{23} &= \frac{\partial U_{tot}^{iso}}{\partial \epsilon_{23}} = 0, \\
 \sigma_{22} &= \frac{\partial U_{tot}^{iso}}{\partial \epsilon_{22}} = 0, & \sigma_{31} &= \frac{\partial U_{tot}^{iso}}{\partial \epsilon_{31}} = 0, \\
 \sigma_{33} &= \frac{\partial U_{tot}^{iso}}{\partial \epsilon_{33}} = 0, & \sigma_{12} &= \frac{\partial U_{tot}^{iso}}{\partial \epsilon_{12}} = 0.
 \end{aligned} \tag{23}$$

Note that this is a quite general uniaxial stress. Since the material is isotropic, the coordinate system can always be chosen with the x -axis along the stress axis.

The expressions in (23) also apply to the case of Figure 1 with $\sigma_{11} = 0$. The U_{tot}^{iso} includes the U_{elas} of (9), with the isotropy condition $2C_{44} = C_{11} - C_{12}$, and U_{me}^{iso} of (22). All other terms which make up U_{tot}^{iso} , except the demagnetization energy, are independent of the strains and the direction of the magnetization. Omission of the demagnetization energy presents no difficulty in principle, since one can introduce external stresses to counteract stresses due to the sample shape, or one could use a spherical sample to minimize these stresses.

The solution of (23) for the strains that are the response to the imposed stress shown in Figure 2 is

$$\begin{aligned}
 \epsilon_{11} &= \frac{B[-\alpha_1^2(C_{11} + C_{12}) + (\alpha_2^2 + \alpha_3^2)C_{12}] + \sigma_{11}(C_{11} + C_{12})}{(C_{11} - C_{12})(C_{11} + 2C_{12})}, \\
 \epsilon_{22} &= \frac{B[-\alpha_2^2(C_{11} + C_{12}) + (\alpha_1^2 + \alpha_3^2)C_{12}] - \sigma_{11}C_{12}}{(C_{11} - C_{12})(C_{11} + 2C_{12})}, \\
 \epsilon_{33} &= \frac{B[-\alpha_3^2(C_{11} + C_{12}) + (\alpha_1^2 + \alpha_2^2)C_{12}] - \sigma_{11}C_{12}}{(C_{11} - C_{12})(C_{11} + 2C_{12})}, \\
 \epsilon_{12} &= -B \frac{\alpha_1 \alpha_2}{C_{11} - C_{12}}, \\
 \epsilon_{23} &= -B \frac{\alpha_2 \alpha_3}{C_{11} - C_{12}}, \\
 \epsilon_{31} &= -B \frac{\alpha_3 \alpha_1}{C_{11} - C_{12}}.
 \end{aligned} \tag{24}$$

If the magnetization is along the x -axis, i.e., $\alpha_1 = 1$ and $\alpha_2 = \alpha_3 = 0$, one can see from (24) that a positive B acts like a compressive uniaxial stress along the direction of the magnetization.

The strains appropriate to the case shown in Figure 1, which has no imposed stress, can be obtained from (24) by setting $\sigma_{11} = 0$. After calling the stress-free solution for the strains as ϵ_{ij}^0 , the magnetostriction parameter defined in Figure 1 is given by

$$\begin{aligned}
 \lambda &= \epsilon_{11}^0(\alpha_1 = 1, \alpha_2 = 0, \alpha_3 = 0) - \epsilon_{11}^0(\alpha_1 = 0, \alpha_2 = 1, \alpha_3 = 0) \\
 &= -\left(\frac{B}{C_{11} - C_{12}}\right).
 \end{aligned} \tag{25}$$

Positive magnetostriction corresponds to $\lambda_s > 0$ and $B < 0$. Note also that the volume strain, which is the sum of the diagonal elements of ϵ_{ij}^0 , is $-B/(C_{11} + 2C_{12})$. This is nonzero but independent of the direction of the magnetization, as it should be. It is a shortcoming of the energy density in (22) that it also includes a volume effect.

For the case shown in Figure 2 it is sensible to break U_{me}^{iso} into a part that depends on the equilibrium strains and a part that describes the response to the imposed stress; i.e.,

$$U_{me}^{iso} = U_{0me}^{iso}(\epsilon_{ij}^0) + U_{\Delta me}^{iso}(\epsilon_{ij} - \epsilon_{ij}^0, \alpha_i). \tag{26}$$

This is only possible because the magnetoelastic energy is linear in strain, and for an isotropic material this energy cannot depend on the direction of the magnetization relative to the crystal lattice. The explicit expressions for these energy density

terms are

$$U_{0me}^{iso} = \frac{-B^2(C_{11} + C_{12})}{(C_{11} - C_{12})(C_{11} + 2C_{12})} = \frac{-B^2}{E}, \quad (27)$$

and

$$\begin{aligned} U_{\Delta me}^{iso} &= \frac{-B\sigma_{11}C_{12}}{(C_{11} - C_{12})(C_{11} + 2C_{12})} + \frac{B\sigma_{11}\alpha_1^2}{C_{11} - C_{12}} \\ &= \frac{-B\sigma_{11}\sigma_p}{E} - \frac{3}{2}\lambda_s\sigma_{11}\cos^2\theta, \end{aligned} \quad (28)$$

where the second form of (27) and (28) has been obtained by using Young's modulus $E = (C_{11} - C_{12})(C_{11} + 2C_{12})/(C_{11} + C_{12})$, Poisson's ratio $\sigma_p = C_{12}/(C_{11} + C_{12})$, and by using (25) together with the relation $\lambda_s = 2\lambda/3$. The last term on the second line of (28) exhibits the magnetoelastic energy density associated with the direction of the magnetization and was used in the older literature to represent the magnetoelastic interaction. Note as well that the elastic energy density also has terms similar to those of (28) due to the static strains of (24).

The first terms of the modern expansion of the magnetoelastic energy, entirely equivalent to (22), can be expressed in terms of η^α and η_i^γ . The α_i are grouped in the spirit of (8) with the substitution $\eta_{ij} \rightarrow \alpha_i\alpha_j$ in the P 's; i.e., $P^{\alpha,0}(\alpha_i) = (\alpha_1^2 + \alpha_2^2 + \alpha_3^2)/\sqrt{3} = 1\sqrt{3}$, $P_1^{\gamma,2}(\alpha_i) = 2\alpha_3^2 - \alpha_1^2 - \alpha_2^2/\sqrt{6}$, etc. Then [2],

$$U_{me}^{iso} = b^{\alpha,0}P^{\alpha,0}(\alpha_i)\eta^\alpha + b^{\gamma,2}\sum_{j=1}^{j=5}P_j^{\gamma,2}(\alpha_i)\eta_j^\gamma, \quad (29)$$

where $b^{\alpha,0}$, a function of magnetization, entropy and volume, is the term describing the volume magnetostriction while $b^{\gamma,2} = B$ describes the magnetostriction. Note that $P^{\alpha,0}(\alpha_i)$ and $P_j^{\gamma,2}(\alpha_i)$ are not Legendre polynomials.

For cubic symmetry, one requires two material parameters to characterize the magnetoelastic interaction energy density in lowest order. The traditional form of this energy density is similar to (22) with different parameters for axes along the cube edge and along the body diagonals. This energy density is

$$U_{me}^{cub} = B_1(\epsilon_{11}\alpha_1^2 + \epsilon_{22}\alpha_2^2 + \epsilon_{33}\alpha_3^2) + 2B_2(\epsilon_{23}\alpha_2\alpha_3 + \epsilon_{31}\alpha_3\alpha_1 + \epsilon_{12}\alpha_1\alpha_2), \quad (30)$$

where B_1 and B_2 are the material's magnetoelastic parameters. The α_i of (29) and (30) are the same as those defined in (19), and the coordinate system used to describe ϵ_{ij} is chosen with axes along the cube edges.

Imposing a uniaxial stress of magnitude σ on a material with the stress axis along the direction described by the direction cosines n_1 , n_2 , and n_3 results in the stress tensor $\sigma_{ij} = \sigma n_i n_j$ [6]. Setting these stress components equal to the derivative of the total energy density with respect to strains as in (23) leads to the following equilibrium strains:

$$\begin{aligned}\epsilon_{11} &= \frac{(\sigma n_1^2 - B_1 \alpha_1^2)(C_{11} + 2C_{12}) - (\sigma - B_1)C_{12}}{(C_{11} - C_{12})(C_{11} + 2C_{12})}, & \epsilon_{23} &= \frac{\sigma n_2 n_3 - B_2 \alpha_2 \alpha_3}{2C_{44}}, \\ \epsilon_{22} &= \frac{(\sigma n_2^2 - B_1 \alpha_2^2)(C_{11} + 2C_{12}) - (\sigma - B_1)C_{12}}{(C_{11} - C_{12})(C_{11} + 2C_{12})}, & \epsilon_{31} &= \frac{\sigma n_3 n_1 - B_2 \alpha_3 \alpha_1}{2C_{44}}, \\ \epsilon_{33} &= \frac{(\sigma n_3^2 - B_1 \alpha_3^2)(C_{11} + 2C_{12}) - (\sigma - B_1)C_{12}}{(C_{11} - C_{12})(C_{11} + 2C_{12})}, & \epsilon_{12} &= \frac{\sigma n_1 n_2 - B_2 \alpha_1 \alpha_2}{2C_{44}}.\end{aligned}\quad (31)$$

In the absence of an applied stress the material is strained into a state described by ϵ_{ij}^0 . This strain can be found from (31) by setting $\sigma = 0$. Substituting these strains into the expressions for the elastic and magnetoelastic energy densities yields, after some tedious algebra [11],

$$\begin{aligned}U_{0elas}^{cub} + U_{0me}^{cub} &= \left(\frac{B_1^2}{C_{11} - C_{12}} - \frac{B_2^2}{2C_{44}} \right) (\alpha_1^2 \alpha_2^2 + \alpha_2^2 \alpha_3^2 + \alpha_3^2 \alpha_1^2) \\ &+ \dots \text{(other terms)}.\end{aligned}\quad (32)$$

The term exhibited in (32) has the same dependence on the direction of the magnetization as the K_1 term in (19). This reflects the fact that the equilibrium magnetostrictive strain results in the lowest energy when the magnetization is aligned along certain crystal axes, and the energy density associated with this term has the same form as that of a magnetic anisotropy.

The modern form of the magnetoelastic interaction is

$$\begin{aligned}U_{me}^{cub} &= b^{\alpha,0} P^{\alpha,0}(\alpha_i) \eta^\alpha + b^{\gamma,0} \left[P_1^{\gamma,0}(\alpha_i) \eta_1^\gamma + P_2^{\gamma,0}(\alpha_i) \eta_2^\gamma \right] \\ &+ b^{\epsilon,0} \left[P_1^{\epsilon,0}(\alpha_i) \eta_1^\epsilon + P_2^{\epsilon,0}(\alpha_i) \eta_2^\epsilon + P_3^{\epsilon,0}(\alpha_i) \eta_3^\epsilon \right],\end{aligned}\quad (33)$$

where $P^{\alpha,0}(\alpha_i)$, $P_k^{\gamma,0}(\alpha_i)$, and $P_k^{\epsilon,0}(\alpha_i)$ are functions of α_i as defined in (8) with $\eta_{ij} \rightarrow \alpha_i \alpha_j$ and $\eta^\alpha \rightarrow P^{\alpha,0}(\alpha_i)$, etc. As with the isotropic case, $b^{\alpha,0}$ is associated with the volume magnetostriction; furthermore, $b^{\gamma,0} = B_1$ and $b^{\epsilon,0} = B_2$.

2.4 Total magnetoelastic interaction energy density

The foregoing sections outline the historical energy density terms used to describe magnetoelasticity together with equivalent modern statements of the same terms. Although complete enough to describe most currently recognized phenomenons,

the nomenclature is misleading since U_{me} does not represent all the magnetoelastic interaction. For example, consider a cubic ferromagnetic crystal that is nominally nonmagnetostrictive with $B_1 = B_2 = 0$ but with nonzero anisotropy. A spherical sample of this material will experience a torque when exposed to a large external magnetic field. This arises because the magnetization favors certain directions in the crystal and, with the direction of the magnetization held in place by the external magnetic field, there is a tendency for the crystal to rotate so that a favored axis is aligned with the magnetization. Indeed, the magnetic anisotropy parameters (the K 's) can be found by measuring this torque. The external torque Γ required to hold the crystal in place is

$$\Gamma = \frac{\partial U_{tot}}{\partial \phi}, \quad (34)$$

where the angle ϕ is a rotation of the crystal about the axis along which Γ is measured. The anisotropy energy density, through its dependence on the direction of the magnetization with respect to the crystal axes, enters (34) as does the Zeeman energy, $-\mathbf{M} \cdot \mathbf{B}$. Some care is required in carrying out the differentiation, and this is examined in the next section.

Even the shape of the sample must be taken into account in (34). For example, a rodlike ferromagnet with no anisotropy or magnetostriction will tend to align with its long axis parallel an externally applied magnetic field. In this configuration, the demagnetizing energy is minimized. For the same reason, a spherical sample will elastically deform and elongate in the direction of an external field. This latter phenomenon is called the form effect.

Thus, the total elastic response of a ferromagnet involves not just the elastic and magnetoelastic energy densities but also the Zeeman, shape demagnetizing, and anisotropy energy densities.

3 Static and dynamic measurements

The material parameters describing magnetoelasticity can be determined from experiments that fall into two categories: static (or quasi-static) and dynamic. With static measurements, one typically controls a parameter such as applied magnetic field or stress, and measures a response such as a change in the strain or the magnetization of the ferromagnet. In effect, one measures properties of the equilibrium state of the ferromagnet that depend on the first derivative of U_{tot} with respect to the coordinates that are allowed to change in the experiment. Usually, but not always, these measurements are made with the ferromagnet in equilibrium at a constant temperature. With dynamic measurements one holds the ferromagnet in an equilibrium reference state and makes measurements of small oscillations about that equilibrium. For example, one could make ultrasonic measurements such as determining the speed of sound [12], or one could excite the ferromagnet with microwaves and measure the frequency and magnetic field required for ferromagnetic

resonance (FMR) absorption to occur. This type of experiment explores the second derivative of U_{tot} with respect to small deviations of the strain, magnetization, etc., from equilibrium values. Dynamic measurements are usually, but not always [13], done under adiabatic conditions.

The extraction of material parameters such as the B 's and the K 's from experiments yields results that depend on whether the experiments were performed under isothermal or adiabatic conditions. Thermodynamics allows us to consistently relate these results. It is unfortunate that the B 's and K 's have become known as the magnetostriction and anisotropy *constants* since, even in the simplest case and just as for the C 's, they are functions of temperature, pressure and magnetic field (or entropy, volume and magnetization).

Common practice has led to a serious error in evaluating the anisotropy parameters. The typical determination of K is made in a static, quasi-stress-free experiment. An applied magnetic field $\mathbf{H}$ causes a spherical sample's magnetization to experience a torque, $\mathbf{M} \times \mathbf{H}$, which can be nonzero because, in general, the anisotropy is such that the magnetization's equilibrium direction is not parallel to $\mathbf{H}$. This torque is communicated to the elastic medium through the anisotropy and acts throughout the volume of the sample. One measures an imposed torque necessary to balance the torque caused by the magnetic field and maintain the sample at rest. This imposed torque is typically applied to a small portion of the sample's surface. Even though the torques are not applied at the same points, they are assumed to result in a stress-free sample, provided the sample is sufficiently rigid. Relaxing this assumption will not alter (35) below. Measurements of the balancing torque are made for many orientations of the crystal axes, and one infers the values of the K 's by requiring the angular derivative of U_{tot} calculated to match the measured torque. For example, K_1 and K_2 enter into terms found by differentiating the contribution to U_{tot} of U_{an} in (19). Under isothermal conditions one actually differentiates the free energy found by performing a Legendre transformation on U_{tot} . Since the experiment is carried out at zero net stress, the equilibrium strains obtained from (31) with $\sigma = 0$ change as the magnetization is reoriented and a term such as the first one in (32) is also present in the energy density U_{tot} . Since both terms have the angular variation $\alpha_1^2\alpha_2^2 + \alpha_2^2\alpha_3^2 + \alpha_3^2\alpha_1^2$, K_1^{static} obtained from the measurement is [11]

$$K_1^{static} = K_1 + \frac{B_1^2}{C_{11} - C_{12}} - \frac{B_2^2}{2C_{44}}. \quad (35)$$

Similar considerations apply to K_2^{static} , which is related to K_2 through higher-order magnetoelastic parameters. Note that static measurements are usually made at constant temperature and it is the isothermal K 's, C 's, and B 's that enter into (35). In the sequel, these parameters will have the superscript T whenever they must be distinguished from adiabatic quantities. There is a tendency to ignore the difference between K_1^{static} and K_1 since it is typically less than 10% of K_1^{static} [11]. This is intolerable for Terfenol-D, a material designed to have a small, negative K_1^{static} near room temperature and for which $K_1 \approx -3K_1^{static}$ [14].

In a dynamic measurement, the equilibrium strain ϵ_{ij}^0 is fixed and one measures $K^{dynamic}$. For example, in a FMR measurement one excites the magnetization from its equilibrium direction and ascribes to anisotropy any restoring torques on the magnetization associated with crystal orientation. The $K^{dynamic}$ one measures is different from K^{static} as well as the K 's describing the anisotropy energy.

The simplest case that clearly demonstrates the difference between $K^{dynamic}$ and the other K 's is that of a magnetostrictive isotropic ferromagnet. The relations developed for cubic ferromagnets apply in this case, provided

$$C_{11} - C_{12} = 2C_{44}, \quad B_1 = B_2 = B, \quad K_1 = K_2 = 0. \quad (36)$$

This is just the case illustrated in Figure 1. If the magnetization is slowly rotated, the elastic bulge associated with the direction of the magnetization follows along, and both the elastic energy density U_{elas} and the magnetoelastic energy density U_{me} are unchanged. Thus, one does not require an external torque to maintain the sample in mechanical equilibrium, and one infers that there is no magnetic anisotropy; i.e., $K^{static} = 0$ as substitution of (36) into (35) predicts. In a dynamic measurement, however, one quickly changes the orientation of the magnetization without allowing the elastic bulge to move. The magnetization will then experience a torque caused by the static strain and one would detect an anisotropy; i.e., $K^{dynamic} \neq 0$. In effect, one disregards the Joule magnetostriction and demands that all the anisotropy be described by the K 's.

The $K^{dynamic}$ one would find for this isotropic case, i.e., $K^{static} = K_1 = K_2 = 0$ but $B \neq 0$, is due to the static strain imposed on the lattice. Changing the direction of the magnetization has no effect on U_{elas} since the static, isothermal strain $e_{ij}^{0,T}$ that enters (9) is constant. In fact, any deviations of the strain from $e_{ij}^{0,T}$ describe a sound wave, whereas the FMR measurement primarily involves electromagnetic waves. In this case the magnetoelastic energy contributes the following to U_{tot} :

$$U_{me} = B \left[\epsilon_{11}^{0,T} \alpha_1^2 + \epsilon_{22}^{0,T} \alpha_2^2 + \epsilon_{33}^{0,T} \alpha_3^2 + 2 \left(\epsilon_{23}^{0,T} \alpha_2 \alpha_3 + \epsilon_{31}^{0,T} \alpha_3 \alpha_1 + \epsilon_{12}^{0,T} \alpha_1 \alpha_2 \right) \right]. \quad (37)$$

This energy is then added to the anisotropy energy ($K_1^S \neq 0$ even though $K_1^T = 0$), and magnetoelastic terms are folded into the K 's to yield $K^{dynamic}$. Consider, for example, the case in which the equilibrium magnetization has the direction cosines $\alpha_1^0 = 1$, $\alpha_2^0 = 0$, and $\alpha_3^0 = 0$. With the instantaneous magnetization's direction described by $\alpha_i = \alpha_i^0 + \delta\alpha_i$, where $\delta\alpha_i$ represents the deviation of the magnetization from its equilibrium direction,

$$U_{me} + U_{an} = B^S \epsilon_{11}^{0,T} + \left(K_1^S + \frac{B^S B^T}{C_{11}^T - C_{12}^T} \right) \left[(\delta\alpha_2)^2 + (\delta\alpha_3)^2 \right] + \dots, \quad (38)$$

where there is no term in K_2^S of order $\delta\alpha^2$ and the constraint $\alpha_1^2 + \alpha_2^2 + \alpha_3^2 = 1$ has been used to eliminate $\delta\alpha_1$. The superscript S denotes the constraint that a material parameter is determined at constant entropy provided, of course, that the experiment was carried out a frequency for which the adiabatic condition held. Thus, one finds *for this orientation of the magnetization* that

$$K_1^{\text{dynamic}} = K_1^S + \frac{B^S B^T}{C_{11}^T - C_{12}^T}. \quad (39)$$

Even though the material was assumed to be magnetically isotropic with respect to static measurements, i.e., $K^{\text{static}} = 0$, it has a nonzero dynamic anisotropy.

It is common to find that K^{static} has been misapplied to the dynamic case [15]. It is also a poor practice to fold the magnetostrictive strain into the anisotropy in analyzing dynamic measurements. The magnetostrictive energy is second order in the direction cosines whereas the first cubic anisotropy term is fourth order. Combining these terms is contrary to the spirit of modern notation. In essence, the magnetization causes the lattice to strain, and any attempt to dynamically reorient the magnetization will cost some energy with respect to that strain. The static strain behaves like a uniaxial anisotropy with an easy axis parallel to the equilibrium direction of the magnetization. The intrinsic magnetocrystalline anisotropy and magnetostrictive anisotropy should be treated separately.

4 Villari and ΔE effects

The Villari and ΔE effects both involve the mutual influence of imposed stresses and the state of magnetization of ferromagnets. In the Villari effect the response of the magnetization to an external magnetic field is affected by the stress, while in the ΔE effect the elastic response of the ferromagnet is affected by the state of the magnetization. Both effects can be understood in terms of the magnetostrictive energy density and micromagnetics.

These effects are apparent when the ferromagnet is not magnetically saturated. Although the details of the transition from a demagnetized state in which a material is broken into domains to a uniformly magnetized state are quite complex [16], the general outline is straightforward. With the Villari effect, one imposes a large stress (or strain) on a ferromagnetic sample and then measures either M versus H or $\partial M / \partial H$ versus H as the external magnetic field H is increased from zero to some large, finite value. If the imposed strain is large enough, a uniaxial anisotropy of the sort described by (30) or (33) will induce many of the magnetic domains to align with their magnetization either along (positive magnetostriction) or perpendicular (negative magnetostriction) to that axis while still maintaining a net magnetization of zero for $H = 0$. Depending on whether this strain-induced anisotropy aids or hinders, the reorientation of the individual magnetic domains in the direction of the subsequently applied H determines whether the net magnetization is larger or smaller than it would otherwise be without the imposed strain. This effect is

technologically important for Terfenol-D actuators [17]. A large prestress is used to preferentially orient much of the magnetization perpendicular to the actuator's line of action, and a relatively small applied magnetic field is used to flip these domains' magnetization parallel to the line of action, with a resulting change in the strain parallel to that line.

The E in ΔE refers to Young's modulus. If one applies a positive stress along the axis of a rod, it will elongate in the direction of the stress and the ratio of stress to strain is Young's modulus E . For a uniformly magnetized magnetostrictive material, $1/E$ is the coefficient of σ_{11} in the expression for ϵ_{11} in (24). This is identical to the expression for a nonmagnetic material and demonstrates that there is no ΔE for uniformly magnetized magnetic materials. However, with a partially magnetized rod the application of the stress will cause some domains to flip to a lower energy configuration, which invariably results in a larger strain in the direction of the imposed stress—the material is elastically softer. Thus the material has a smaller Young's modulus E , provided domains are available to reorient as the stress is applied. A description of this effect can be found in Bozorth's book [10].

5 Wiedemann effect

The Wiedemann effect refers to the tendency of a current-carrying wire to twist under the influence of an applied magnetic field. Unlike the Villari and ΔE effects, this effect persists when the magnetic field is large enough to magnetically saturate the wire. Related effects are the inverse Wiedemann effect and Matteucci effect. In the inverse Wiedemann effect (of which there are several variations [2]), a longitudinally magnetized wire or one carrying a current exhibits, under the influence of an externally applied torque, changes in magnetization along the wire's axis as well as perpendicular to that axis. The Matteucci effect refers to the development of a voltage across the ends of the wire as the magnetization responds to the applied torque.

The Wiedemann effect can be understood with the aid of Figure 3. As shown, a current-carrying wire has a large magnetic field imposed parallel to the wire. In the absence of any magnetostriction or anisotropy, the magnetization aligns with the total magnetic field, which, of course, includes a circulating component caused by the magnetic field produced by the current. On the axis of the wire the magnetization aligns parallel to the axis. At a nonzero distance from the axis the magnetization is helical and the pitch of the helix becomes tighter as the distance from the axis increases. If the wire is magnetostrictive, the magnetization is slightly modified from this simple nonmagnetostrictive case but qualitatively unchanged.

An element of volume of a positive magnetostrictive material elongates in the direction of the magnetization. Since the elongation has an azimuthal component, the wire twists. It is this twist that constitutes the Wiedemann effect, but the effect

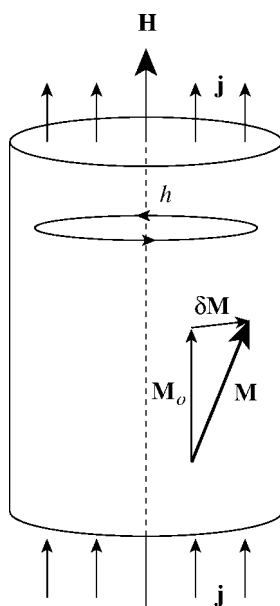


Figure 3 The Wiedemann effect occurs in the situation illustrated. A wire with a static magnetic field $\mathbf{H}$ along its axis also carries a current density $\mathbf{j}$. In the absence of the current, the equilibrium magnetization $\mathbf{M}_0$ aligns with the static magnetic field. The magnetic field h due to the current causes the magnetization to develop an azimuthal component $\delta\mathbf{M}$, which results in a total magnetization $\mathbf{M}$ that spirals around the wire's axis. In a positive magnetostrictive material, the elongation of elements of the wire in the direction of $\mathbf{M}$ leads to a twist of the wire.

is quite small. For example, the twist of a nickel wire carrying a few A/mm² in a moderate magnetic field is of the order of 0.03° per cm [18].

6 Conclusion

The energy densities described in this chapter provide an adequate basis for understanding most phenomenons involving interactions between the elastic and magnetic degrees of freedom (the strains and the components of the magnetization) of a ferromagnet. The historical development of these energy densities has an *ad hoc* nature wherein, for example, Joule magnetostriction and magnetic anisotropy were viewed, erroneously, as independent phenomenons. Modern notation based on group theory restates these interactions in a rational manner that allows higher-order interactions to be treated consistently. However, the onus is on individual researchers to carefully state the conditions under which measurements are made and state their results consistently so that one does not have to deal with, for example, K_1^{static} and $K_1^{dynamic}$, which are really combinations of anisotropy, magnetostriction, and elastic parameters extracted from measurements made under isothermal or adiabatic conditions.

The stresses that are attributable to the shape of a ferromagnet are implicitly included in the demagnetization energy term of (15). A more transparent expression would be welcome.

Finally, there is a serious shortcoming of the description of magnetoelasticity given here which manifests itself in two ways, both involving torques. Classical elasticity requires only that the stress satisfy (13) and (14). There is enough freedom in these requirements that the stress can always be made symmetrical, i.e., $\sigma_{ij} = \sigma_{ji}$. Even when a torque is introduced, such as the twisting of a rod about its axis, a symmetrical stress tensor can be found [6]. However, the derivation of the twist in the Wiedemann effect is presented, as for example by du Trémolet de Lachesserie [2], in an *ad hoc* fashion with a nonsymmetrical stress tensor. I am not aware of any derivations using a symmetrical stress tensor. Is the use of the nonsymmetrical stress a convenience or a necessity?

A more severe difficulty arises in the description of the Einstein–de Hass effect. When a rod, magnetized uniformly along its axis and suspended at rest so that it is free to rotate about its axis, has its magnetization reversed, the rod will rotate. This happens because there is angular momentum associated with the magnetization and, when the magnetization together with its angular momentum are reversed, the elastic material of the rod must recoil with an equal and opposite change in its angular momentum [19–20]. While the magnetization is reversing there is a flow of angular momentum into the elastic material, which can be described as a body couple acting on the material. A couple also arises in considering the so-called spin pinning boundary condition. This boundary condition is invoked to describe the excitation of spin waves in microwave experiments and it implies that a surface torque is exerted on the magnetization. The reaction of this surface couple on the elastic medium is ignored.

Theories of elasticity incorporating couples are not fully developed and some have been shown to violate conservation of energy and momentum [21]. Correct theories including couples, such as the one described by Mindlin and Tiersten [22] and by Toupin [23], have not been developed to the point where they can handle a time-dependent intrinsic angular momentum. For example, see the restriction imposed by Toupin [23, Eq. 5.21]. Thus, the description of magnetoelastic phenomena is not yet complete.

References

- [1] R.N. Thurston, Wave propagation in fluids and normal solids, *Physical acoustics, Vol. I Part A* (W.P. Mason, ed), Academic Press, New York, NY, USA, 1964.
- [2] E. du Trémolet de Lachesserie, *Magnetostriction*, CRC Press, Boca Raton, FL, USA, 1993.
- [3] G. Dewar, Electromagnetic generation of gigahertz sound in ferromagnetic metals, *Phys Rev B* **36** (1987), 7805–7815.

- [4] A.E.H. Love, *A treatise on the mathematical theory of elasticity*, Dover Press, New York, NY, USA, 1944.
- [5] C. Kittel, *Introduction to solid state physics*, Wiley, New York, NY, USA, 1996.
- [6] L.D. Landau and E.M. Lifshitz, *Theory of elasticity*, Pergamon Press, New York, NY, USA, 1986.
- [7] G. Dewar, Magnetostriction and Terfenol-D, *Complex mediums* (A. Lakhtakia, W.S. Weiglhofer and R.F. Messier, eds), Proc SPIE **4097**, Bellingham, WA, USA, 2000, 134–144.
- [8] E.R. Callen and H.B. Callen, Anisotropic magnetization, *J Phys Chem Solids* **16** (1960), 310–328.
- [9] F.C. Von der Lage and H.A. Bethe, A method for obtaining electronic eigenfunctions and eigenvalues in solids with an application to sodium, *Phys Rev* **71** (1947), 612–622.
- [10] R.M. Bozorth, *Ferromagnetism*, Van Nostrand, New York, NY, USA, 1951.
- [11] C. Kittel, Physical theory of ferromagnetic domains, *Rev Mod Phys* **21** (1949), 541–583.
- [12] R.C. Le Craw and R.L. Comstock, Magnetoelastic interactions in ferromagnetic insulators, *Physical acoustics, Vol. III Part B* (W.P. Mason, ed), Academic Press, New York, NY, USA, 1965.
- [13] G. Dewar, The effect of the entropy wave on sound and electromagnetic waves in ferromagnetic metals, *J Appl Phys* **64** (1988), 5873–5875.
- [14] G. Dewar, S. Pagel and P. Sourivong, Ferromagnetic resonance in Terfenol-D at 17 GHz, *Int J Mod Phys B* **15** (2001), 3266–3269.
- [15] C. Vittoria, J.N. Craig and G.C. Bailey, General dispersion law in a ferromagnetic cubic magnetoelastic conductor, *Phys Rev B* **10** (1974), 3945–3956.
- [16] W.F. Brown, Jr., *Micromagnetics*, Interscience, New York, NY, USA, 1963.
- [17] A.E. Clark, J.P. Teter and O.D. McMasters, Magnetostriction “jumps” in twinned Tb_{0.3}Dy_{0.7}Fe_{1.9}, *J Appl Phys* **63** (1988), 3910–3912.
- [18] H.A. Pidgeon, Magnetostriction with special reference to pure cobalt, *Phys Rev* **13** (1919), 209–237.
- [19] C. Kittel, On the gyromagnetic ratio and spectroscopic splitting factor of ferromagnetic substances, *Phys Rev* **76** (1949), 743–748.
- [20] J.H. Van Vleck, Concerning the theory of ferromagnetic resonance absorption, *Phys Rev* **78** (1950), 266–274.
- [21] R.D. Mindlin, *Report US Signal Corps contract DA 36-039 SC-87414*, 1961.
- [22] R.D. Mindlin and H.F. Tiersten, Effects of couple-stresses in linear elasticity, *Arch Rat Mech Anal* **11** (1962), 415–448.
- [23] R.A. Toupin, Elastic materials with couple-stresses, *Arch Rat Mech Anal* **11** (1962), 385–414.



Graeme Dewar was born in Ontario, Canada, and earned a Ph.D. in Physics in 1980 from Simon Fraser University. He is currently an Associate Professor of Physics at the University of North Dakota. Most of his research projects have involved the experimental investigation of the interaction of electromagnetic radiation, principally microwaves, with complex mediums, which has resulted in more than 30 publications.

Frequency Shifts Induced by a Time-Varying Magnetoplasma Medium

Dikshitulu K. Kalluri

Abstract

An electromagnetic wave is transformed in a remarkable way by a time-varying magnetoplasma medium. The main effect of the temporal change in the parameters of the medium is the splitting of a source wave into new waves whose frequencies are different from the source wave frequency. Several problems involving either slow/fast creation or slow/fast collapse of the plasma medium in the presence of a static magnetic field are discussed. The more important results are (i) frequency upshifting with power intensification of a whistler wave by a collapsing plasma medium, and (ii) conversion of a whistler wave into a controllable helical wiggler magnetic field. A switched magnetoplasma medium can act like a frequency transformer.

1 Introduction

This book deals with *complex mediums* of various kinds and their applications. Some of these complexities are nonhomogeneity, dispersion, biisotropy, anisotropy, and combinations thereof. The medium complexity is expressed through constitutive relations. A comprehensive and rigorous presentation of the constitutive relations for various kinds of complex mediums is given elsewhere in this book [1].

An additional aspect of the medium complexity that is of current research interest [2] arises out of the time-varying parameters of the medium. The transformation of the frequency of an electromagnetic wave by a general time-varying medium is rarely discussed in books on electromagnetics, even though the Doppler effect (a frequency change due to a moving medium) is a standard topic in many books. The moving-medium problem is a particular case of a time-varying medium.

The frequency change in a nonmoving medium is contrary to usual experiences. The chapter deals with frequency shifts of several orders of magnitude that can be achieved by adding the complexity of time-varying parameters on top of the complexity of an anisotropic medium—in particular, the magnetoplasma medium. A time-varying magnetoplasma medium can act as a frequency transformer with a large frequency transformation ratio of the frequency of the output wave to the input wave (frequency upshifting) or a very small frequency transformation ratio (frequency downshifting). This remarkable effect is discussed in this chapter by using simple ideal models for the geometry of the problem as well as the constitutive relations of the magnetoplasma medium. The subject thus becomes more accessible and the focus is on the effect rather than on the achievement of high accuracy in the results. Accurate results can be obtained by using the finite difference time domain (FDTD) method for the numerical simulation of three-dimensional problems [3], which is not discussed here.

The chapter is organized as follows. In Section 2 the effect of a temporal discontinuity as opposed to a spatial discontinuity in the properties of the medium is discussed, and a simple explanation for the frequency shift caused by a temporal discontinuity is given. In Section 3 the constitutive relations for a time-varying plasma medium are discussed. The sudden switching of an unbounded plasma medium is considered in Section 4. Section 5 deals with the more realistic problem of switching a plasma slab, and Section 6 discusses applications under development of the frequency shifting research. In Section 7, the theory of wave propagation in a time-varying magnetoplasma medium and the possibility of effecting a big change in the relative permittivity of the medium by changing the ionization level or the background magnetic field are explained. Such a big change in the relative permittivity leads to a large frequency shift.

2 Frequency change due to a temporal discontinuity in the medium properties

Let us consider normal incidence on a spatial discontinuity in the dielectric properties of a medium, of a plane wave propagating in the z -direction. The spatial step profile of the permittivity ε is shown at the top of Figure 1(a).

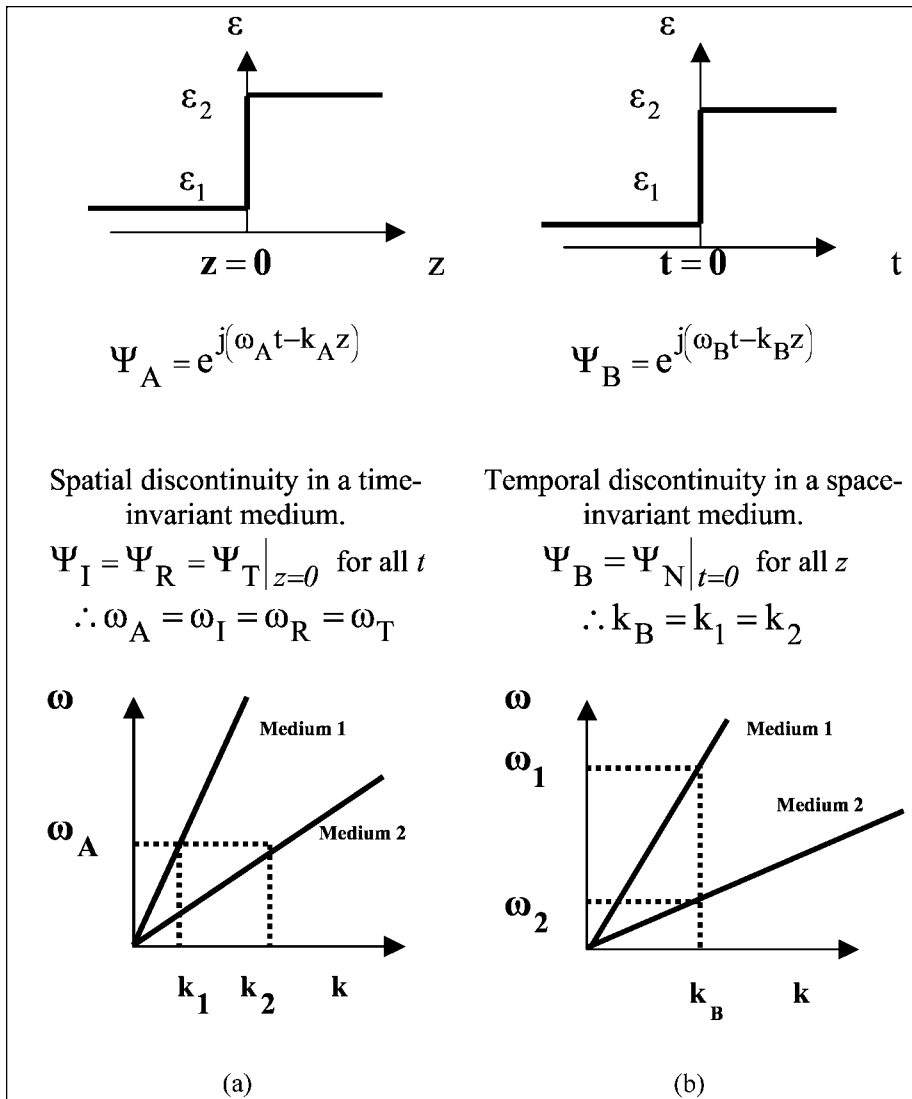


Figure 1 Comparison of the effects of temporal and spatial discontinuities.

The permittivity suddenly changes from ϵ_1 to ϵ_2 at $z = 0$. Let us also assume that the permittivity profile is time-invariant. The phase factors of the incident, reflected, and transmitted waves are expressed as $\psi_A = e^{j(\omega_A t - k_A z)}$, where the subscript $A = I$ for the incident wave, $A = R$ for the reflected wave, and $A = T$ for the transmitted wave. The boundary condition of the continuity of the tangential component of the electric field at $z = 0$ for all t requires

$$\omega_I = \omega_R = \omega_T = \omega_A. \quad (1)$$

This result may be stated as follows: the frequency ω is conserved across a spatial discontinuity in the properties of the electromagnetic medium. As the wave

crosses from one medium to the other in space, the wavenumber k changes as dictated by the change in the phase velocity, not considering absorption here. The bottom part of Figure 1(a) illustrates this aspect graphically. The slopes of the two straight lines in the ω - k diagram are the phase velocities in the two mediums. Conservation of ω is implemented by drawing a horizontal line, which intersects the two straight lines. The k values of the intersection points give the wavenumbers in the two mediums.

A dual problem can be created by considering a temporal discontinuity in the properties of the medium. Let an unbounded medium (in space) undergo a sudden change in its permittivity at $t = 0$. The continuity of the electric field at $t = 0$ now requires that the phase factors of the wave existing before the discontinuity occurs, called a source wave, must match with phase factors, Ψ_N , of the newly created waves in the altered or switched medium, when $t = 0$ is substituted in the phase factors. This must be true for all values of z . Thus comes the requirement that k is conserved across a temporal discontinuity in a spatially invariant medium. Conservation of k is implemented by drawing a vertical line in the ω - k diagram as shown in the bottom part of the Figure 1(b). The ω values of the intersection points give the frequencies of the newly created waves [4]-[7].

3 Time-varying plasma medium

Any plasma is a mixture of charged particles and neutral particles. The mixture is characterized by two independent parameters for each of the particle species. These are the particle density N and the temperature T . There is a vast amount of literature on plasmas. A few references of direct interest to the reader of this chapter are provided here [8]-[10]. These deal with modeling of a magnetoplasma as an electromagnetic medium. The models are adequate in exploring some of the applications where the medium may be considered to have time-invariant electromagnetic parameters.

There are some applications in which the thermal effects are unimportant. Such a plasma is called a *cold plasma*. A *Lorentz plasma* [8] is a further simplification of the medium. It is assumed that the electrons interact with each other in a Lorentz plasma only through collective space-charge forces and that the heavy positive ions and neutral particles are at rest. The positive ions serve as a background that ensures the overall charge-neutrality of the mixture. In this chapter the Lorentz plasma is the dominant model used to explore the major effects of a nonperiodically time-varying electron density profile $N(t)$.

The constitutive relations for this simple model viewed as a dielectric medium are given by

$$\mathbf{D} = \varepsilon_0 \varepsilon_p \mathbf{E}, \quad (2)$$

where

$$\varepsilon_p = 1 - \frac{\omega_p^2}{\omega^2}, \quad (3)$$

and

$$\omega_p^2 = \frac{q^2 N}{m \epsilon_o}. \quad (4)$$

In these equations, q and m are the absolute values of the charge and mass of the electron, respectively, and ω_p^2 is the square of the plasma frequency proportional to the electron density.

A sketch of ϵ_p versus ω is given in Figure 2. The relative permittivity ϵ_p is real-valued only if the signal frequency is larger than the plasma frequency. Hence, ω_p is a cutoff frequency for the isotropic plasma. Above cutoff, a Lorentz plasma behaves as a dispersive dielectric medium with the relative permittivity lying between 0 and 1.

The relative permittivity of the medium can be changed by changing the electron density N , which in turn can be accomplished by changing the ionization level. The sudden change in the permittivity shown in Figure 1(b) is an idealization of a rapid ionization. Quantitatively, the sudden-change approximation can be used if the period of the source wave is much larger than the rise time of the temporal profile of the electron density. A step change in the profile is referred to in the literature as sudden creation [2], [5] or flash ionization [2].

Experimental realization of a small rise time is not easy. A large region of space has to be ionized uniformly at a given time. Joshi *et al.* [11], Kuo [12], Kuo and Ren [13], as well as Rader and Alexeff [14] developed ingenious experimental techniques to achieve these conditions and demonstrated the principle of frequency shifting using an isotropic plasma. One of the earliest pieces of experimental evidence of frequency-shifting quoted in the literature is a seminal paper by Yablonovitch [15]. Savage *et al.* [16] used the ionization front to upshift the frequencies. Ionization front is a moving boundary between unionized medium and the plasma [17]. Such a front can be created by a source of ionizing radiation pulse, say, a strong laser pulse. As the pulse travels in a neutral gas, it converts it into a plasma, thus creating a moving boundary between the plasma and the unionized medium. However, the ionization-front problem is somewhat different from the

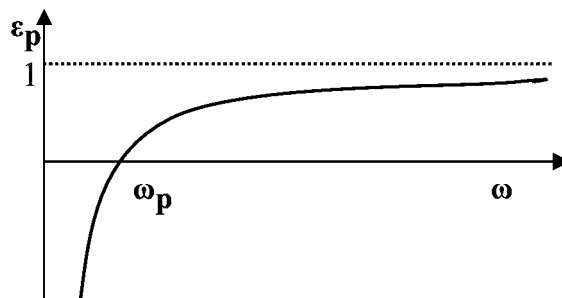


Figure 2 Relative permittivity ϵ_p vs. angular frequency ω for a plasma medium.

moving-plasma problem. In the front problem the boundary alone is moving and the plasma is not moving with the boundary.

The constitutive relation (2), based on the dielectric model of a plasma, does not explicitly involve the current density $\mathbf{J}$ in the plasma. The constitutive relations that involve the plasma current density $\mathbf{J}$ are given by

$$\mathbf{D} = \varepsilon_o \mathbf{E}, \quad (5)$$

$$\mathbf{J} = -qN\mathbf{v}. \quad (6)$$

The velocity $\mathbf{v}$ of the electrons is given by the force equation

$$m \frac{d\mathbf{v}}{dt} = -q\mathbf{E}. \quad (7)$$

In (7) the magnetic force due to the wave's magnetic field $\mathbf{H}$ is neglected, since it is much smaller [8] than the force due to the wave's electric field. The magnetic force term $(-q\mathbf{v} \times \mathbf{H})$ is nonlinear. Stanic [18] studied this problem as a weakly nonlinear system.

Since ion motion is neglected, (6) does not contain ion current. Such an approximation is called the radio approximation [9]. It is used in the study of radio wave propagation in the ionosphere. Low-frequency wave propagation studies take the ion motion into account [9].

Substituting (5), (6) and (7) in the Ampere–Maxwell equation

$$\nabla \times \mathbf{H} = \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t}, \quad (8)$$

we obtain

$$\nabla \times \mathbf{H} = j\omega\varepsilon_o\varepsilon_p(\omega)\mathbf{E}, \quad (9)$$

where $\varepsilon_p(\omega)$ is given by (3) and an $\exp(j\omega t)$ time-dependence has been assumed. For an arbitrary temporal profile of the electron density $N(t)$, (6) is not valid [2], [19], [20]. The electron density $N(t)$ increases because of the new electrons born at different times. The newly born electrons start with zero velocity and are subsequently accelerated by the fields. Thus, all the electrons do not have the same velocity at a given time during the creation of the plasma. Therefore,

$$\mathbf{J}(t) \neq -qN(t)\mathbf{v}(t), \quad (10)$$

but

$$\Delta \mathbf{J}(t) = -q\Delta N_i \mathbf{v}_i(t), \quad (11)$$

instead. Here ΔN_i is the electron density added at t_i , and $\mathbf{v}_i(t)$ is the velocity at time t of these ΔN_i electrons born at time t_i . Thus, $\mathbf{J}(t)$ is given by the integral

of (11) with respect to time and not by (10). The integral of (11), when differentiated with respect to t , gives the constitutive relation between $\mathbf{J}$ and $\mathbf{E}$ as follows:

$$\frac{d\mathbf{J}}{dt} = \varepsilon_o \omega_p^2(\mathbf{r}, t) \mathbf{E}(\mathbf{r}, t). \quad (12)$$

Equations (8), (12) and the Faraday equation

$$\nabla \times \mathbf{E} = -\mu_o \frac{\partial \mathbf{H}}{\partial t}, \quad (13)$$

are needed to describe the electromagnetics of isotropic plasmas.

Propagation of an electromagnetic wave traveling in the z -direction with $\mathbf{E} = \hat{\mathbf{x}}E$ and $\mathbf{H} = \hat{\mathbf{y}}H$ can be studied by assuming that the components of the field variables have harmonic space variation; i.e.,

$$F(z, t) = f(t)e^{-jkz}. \quad (14)$$

Substituting (14) in (8), (12) and (13), we obtain the wave equations

$$\frac{d^2 E}{dt^2} + [k^2 c^2 + \omega_p^2(t)] E = 0, \quad (15)$$

and

$$\frac{d^3 H}{dt^3} + [k^2 c^2 + \omega_p^2(t)] \frac{dH}{dt} = 0, \quad (16)$$

for E and H .

4 Sudden creation of an unbounded plasma medium

The geometry of the problem is shown in Figure 3. A plane wave of frequency ω_o is propagating in free space in the z -direction. Suddenly at $t = 0$, an unbounded

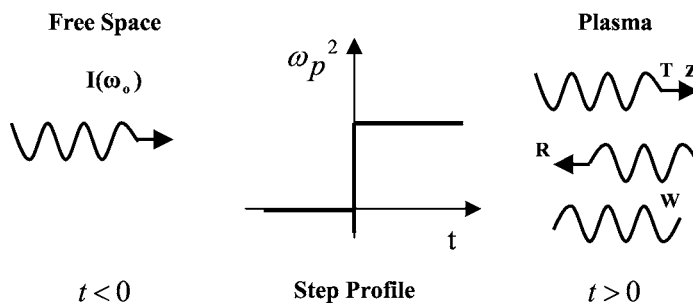


Figure 3 Suddenly created unbounded plasma medium.

plasma medium of plasma frequency ω_p is created. Thus arises a temporal discontinuity in the properties of the medium. The solution of (16), when ω_p is a constant, can be obtained as

$$H(t) = \sum_{m=1}^3 H_m \exp(j\omega_m t), \quad (17)$$

where

$$\omega_m [\omega_m^2 - (k^2 c^2 + \omega_p^2)] = 0. \quad (18)$$

The ω - k diagram [2] for the problem under discussion is obtained by graphing (18). Figure 4 shows the ω - k diagram, where the top and the bottom branches are due to the factor in the square brackets equated to zero, and the horizontal line is due to the factor $\omega = 0$. The line $k = \text{constant}$ is a vertical line that intersects the ω - k diagram at the three points marked as 1, 2, and 3. The third mode is the *wiggler* mode [2], [5], [21]. Its real-valued fields are

$$\mathbf{E}_3(x, y, z, t) = \mathbf{0}, \quad (19)$$

$$\mathbf{H}_3(x, y, z, t) = \hat{\mathbf{y}} H_o \frac{\omega_p^2}{\omega_o^2 + \omega_p^2} \cos(kz), \quad (20)$$

$$\mathbf{J}_3(x, y, z, t) = \hat{\mathbf{x}} H_o k \frac{\omega_p^2}{\omega_o^2 + \omega_p^2} \sin(kz). \quad (21)$$

It is of zero frequency but varies in space. (Such wiggler fields are used in a free electron laser (FEL) to generate coherent radiation [22].) Its electric field is zero but has a magnetic field due to the plasma current J_3 . In the presence of a static

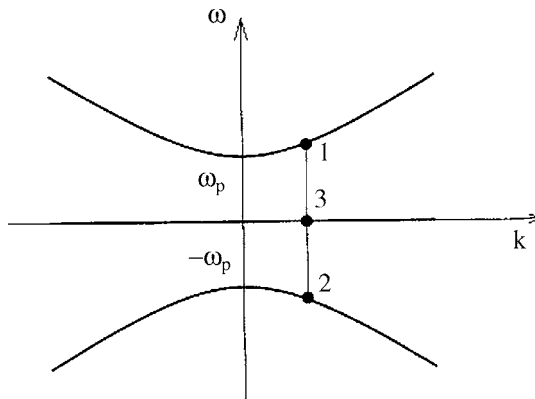


Figure 4 ω - k diagram and wiggler magnetic field.

magnetic field in the z -direction, the third mode becomes a traveling wave with a downshifted frequency. This aspect is discussed in Section 7.5.

The modes 1 and 2 have frequencies given by

$$\omega_{1,2} = \pm \sqrt{(\omega_p^2 + k^2 c^2)}, \quad (22)$$

where ω_2 has a negative value. Since the harmonic variations in space and time are expressed in the phase factor $\exp j(\omega t - kz)$, a negative value for ω gives rise to a wave propagating in the negative z -direction. It is a backward propagating wave, which for convenience can be referred to as a wave reflected by the discontinuity. The modes 1 and 2 have higher frequencies than the source wave. These are upshifted waves.

5 Switched plasma slab

The interaction of an electromagnetic wave with a plasma slab is experimentally more realizable than with an unbounded plasma medium. When an incident wave enters a pre-existing plasma slab, the wave must experience a spatial discontinuity. If the plasma frequency is lower than the incident wave frequency, then the incident wave is partially reflected and transmitted. When the plasma frequency is higher than that of the incident wave, the wave is totally reflected because the relative permittivity of the plasma is less than zero.

However, if the plasma slab is sufficiently thin, the wave can be transmitted due to a tunneling effect [2]. For this time-invariant plasma, the reflected and transmitted waves have the same frequency as the source wave frequency; and we refer to these as *A* waves. The wave inside the plasma has a different wavenumber but the same frequency due to the requirement of the boundary conditions.

When a source wave is propagating in free space and suddenly a plasma slab is created, the wave inside the slab region experiences a temporal discontinuity in the properties of the medium. Hence, the switching action generates new waves whose frequencies are upshifted, and then the waves propagate out of the slab. We call these waves *B* waves. The phenomenon is illustrated in Figure 5. In Figure 5(a) the source wave of frequency ω_o is propagating in free space. At $t = 0$, a slab of the plasma frequency ω_p is created. The *A* waves in Figure 5(b) have the same frequency as that of the source wave. The *B* waves are newly created waves due to the sudden switching of the plasma slab and have upshifted frequencies $\omega_1 = \sqrt{(\omega_o^2 + \omega_p^2)} = -\omega_2$. The negative value for the frequency of the second *B* wave shows that it is a backward propagating wave. These waves, however, have the same wavenumber as that of the source wave as long as they remain in the slab region. As the *B* waves come out of the slab, they encounter a spatial discontinuity, and therefore the wavenumber changes accordingly. The *B* waves are only created at the time of switching and hence exist for a finite time.

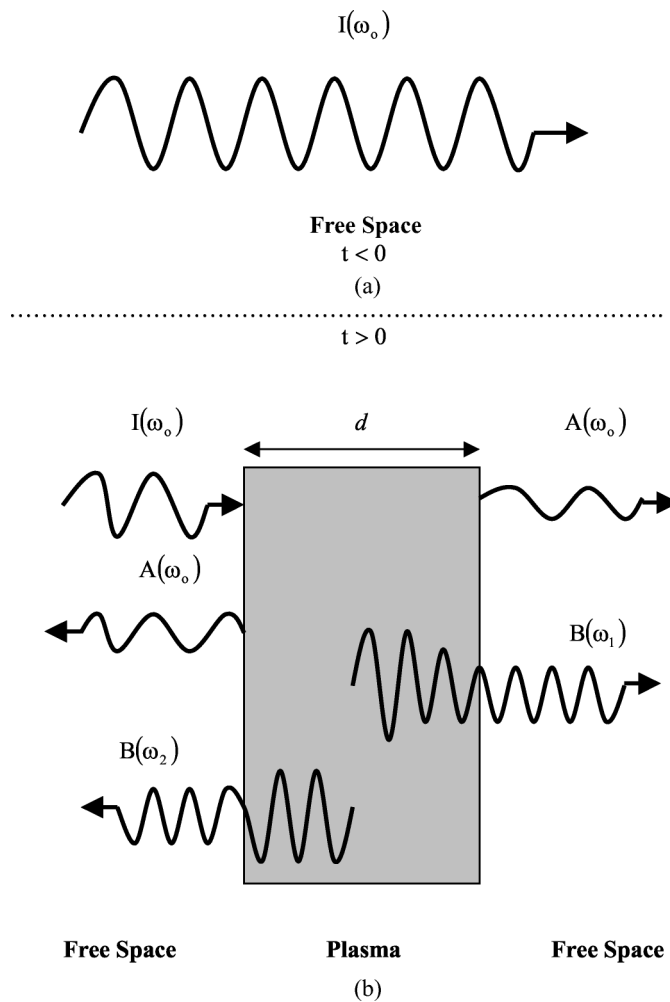


Figure 5 Effect of switching an isotropic plasma slab. *A* waves have the same frequency as the incident wave frequency (ω_o), and *B* waves have upshifted frequency $\omega_1 = \sqrt{(\omega_o^2 + \omega_p^2)} = -\omega_2$.

6 Applications

Many applications for frequency-shifting research are developing. An obvious application is for frequency transformers. The source wave can be generated in a frequency band using standard equipment, and the switched plasma device converts the source wave into a new wave in a frequency band not easily accessible by other methods. The frequency-shifting mechanism can be applied for plasma cloaking of satellites and aircrafts, and for producing short-chirped-pulses as ultrawide-band signals [23]. Application to photonics has been dealt in detail by Nerukh *et al.* [24].

7 Time-varying magnetoplasma medium

A plasma medium in the presence of a static magnetic field behaves like an anisotropic dielectric medium [1-2]. Therefore, the theory of electromagnetic wave propagation in this medium is similar to the theory of light waves in crystals. Of course, in addition, account has to be taken of the highly dispersive nature of the plasma medium.

A cold magnetoplasma is described by two parameters: the electron density N and the quasi-static magnetic field. The first parameter is usually given in terms of the plasma frequency ω_p . The strength and the direction of the quasi-static magnetic field have significant effect on the dielectric properties of the plasma. The parameter that is proportional to the static magnetic field is the electron gyrofrequency ω_b defined in Section 7.1. The cutoff frequency of the magnetoplasma is influenced by ω_p as well as ω_b . An additional important aspect of the dielectric properties of the magnetoplasma medium is the existence of a resonant frequency. At the resonant frequency, the relative permittivity ε_p goes to infinity. As an example, for longitudinal propagation defined in Section 7.2, resonance occurs when the signal frequency ω is equal to the electron gyrofrequency ω_b . For the frequency band $0 < \omega < \omega_b$, $\varepsilon_p > 1$ and can have very high values for certain combinations of ω , ω_p and ω_b ; for instance, when $f_p = \omega_p/(2\pi) = 10^{14}$ Hz, $f_b = \omega_b/(2\pi) = 10^{10}$ Hz, and $f = 10$ Hz, $\varepsilon_p = 9 \times 10^{16}$ [9]. A big change in ε_p can thus be obtained by collapsing the electron density, thereby converting the magnetoplasma medium into free space. A big change in ε_p can also be obtained by collapsing the background quasi-static magnetic field, thus converting the magnetoplasma medium into an isotropic plasma medium. These aspects are discussed in the remaining parts of this section.

7.1 Basic field equations

The electric field $\mathbf{E}(\mathbf{r}, t)$ and the magnetic field $\mathbf{H}(\mathbf{r}, t)$ satisfy the Maxwell curl equations:

$$\nabla \times \mathbf{E} = -\mu_0 \frac{\partial \mathbf{H}}{\partial t}, \quad (23)$$

$$\nabla \times \mathbf{H} = \varepsilon_0 \frac{\partial \mathbf{E}}{\partial t} + \mathbf{J}. \quad (24)$$

In the presence of a quasi-static magnetic field $\mathbf{B}_0$, the constitutive relation for the current density is given by

$$\frac{d\mathbf{J}}{dt} = \varepsilon_0 \omega_p^2(\mathbf{r}, t) \mathbf{E} - \mathbf{J} \times \boldsymbol{\omega}_b(\mathbf{r}, t), \quad (25)$$

where

$$\boldsymbol{\omega}_b = \frac{q\mathbf{B}_0}{m} = \omega_b \hat{\mathbf{B}}_0. \quad (26)$$

Therein, $\hat{\mathbf{B}}_0$ is a unit vector in the direction of the quasi-static magnetic field, ω_b is the gyrofrequency, and

$$\omega_p^2(\mathbf{r}, t) = \frac{q^2 N(\mathbf{r}, t)}{m \varepsilon_0}. \quad (27)$$

7.2 Characteristic waves

Next, the solution for a plane wave propagating in the z -direction in a homogeneous, time-invariant unbounded magnetoplasma medium can be obtained by assuming

$$f(z, t) = \exp[j(\omega t - kz)], \quad (28)$$

$$\omega_p^2(z, t) = \omega_p^2, \quad (29)$$

$$\omega_b(z, t) = \omega_b, \quad (30)$$

where f stands for the components of the field variables $\mathbf{E}$, $\mathbf{H}$, or $\mathbf{J}$.

The well-established *magnetoionic theory* [8]-[10] is concerned with the study of plane-wave propagation of an arbitrarily polarized plane wave in a cold anisotropic plasma, where the direction of phase propagation of the plane wave is at an arbitrary angle to the direction of the static magnetic field. As the plane wave travels in such a medium, the polarization state continuously changes. However, there are specific normal modes of propagation in which the state of polarization is unaltered. Plane waves with left (L wave) or right (R wave) circular polarization are the normal modes in the case of wave propagation along the quasi-static magnetic field. Such propagation is labeled as *longitudinal propagation*. The ordinary wave (O wave) and the extraordinary wave (X wave) are the normal modes for *transverse propagation*, where the direction of propagation is perpendicular to the static magnetic field. In this chapter, propagation of the R wave in a time-varying plasma is discussed. An analysis of the propagation of other characteristic waves can be found elsewhere [2].

7.3 R-wave propagation

The relative permittivity for R-wave propagation is [2], [8]

$$\varepsilon_{pR} = 1 - \frac{\omega_p^2}{\omega(\omega - \omega_b)} = \frac{(\omega + \omega_{c1})(\omega - \omega_{c2})}{\omega(\omega - \omega_b)}, \quad (31)$$

where ω_{c1} and ω_{c2} are the cutoff frequencies given by

$$\omega_{c1, c2} = \mp \frac{\omega_b}{2} + \sqrt{\left(\frac{\omega_b}{2}\right)^2 + \omega_p^2}. \quad (32)$$

Equation (31) is obtained by eliminating $\mathbf{J}$ from (24) with the help of (25) and recasting it in the form of (9).

The dispersion relation is obtained from

$$k_R^2 c^2 = \omega^2 \varepsilon_{pR} = \frac{\omega(\omega + \omega_{c1})(\omega - \omega_{c2})}{(\omega - \omega_b)}, \quad (33)$$

where k_R is the wavenumber for the R wave and c is the speed of light in free space. When expanded, this equation becomes

$$\omega^3 - \omega_b \omega^2 - (k_R^2 c^2 + \omega_p^2) \omega + k_R^2 c^2 \omega_b = 0. \quad (34)$$

Figure 6 shows a graph of ε_p versus ω , while Figure 7 gives the ω - k diagram for R -wave propagation. The R wave is a characteristic wave of longitudinal propagation. For this wave, the medium behaves like an isotropic plasma, except that ε_p is influenced by the strength of the quasi-static magnetic field. Particular attention is drawn to the specific feature, visible in Figure 6, showing $\varepsilon_p > 1$ for the

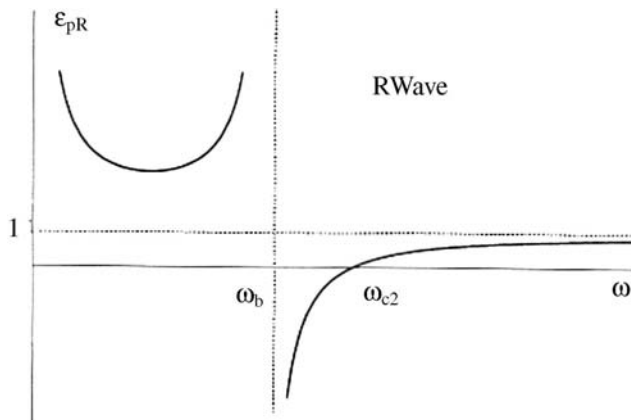


Figure 6 Relative permittivity for R -wave propagation.

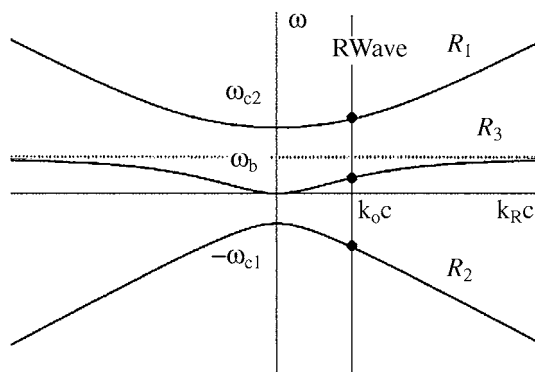


Figure 7 ω - k diagram for R -wave propagation.

R wave in the frequency band $0 < \omega < \omega_b$. This mode of propagation is called *whistler mode* in the literature on ionospheric physics and *helicon mode* in the literature on solid state plasmas. Sections 7.6 and 7.7 deal with the transformation of the whistler wave by a transient magnetoplasma medium and the consequences of such a transformation. An isotropic plasma medium does not support a whistler wave.

7.4 Sudden creation

In this section the problem of sudden creation of the plasma in the presence of a static magnetic field in z -direction is analyzed. The geometry of the problem is given in Figure 8. The source wave is assumed to be an R wave. The sudden creation is equivalent to creating a temporal discontinuity in the dielectric properties of the medium. In such a case, the wavenumber k_0 is conserved across the temporal discontinuity. For a given k_0 of the source wave, we draw a vertical line that intersects the branches in the ω - k diagram in Figure 7 at three points. The frequencies of these waves are different from the source frequency. The medium switching in this case creates three R waves labeled as R_1 , R_2 , and R_3 . Whereas R_1 and R_3 are transmitted waves, R_2 is a reflected wave.

A physical interpretation of the waves may be given in the following way. The electric and magnetic fields of the incident wave and the quasi static magnetic field accelerate the electrons in the newly created magnetoplasma, which in turn radiate new waves. The frequencies of the new waves and their fields can be obtained by adding contributions from the many electrons whose positions and motions are correlated by the collective effects supported by the magnetoplasma medium. Such a detailed calculation of the radiated fields seems to be quite involved. A simple but less accurate description of the plasma effect is obtained by modeling the magnetoplasma medium as a dielectric medium whose refractive index is computed through

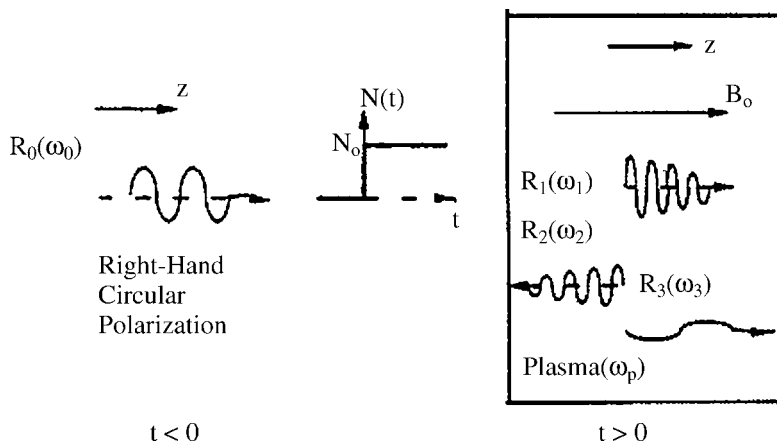


Figure 8 Effect of switching an unbounded magnetoplasma medium. Sketches of the B waves generated in the plasma are given for R -incidence.

magnetoionic theory [9]. The frequencies of the new waves are constrained by the requirements that the wavenumber k_0 is conserved over the temporal discontinuity, and the refractive index n is applicable to the type of wave propagation in the magnetoplasma. This gives a conservation law [25] $k_0 c = \omega_0 = n(\omega)$ from which ω may be determined. Solution of the associated electromagnetic initial value problem gives the electric and magnetic fields of the new waves.

7.5 Frequency-shifting characteristics of various R waves

The shift ratio and the efficiency of the frequency-shifting operation may be controlled by the parameters ω_p and ω_b . The results are presented by normalizing all frequency variables with respect to the source wave frequency ω_0 . This normalization is achieved by taking $\omega_0 = 1$ in numerical work.

For R waves, the curves of $\omega - \omega_p$ and $\omega - \omega_b$ are sketched in Figure 9(a) and Figure 9(b), respectively. In Figure 10, results are presented for the R_1 wave; the values on the vertical axis give the frequency-shift ratio since the frequency variables are normalized with respect to ω_0 . This is an upshifted wave, and the shift ratio increases with ω_p as well as ω_b . From Figure 10 it appears that by a suitable choice of ω_p and ω_b , one can obtain any desired large frequency shift. However, the wave generated may have weak fields associated with it, and the power density S_1 may be low. This point is illustrated in Table 1 by considering two sets of values for the parameters (ω_p, ω_b) . For the set $(0.5, 0.5)$ the shift ratio is 1.2, but

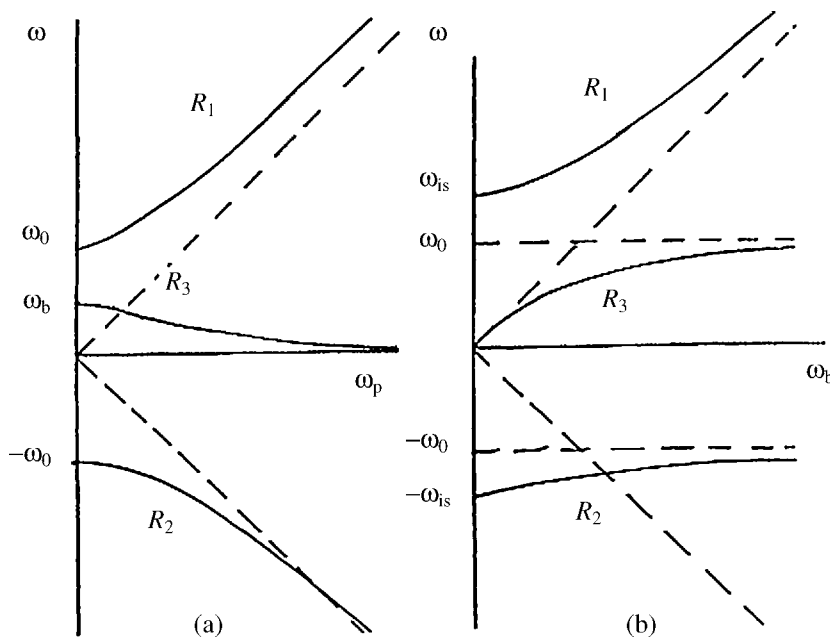


Figure 9 Frequency shifting of R waves. Sketches of (a) ω versus ω_p , (b) ω versus ω_b ; $\omega_{is} = (\omega_0^2 + \omega_p^2)^{1/2}$.

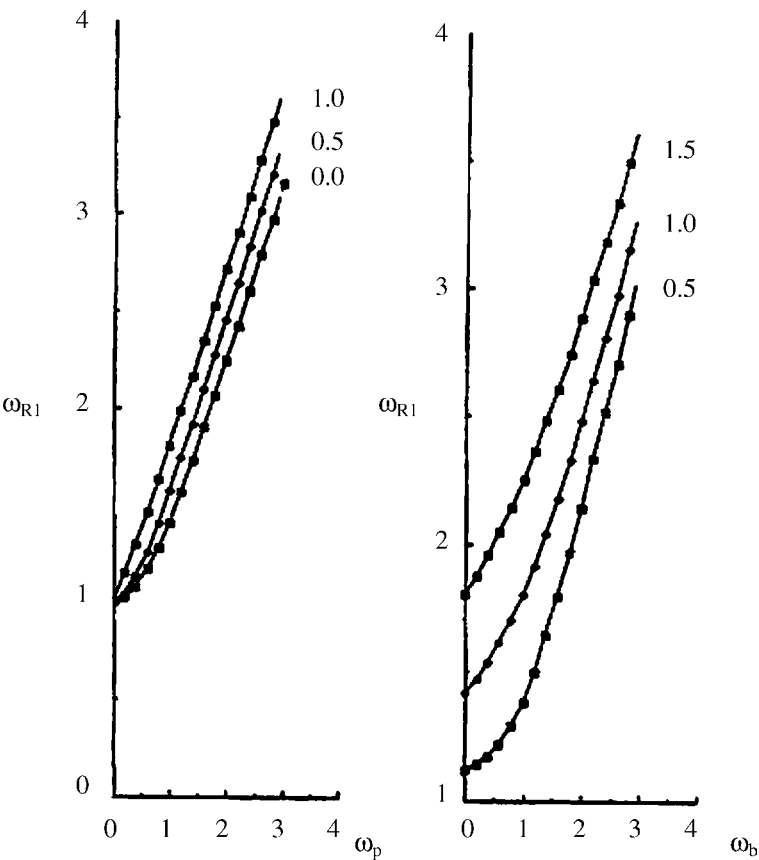


Figure 10 Frequency-shifting of the R_1 wave. The frequency variables are normalized with respect to the source wave frequency by taking $\omega_0 = 1$. Shown is the frequency shift ratio versus (a) ω_p and (b) ω_b . The numbers on the curves are (a) ω_b and (b) ω_p .

Table 1 R_1 -wave shift ratio and power density for two sets of (ω_p, ω_b) .

ω_0	ω_p	ω_b	E_1/E_o	H_1/H_o	S_1/S_o	ω_{R1}/ω_o
1	0.5	0.5	0.83	0.69	0.57	1.20
1	2.0	2.0	0.39	0.18	0.07	3.33

the power density ratio S_1/S_0 is 0.57; whereas for the set (2.0, 2.0) the shift ratio is 3.33, but the power density ratio is only 0.07. Similar remarks apply to other waves.

The R_2 wave is a reflected wave. This is an upshifted wave, and the shift ratio increases with ω_p but decreases with ω_b [2]. The R_3 wave in Figure 11 is a transmitted wave which is downshifted; the shift ratio decreases with ω_p and increases with ω_b . When $\omega_b = 0$, ω_{R3} becomes zero. The electric field E_3 also becomes zero, and the magnetic field degenerates to the wiggler magnetic field [26]. This result is in conformity with the result for the isotropic case discussed in Section 4.

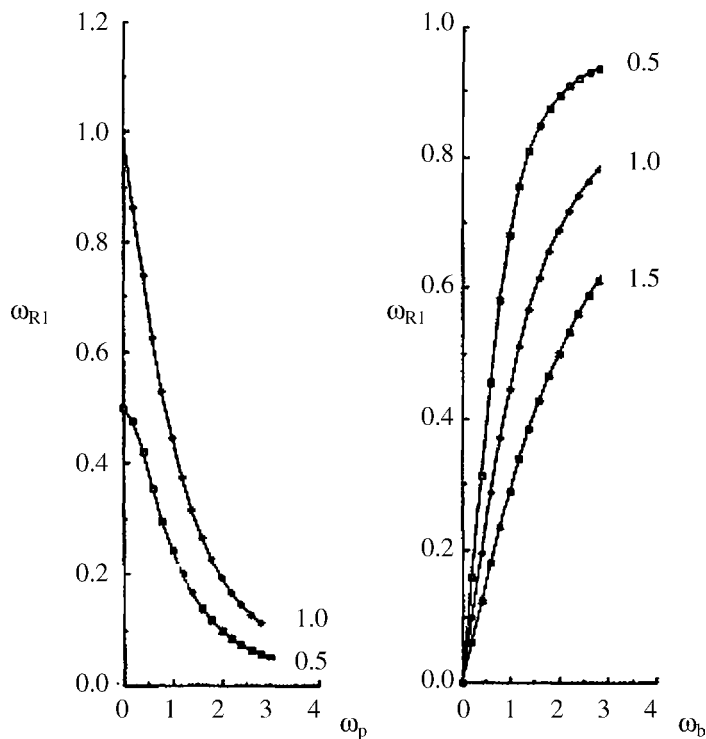


Figure 11 Frequency-shifting of the R_3 wave; $\omega_0 = 1$. Shown is the frequency shift ratio versus (a) ω_p (b) ω_b . The numbers on the curves are (a) ω_b (b) ω_p . Shown is the frequency shift ratio versus (a) ω_p and (b) ω_b . The numbers on the curves are (a) ω_b and (b) ω_p .

7.6 Frequency upshifting with power intensification

Suppose that a source wave is present in an unbounded magnetoplasma medium. The parameters are such that the refractive index is greater than 1. A wave propagating in the *whistler mode* is an example of such a wave. By collapsing the ionization, which can be achieved by removing the source of ionization, it is possible to obtain a frequency-upshifted wave with power intensification. The two limiting cases of (i) sudden collapse and (ii) slow decay are considered in this section. The starting point is the wave equation for the magnetic field of the R wave in a time-varying magnetoplasma medium obtained from (23), (24) and (25) as follows:

$$\frac{d^3 H}{dt^3} - j\omega_b(t) \frac{d^2 H}{dt^2} + [k_R^2 c^2 + \omega_p^2(t)] \frac{dH}{dt} - j\omega_b(t) k_R^2 c^2 H = 0. \quad (35)$$

The sudden collapse case is solved as an initial value problem and the slow decay case is solved using an adiabatic approximation [2]. In either case, the final upshifted frequency is the source frequency multiplied by the refractive index.

When the source frequency is much less than the electron gyrofrequency, the refractive index of the whistler wave n_w is quite large; and the electric field is

intensified by a factor of $n_w/2$ for the case of sudden collapse and by a factor of $n_w/\sqrt{2}$ for the case of slow decay. The corresponding intensification factors for the power density are $n_w/4$ and $n_w/2$, respectively. A physical explanation of the results, based on energy balance, is as follows. The energy in the whistler mode is predominantly the magnetic energy due to the plasma current. After the plasma collapses, the plasma current collapses and the magnetic energy due to plasma current is converted into wave electric and magnetic energies giving rise to the frequency-upshifted waves with enhanced electric fields and power density. Recent work by Bakov *et al.* [27] on energy relations in a general time-varying magnetoplasma can be applied to the study of frequency-shifts of the other modes.

Figure 12 illustrates the results for an exponential decay profile:

$$\omega_p^2(t) = \omega_{p0}^2 e^{-bt}. \quad (36)$$

All variables are normalized with reference to the source wave quantities. The source wave frequency $\omega_0 = 1$, so that ω_1 is the frequency upshift ratio. The parameters ω_b and ω_{p0} are assigned the values 100 and 150, respectively. A value of 0.01 is assigned to b , which therefore describes a slow decay, since $100/(2\pi)$ cycles of the source wave are accommodated in one time constant. The independent variable, time, is normalized with respect to the period of the source wave. Figure 12 also contains, for comparison, horizontal lines which are the results for sudden collapse. It is clear that frequency upshifting with power intensification occurs.

7.7 Generation of a controllable helical wiggler magnetic field

A plasma in the presence of a quasi-static magnetic field supports a whistler wave. When the static magnetic field is switched off, the energy of the whistler wave is converted into the energy of a helical wiggler magnetic field [2]. The analysis is based on the solution of (35) in which ω_p^2 is assumed to be constant but ω_b is a function of time.

Figure 13 illustrates the results for an exponential decay profile

$$\omega_b(t) = \omega_{b0} e^{-bt}. \quad (37)$$

All variables are normalized with frequency with reference to the source wave quantities. The parameters ω_{b0} and ω_p are assigned the values 100 and 1000, respectively. A value of 0.01 is assigned to b . Figure 13 shows also, for comparison, horizontal lines that are the results of sudden collapse. Irrespective of the rate of collapse of B_0 , a strong wiggler field is generated.

8 Conclusion

In this chapter, an introduction to the electromagnetics of the time-varying magnetoplasma medium is given by using idealized models. Mention has been made

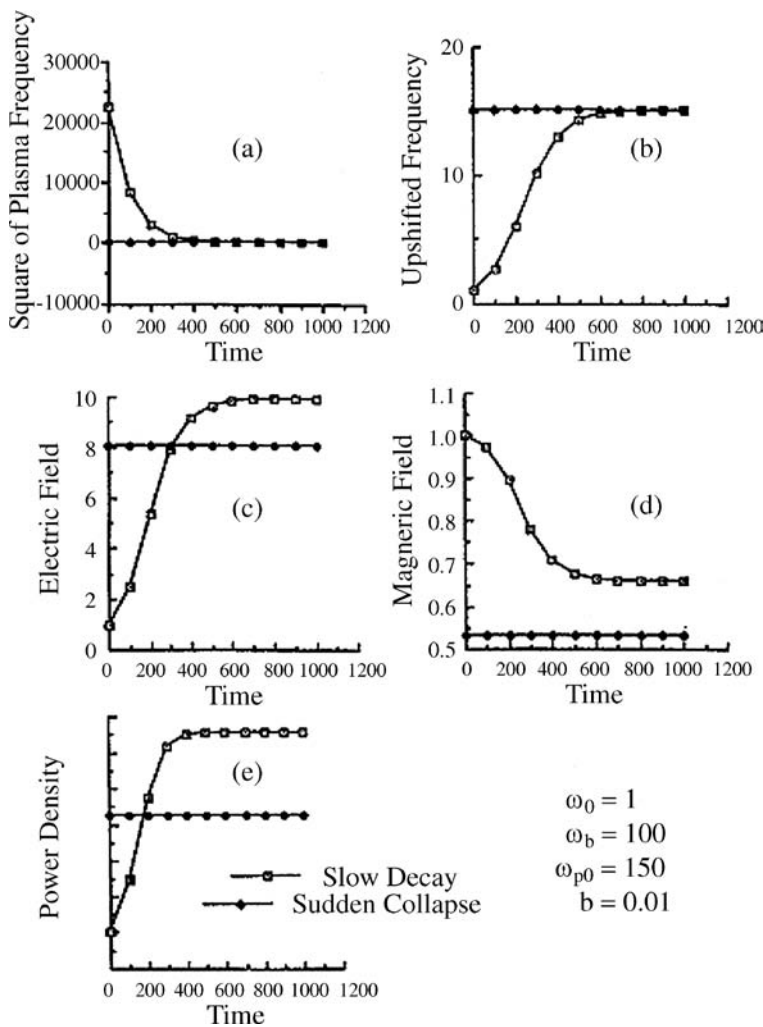


Figure 12 Frequency upshifting with power intensification of a whistler wave. The values on the vertical axes are normalized with respect to source wave quantities. The source frequency ω_o is taken as 1. The horizontal time scale is normalized with respect to the period of the source wave.

of the proof-of-the-principle experimental work performed by several groups [12]-[16], [28]. A parallel development in computer simulation using the FDTD method is also referenced [3]. Work is in progress at several places on the theory, computer simulation and experimental work to study the scalability of the results and the design of commercially viable systems for practical applications.

Acknowledgment

The author is thankful to his doctoral student Monzurul Ehsan for his help in preparing the material for this chapter.

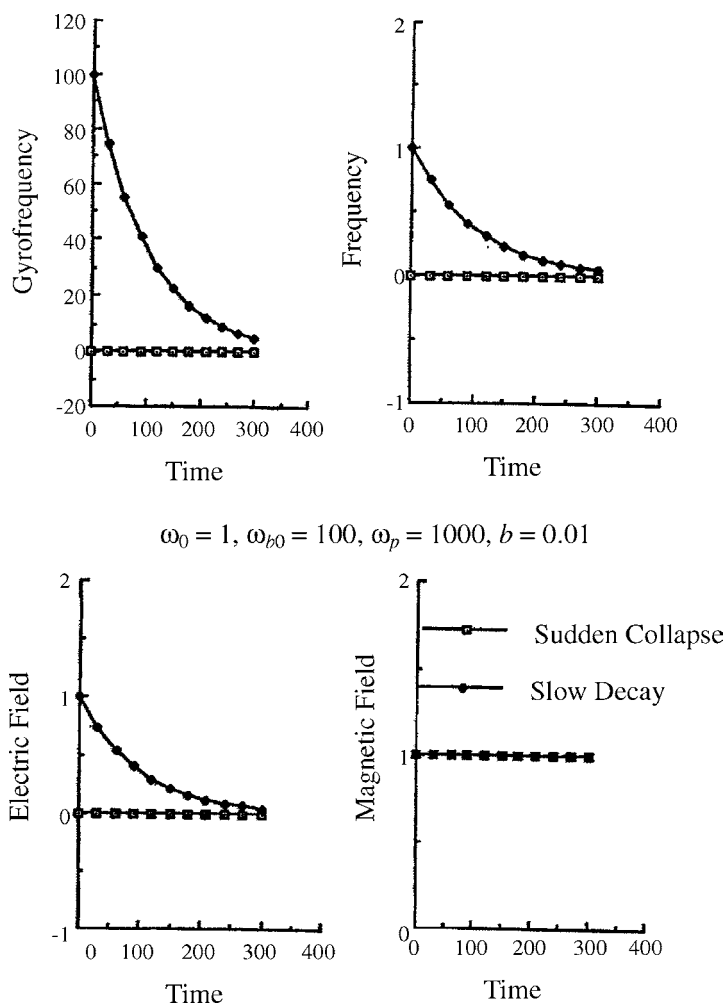


Figure 13 Conversion of a whistler wave into a helical magnetic wiggler field. The values on the vertical axes are normalized with respect to source wave quantities; $\omega_0 = 1$. The horizontal time scale is normalized with respect to the period of the source wave.

References

- [1] W.S. Weiglhofer, Constitutive characterization of simple and complex mediums, *Introduction to complex mediums for optics and electromagnetics* (W.S. Weiglhofer and A. Lakhtakia, eds), SPIE Press, Bellingham, WA, USA, 2003.
- [2] D.K. Kalluri, *Electromagnetics of complex media*, CRC Press, Boca Raton, FL, USA, 1999.
- [3] J.H. Lee and D.K. Kalluri, Three dimensional FDTD simulation of electromagnetic wave transformation in a dynamic inhomogeneous magnetized plasma, *IEEE Trans Antennas Propagat* **47** (1999), 1146–1151.

- [4] B.A. Auld, J.H. Collins and H.R. Zapp, Signal processing in a nonperiodically time-varying magnetoelastic medium, *Proc IEEE* **56** (1968), 258–272.
- [5] C.L. Jiang, Wave propagation and dipole radiation in a suddenly created plasma, *IEEE Trans Antennas Propagat* **23** (1975), 83–90.
- [6] B.L. Felsen and G.M. Whitman, Wave propagation in time-varying media, *IEEE Trans Antennas Propagat* **18** (1970), 242–253.
- [7] R.L. Fante, Transmission of electromagnetic waves into time-varying media, *IEEE Trans Antennas Propagat* **19** (1971), 417–424.
- [8] M.A. Heald and C.B. Wharton, *Plasma diagnostics with microwaves*, Wiley, New York, NY, USA, 1965.
- [9] H.G. Booker, *Cold plasma waves*, Kluwer, Higham, MA, USA, 1984.
- [10] B.S. Tannenbaum, *Plasma physics*, McGraw-Hill, New York, NY, USA, 1967.
- [11] C.J. Joshi, C.E. Clayton, K. Marsh, D.B. Hopkins, A. Sessler and D. Whittum, Demonstration of the frequency upshifting of microwave radiation by rapid plasma creation, *IEEE Trans Plasma Sci* **18** (1990), 814–818.
- [12] S.P. Kuo, Frequency up-conversion of microwave pulse in a rapidly growing plasma, *Phys Rev Lett* **65** (1990), 1000–1003.
- [13] S.P. Kuo and A. Ren, Experimental study of wave propagation through a rapidly created plasma, *IEEE Trans Plasma Sci* **21** (1993), 53–56.
- [14] M. Rader, F. Dyer, A. Matas and I. Alexeff, Plasma-induced frequency shifts in microwave beams, *Conf Rec Abstracts, IEEE Int Conf Plasma Sci*, Oakland, CA, 1990, p. 171.
- [15] E. Yablonovitch, Spectral broadening in the light transmitted through a rapidly growing plasma, *Phys Rev Lett* **31** (1973), 877–879.
- [16] R.L. Savage, Jr., C.J. Joshi and W.B. Mori, Frequency up-conversion of electromagnetic radiation upon transmission into an ionization front, *Phys Rev Lett* **68** (1992), 946–949.
- [17] M. Lampe and J.H. Walker, Interaction of electromagnetic waves with a moving ionization front, *Phys Fluids* **21** (1978), 42–54.
- [18] B. Stanic, P. Drljaca and B. Boskoic, Electron plasma waves generation in suddenly created isotropic plasma, *IEEE Trans Plasma Sci* **26** (1998), 1514–1519.
- [19] N.S. Stepanov, Dielectric constant of unsteady plasma, *Sov Radiophys Quant Electron* **19** (1976), 683–689.
- [20] A. Banos, Jr., W.B. Mori and J.M. Dawson, Computation of the electric and magnetic fields induced in a plasma created by ionization lasting a finite interval of time, *IEEE Trans Plasma Sci* **21** (1993), 57–69.
- [21] S.C. Wilks, J.M. Dawson and W.B. Mori, Frequency up-conversion of electromagnetic radiation with use of an overdense plasma, *Phys Rev Lett* **61** (1988), 337–340.
- [22] V.L. Granatstein and I. Alexeff (eds), *High-power microwave sources*, Artech House, Boston, MA, USA, 1987.

- [23] Special issue on generation of coherent radiation using plasmas, *IEEE Trans Plasma Sci* **21** (1993).
- [24] A.G. Nerukh, L.V. Scherbatko and M. Marciniak, *Electromagnetics of modulated media with applications to photonics*, IEEE/LEOS Poland chapter, Warsaw, 2001; printed by National Institute of Telecommunications, Dept. of Transmission and Fiber Technology, Warsaw, Poland.
- [25] D.K. Kalluri, Effect of switching a magnetoplasma medium on a traveling wave: conservation law for frequencies of newly created waves, *Conf Rec Abstracts, IEEE Int Conf Plasma Sci*, Oakland, CA, 1990, 129.
- [26] D.K. Kalluri, Effect of switching a magnetoplasma medium on a travelling wave: longitudinal propagation, *IEEE Trans Antennas Propagat* **37** (1989), 1638–1642.
- [27] M.I. Bakunov and I.S. Grachev, Energy relations for electromagnetic waves in a time-varying magnetoplasma medium, *Complex mediums II: Beyond linear isotropic dielectrics* (A. Lakhtakia, W.S. Weiglhofer and I.J. Hodgkinson, eds), Proc SPIE **4467** (2001), Bellingham, WA, USA, 78–86.
- [28] S.P. Kuo, D. Bivolaru, L. Orlick, I. Alexeff and D.K. Kalluri, A transmission line filled with fast switched periodic plasma as a wideband frequency transformer, *IEEE Trans Plasma Sci* **29** (2001), 365–370.



Dikshitulu K. Kalluri is Professor of Electrical and Computer Engineering at the University of Massachusetts Lowell. Born in Chodavaram, India, he received B.E. in electrical engineering from Andhra University, India; D.I.I. Sc. in high-voltage engineering from the Indian Institute of Science in Bangalore, India; M.S. in electrical engineering from the University of Wisconsin, Madison; and Ph.D. in electrical engineering from the University of Kansas, Lawrence. Dr. Kalluri began his career at the Birla Institute, Ranchi, India, advancing to the rank of Professor, then heading the Electrical Engineering Department, and finally serving as (Dean) Assistant Director of the institute. He collaborated with research groups at the Lawrence Berkeley Laboratory, University of California at Los Angeles, University of Southern California, Polytechnic University of New York, and the University of Tennessee, and worked several summers as a faculty research associate at Air Force Laboratories. Since 1984, he has been with the University of Massachusetts Lowell, where he is coordinator of the doctoral program and co-director of the Center for Electromagnetic Materials and Optical Systems (CEMOS). As part of the center, he recently established the Electromagnetics and Complex Media Research Laboratory. A fellow of the Institute of Electronic and Telecommunication Engineers and a member of Eta Kappa Nu and Sigma Xi, he has published many research papers, technical articles and reviews.

Magnetoimpedance in Multilayered Films for Miniature Magnetic Sensors

Larissa V. Panina and Dmitriy P. Makhnovskiy

Abstract

The discovery of the magnetoimpedance (MI) effect in 1994 had a strong impact on the development of micromagnetic sensors operating in the nano-tesla range. In certain soft magnetic materials, such as composites of amorphous thin wires, the impedance change (MI ratio) is in the range of 50–100% in the megahertz frequency band for external magnetic fields of a few oersted. However, when decreasing the size of the sensor element, the maintenance of such high sensitivity becomes a major concern. Special thin-film structures have been proposed and fabricated to improve the MI performance in miniature elements. This chapter concerns the principal advantages of MI in magnetic/metallic multilayered materials and their applications to magnetic sensing technology. Physical concepts, theoretical analysis based on the field-dependent surface impedance matrix, experimental results, and sensor designs are discussed. These encompass multifold enhancement of the MI ratio, a considerable extension of the operational frequency range, and the field symmetry of the MI behavior. Layered systems allow special types of magnetic anisotropy (either transverse or crossed) to be established in order to realize antisymmetric and asymmetric MI effects. This property is of a particular interest for magnetic-sensor applications.

1 Introduction

Miniature magnetic-sensor technology is a rapidly growing industry. Sensitive and rapid-response sensors are in great demand for advanced intelligent measurement and control systems. This demand has been satisfied in some areas by magneto-resistance (MR), giant magneto-resistance (GMR), fluxgate and other technologies [1]-[3]. A relatively new magnetic sensor technology based on the magnetoimpedance (MI) effect has the potential to overtake some of the other emergent sensor systems in term of performance and low cost [4]-[6]. MI sensors combine such desirable characteristics as high sensitivity, miniature size, fast response, and low power consumption, which make them unique in the magnetic sensor family.

The MI sensing performance is based on enormous changes in the complex-valued impedance at high frequencies under the application of a dc (or low-frequency) magnetic field [7]-[12]. The ratio of the percent change of the impedance, called the MI ratio, reaches a value of up to 150% at megahertz frequencies in amorphous microwires having circumferential (or helical) anisotropy. The characteristic magnetic fields required to cause this impedance change are in the range of 1–5 Oe. Then, MI sensitivity is at least one order of magnitude higher than that of GMR materials. However, the MI wires have much greater dimensions: 10–50 μm in diameter and a few millimeters in length. With decreasing sensor-element size, the maintenance of ultrahigh sensitivity becomes a problem. Special thin-film structures are employed to improve the MI performance in miniature elements (thickness < 1 μm , in-plane size < 200 μm and length < 5000 μm) [14]-[17].

This chapter deals with MI in sandwich-type structures having two outer magnetic layers and an inner conductive lead. The MI ratio in these systems has been proven to be much greater than in a similar ferromagnetic single layer. For example, in CoFeSiB/Cu/CoFeSiB multilayers of several micrometers thickness the impedance change is more than 300% at 10 MHz under the application of a dc magnetic field of about 10 Oe [14]. In contrast, a single CoFeSiB layer under the same conditions exhibits a change of just a few percent.

In electrically uniform materials subjected to a dc magnetic field, MI originates from the dependence of the skin depth on the transverse permeability. In multilayer structures having a highly conductive inner film, a very large change in impedance can occur at quite low frequencies when the inductance caused by the outer magnetic layers becomes sufficiently high. This results in multifold enhancement of the MI ratio and a considerable extension of the operational frequency range. Furthermore, special types of magnetic anisotropy can be realized in layered systems [18]-[20]. Multilayers with a transverse anisotropy produce antisymmetric response manifested as off-diagonal terms in the surface impedance tensor. Cross-anisotropy films exhibit asymmetric MI characteristics. These features are crucially important to obtain a linear voltage response with enhanced sensitivity, without requiring large power consumption, and are of particular interest for magnetic-sensor applications.

2 Analysis of MI in multilayer structures

In this section, we discuss a symmetrical three-layer film model that reflects all the essential features of MI behavior in multilayered materials. We start our analysis by considering the impedance Z of this structure in a usual way as the ratio $Z = V_f/i$, where V_f is the voltage induced in the film by a passing current i . However, in a magnetic system the induced voltage can be of a more complicated origin. For example, the voltage V_f can be induced by the current i_c flowing in the coil attached to the film. Alternatively, the current i is capable of generating the coil voltage V_c . To describe the generalized voltage (V_f , V_c) as response to excitation (i , i_c) it is convenient to introduce the surface impedance tensor $\hat{\zeta}$. We calculate the tensor $\hat{\zeta}$ in the low-frequency approximation, which neglects the skin effect. This approach is reasonable for not very thick films ($<1 \mu\text{m}$) and for not very high frequencies ($<1 \text{ GHz}$), which covers a wide range of applications. Nevertheless, the approach to the exact solution is considered as well. Finally, the effect of the film in-plane size on MI is analyzed. Gaussian units are used since they are more convenient in electromagnetics.

The structure consists of an inner conductive lead (M) of a thickness $2d_1$ and two identical magnetic layers (F) of a thickness d_2 along the x axis. The layers are parallel to the yz plane having a width b (y axis) and length l (z axis). An external dc magnetic field H_{ex} (which is to be sensed) is applied parallel to the z axis. In general, a dc bias current I_b can also be used to modify the static magnetic structure. This model is depicted in Figure 1.

Although the magnetic layers are made of the same alloy, they can be produced with different magnetic anisotropy. In particular, we are interested in the case when the anisotropy axes are directed at an angle $\pm\alpha$ to the z axis, respectively for the upper ($x > 0$) and lower ($x < 0$) magnetic layers. Such anisotropy can be induced, for example, by current annealing in the presence of a longitudinal field. The combination of anisotropy, H_{ex} and I_b , which induces a circulatory magnetic field $H_b(x > 0) = -H_b(x < 0)$, results in the antisymmetric arrangement of dc magnetization $\mathbf{M}_0$ such that $M_{0y}(x > 0) = -M_{0y}(x < 0)$.

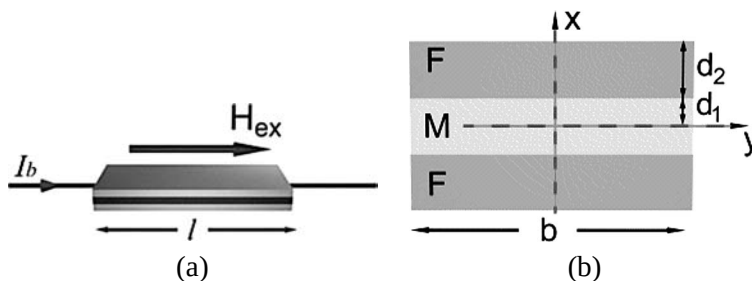


Figure 1 Schematic drawing of three-layer MI film showing principal directions and quantities used. (a) General view, (b) cross-section of the film.

2.1 Impedance of symmetrical three-layer film

Our first goal is to calculate the impedance of the three-layer structure shown in Figure 1 when it is excited by an alternating current $i = i_0 \exp(-j\omega t)$ flowing along the z axis. The layer thicknesses d_1 and d_2 are such that the skin effect can be neglected. If the inner lead conductivity σ_1 is considerably higher than the conductivity σ_2 of the magnetic layers (to be exact, $\sigma_1 d_1 \gg \sigma_2 d_2$), the current mainly flows along the conductive lead. With these assumptions, the expression for the impedance can be written in the form

$$Z = R_m - j\omega\Phi/ci, \quad (1)$$

where $R_m = l/2\sigma_1 d_1 b$ is the resistance of the inner conductor, Φ is the total transverse magnetic flux generated by the driving current i in the magnetic layers, and c is the velocity of light in vacuum. If the edges effects are neglected (i.e., the film is treated as infinitely long in the y and z directions), the ac magnetic field $\mathbf{h}$ in the magnetic layers is uniform. Then,

$$\Phi = (\hat{\boldsymbol{\mu}}\mathbf{h})_y d_2 l, \quad (2)$$

where $\hat{\boldsymbol{\mu}}$ is the ac permeability tensor. In general, the field $\mathbf{h}$ has y - and x -components due to the tensor form of the permeability. The value of the y -component is related to the current as $|h_y| = 2\pi i/cb$, and the relation between h_y and h_x is determined from the condition that the normal magnetic flux is zero [i.e., $(\hat{\boldsymbol{\mu}}\mathbf{h})_x = 0$]; accordingly,

$$\begin{aligned} \Phi &= \tilde{\mu}_{yy} h_y d_2 l = \tilde{\mu}_{yy} d_2 \frac{2\pi i l}{cb}, \\ \tilde{\mu}_{yy} &= \mu_{yy} + \mu_{xy}^2 / \mu_{xx}. \end{aligned} \quad (3)$$

After using (3), the impedance can be written in the form

$$Z = R_m \left(1 - 2j\tilde{\mu}_{yy} \frac{d_1 d_2}{\delta_1^2} \right), \quad (4)$$

where $\delta_1 = c/\sqrt{2\pi\sigma_1\omega}$ is the skin depth in the metallic inner lead. Expression (4) shows that the MI ratio in the sandwich film can be very large even at relatively low frequencies when the skin effect is not substantial, and has a linear dependence on $\tilde{\mu}_{yy}$.

This conclusion can be illustrated as follows. At a frequency of 10 MHz, taking $d_1 = d_2 = 0.5 \mu\text{m}$ and $\sigma_1 = 2 \times 10^{18} \text{s}^{-1}$ (conductivity of copper), we get $d_1/\delta_1 = d_2/\delta_1 = 0.045$. A typical low-frequency change in the permeability $\tilde{\mu}_{yy}$ (having a rotational mechanism) under the application of $H_{ex} \cong H_K$ (where H_K is the

effective anisotropy field) is from 1 to 500; then, the impedance varies over 200% according to (4).

The impedance for the present geometry can be calculated exactly for any frequency, without neglecting the skin effect [18]-[21]. For a film of $1\ \mu\text{m}$ thickness with $\sigma_1 d_1 / \sigma_2 d_2 \approx 50$ (as for a CoSiB/Cu/CoSiB system with amorphous magnetic films), the approximate solution almost coincides with the exact one up to a frequency of 1 GHz.

2.2 Surface impedance tensor

For the considered configuration, the ac current induces both the voltage V_f between the film ends and the coil voltage V_c , as shown in Figures 2(a) and 2(b), since the current flow gives rise to an antisymmetric transverse magnetization (or circulatory magnetization) and a nonzero longitudinal magnetization.

If the film is placed in a variable longitudinal field h_{ex} induced by the coil current i_c , not only the longitudinal magnetization is induced, but also the circulatory magnetization contributing to V_f , as shown in Figure 2(c). A mixed excitation by both the current i in the film and current i_c in the coil can be used. The crossed magnetization processes related to the voltages V_c and V_f are known as the inverse Wiedemann and the Matteucci effects [6]. With increasing frequency, the induced voltages become of the same order and very sensitive to the longitudinal field H_{ex} .

The induced voltages are convenient to describe in terms of the surface impedance tensor $\hat{\zeta}$ [13], [18], [20] that relates the variable electric $\mathbf{e}_t$ and magnetic $\mathbf{h}_t$ fields taken on the external surfaces $x = \pm d$ ($d = d_1 + d_2$). Thus,

$$e_{t\alpha} = \zeta_{\alpha\beta} (\mathbf{h}_t \times \mathbf{n})_\beta, \quad \alpha, \beta = z, y, \quad (5)$$

where $\mathbf{n}$ is a unit vector directed inside the film, while $\mathbf{e}_t$ and $\mathbf{h}_t$ are tangential to the surface. The voltage V_f is determined by the surface value of the longitudinal electric field $e_z(d) = e_z(-d)$ and is related to the fields h_y and $h_z = h_{ex}$ via the components ζ_{zz} and ζ_{zy} as follows:

$$V_f = (\zeta_{zz} h_y - \zeta_{zy} h_{ex}) l, \quad (6)$$

$$h_y = (2\pi / cb) i, \quad h_{ex} = 4\pi n_1 i_c / c, \quad (7)$$

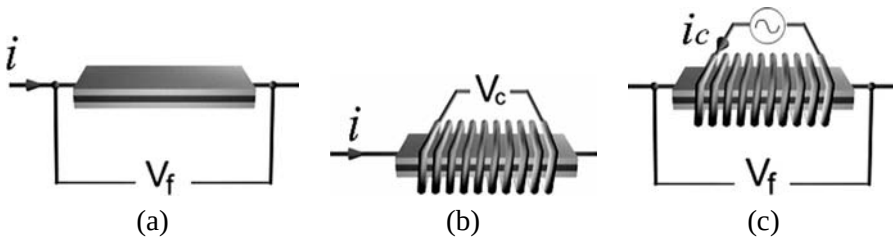


Figure 2 Methods of ac voltage excitation.

where n_1 is the number of turns per unit length in the excitation coil. The coil voltage V_c is associated with the circulatory electric field $e_y(d) = -e_y(-d)$ and can be expressed as

$$V_c = 2bN_2(\varsigma_{yz}h_y - \varsigma_{yy}h_{ex}), \quad (8)$$

where N_2 is the number of turns in the detection coil. In a general case, two coils can be used: the excitation coil for producing h_{ex} and the pick-up coil for detection of V_c . Equations (6) and (8) can be written in terms of currents i and i_c :

$$\begin{pmatrix} V_f/l \\ V_c/2bN_2 \end{pmatrix} = \begin{pmatrix} \varsigma_{zz} & -\varsigma_{zy} \\ \varsigma_{yz} & -\varsigma_{yy} \end{pmatrix} \begin{pmatrix} Ai \\ Bi_c \end{pmatrix}, \quad (9)$$

where the parameters $A = h_y/i$ and $B = h_{ex}/i_c$ are given by (7).

The surface impedance tensor can be easily found in the low-frequency approximation. Here we are interested in analyzing ς_{zz} and $\varsigma_{yz} = \varsigma_{zy}$ components. The parameter ς_{zz} is obtained by comparing (4) and (9) wherein $i_c = 0$:

$$\varsigma_{zz} = Z/A = \varsigma_0 \left(1 - 2j\tilde{\mu}_{yy} \frac{d_1 d_2}{\delta_1^2} \right), \quad \varsigma_0 = \frac{c}{4\pi\sigma_1 d_1}. \quad (10)$$

The off-diagonal component, say ς_{yz} , can be found by considering the coil voltage V_c induced by the film current i . This voltage is related to the longitudinal magnetic flux:

$$V_c = -j\omega N_2(2d_2 b)(\hat{\mu}h)_z. \quad (11)$$

When comparing (11) and (9) taken with $i_c = 0$ one can obtain

$$\begin{aligned} \varsigma_{yz} &= V_c/2bN_2 Ai = -j\omega d_2 \tilde{\mu}_{zy}/c, \\ \tilde{\mu}_{zy} &= \mu_{zy} - \mu_{zx}\mu_{xy}/\mu_{xx}. \end{aligned} \quad (12)$$

2.3 Exact solution for the surface impedance tensor

The calculation of $\hat{\varsigma}$ valid for any frequency is based on the solution of the Maxwell equations for the fields $\mathbf{e}$ and $\mathbf{h}$, together with the equation of motion for the magnetization vector $\mathbf{M}$. The theoretical aspects of the problem can be very complicated and generally require approximations. Typically, a linear approximation with respect to the time-variable parameters $\mathbf{e}$, $\mathbf{h}$ and $\mathbf{m} = \mathbf{M} - \mathbf{M}_0$ (where $\mathbf{M}_0$ is the static magnetization) is used. On assuming a local relationship between $\mathbf{m}$ and $\mathbf{h}$ ($\mathbf{m} = \hat{\chi}\mathbf{h}$), the problem is simplified to finding the solutions of Maxwell's equations with a given ac permeability tensor $\hat{\mu} = 1 + 4\pi\hat{\chi}$, which corresponds to neglecting the exchange effects.

Further assumptions about $\hat{\mu}$ are needed. The permeability depends on many factors, including the domain configuration, anisotropy and stress distribution, and the mode of magnetization (domain wall motion or magnetization rotation). These factors can be very complicated in real materials, making modeling very difficult. The problem is simplified if the domain structure is not considered (in practice, it can be eliminated by a proper dc bias) and if $\mathbf{M}_0$ is constant in the films. Then, $\hat{\mu}$ is determined by the magnetic moment rotation and is independent of the position. Furthermore, the tensor $\hat{\mu}$ has a general form that reduces to a quasi-diagonal form in the primed coordinate system with the axis $z' \parallel \mathbf{M}_0$, shown in Figure 3.

In the case of a uniform precession of the total magnetization vector $\mathbf{M}$ around $\mathbf{M}_0$, the susceptibility tensor in the primed coordinates (x, y', z') is of the form

$$\hat{\chi} = \begin{pmatrix} \chi_1 & -j\chi_a & 0 \\ j\chi_a & \chi_2 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \quad (13)$$

This form can be easily obtained from the linearized Landau–Lifshitz equation describing the magnetization dynamics. The expressions for χ_1 , χ_2 , χ_a depend on the specific magnetic configuration [19].

Here we are restricted to the consideration of magnetization rotation dynamics only, assuming that the domain wall displacements are strongly damped. Then, (13) determines the permeability tensor $\hat{\mu} = 1 + 4\pi \hat{\chi}$. To calculate the normalized parameters $\tilde{\mu}_{yy}$ and $\tilde{\mu}_{zy}$, which appear in expressions (10) and (12) for the impedance components, the tensor $\hat{\mu}$ has to be written in the original coordinates x, y, z , which yields

$$\tilde{\mu}_{yy} = 1 + 4\pi \chi \cos^2 \theta, \quad (14)$$

$$\tilde{\mu}_{zy} = -4\pi \chi \sin \theta \cos \theta, \quad (15)$$

$$\chi = \chi_2 - \frac{4\pi \chi_a^2}{1 + 4\pi \chi_1},$$

where θ is the angle between the dc magnetization $\mathbf{M}_0$ and the z axis.

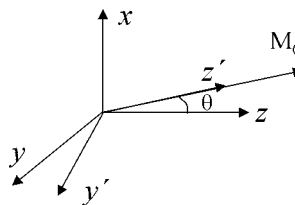


Figure 3 Original xyz and primed $xy'z'$ coordinate systems. Using primed coordinates reduces the permeability tensor to a quasi-diagonal form.

2.4 MI in a narrow sandwich (width effect)

For practical device use, the effect of in-plane sandwich width on MI has to be studied. If the edge effect is neglected (approximation of an infinite width), the magnetic flux generated by the current flowing along the inner lead is confined within the outer magnetic layers. In a sandwich of finite width, the flux leaks across the inner conductor [21–22], [24]. This process eventually results in the considerable drop in MI ratio, if the film width is smaller than some critical value depending on the transverse permeability and the thicknesses of the magnetic and conductive layers. This process is similar to that resulting in a drop in the efficiency of inductive recording heads [25–26].

The effect of the flux leakage across the conductive layer is found to be described by an effective “ac demagnetizing” factor g that reduces the permeability and makes it less sensitive to the dc field H_{ex} when the film width b is decreased. In the low-frequency approximation, when the skin effect is neglected, the parameter g is given by

$$g(k) = \frac{1}{(k_1\lambda)^2} \left[1 - (k\lambda)^2 \frac{kb \sinh kb}{kb \sinh kb + 2(k_1\lambda)^2(1 - \cosh kb)} \right], \quad (16)$$

$$k^2 = 1/\lambda^2 - k_1^2, \quad \lambda^2 = d_1 d_2 \tilde{\mu}_{yy}, \quad k_1 = (1 + j)/\delta_1.$$

For $kb \gg 1$ (or $b/\lambda \gg 1$) the function g tends to be unity. It means that the parameter $b^* = 1/k$ plays the role of a critical width: for $b < b^*$ all the film dimensions b , d_1 and d_2 influence the value of the impedance. In the low-frequency limit, $b^* = \sqrt{d_1 d_2 \tilde{\mu}_{yy}}$. Typical parameters for the structures of interest are $d_1 \approx d_2 \approx 0.1\text{--}0.5 \mu\text{m}$, $2b \approx 10\text{--}50 \mu\text{m}$ and $\mu_t \approx 10^3$, for which $b^* \approx 3\text{--}15 \mu\text{m}$ is comparable to the half-width, which means that the size effects cannot be neglected.

Figure 4 shows the plots of the impedance change vs. frequency, with the film width b as a parameter, for magnetic films with transverse anisotropy [21].

It is convenient to define the impedance change ratio as $\Delta Z/Z = |Z(H_K) - Z(0)|/|Z(0)| \times 100\%$, where H_K is the anisotropy field (for transverse anisotropy and $I_b = 0$, the magnetic layers are magnetized along the z axis when $H_{ex} = H_K$). The function $|Z(H_{ex})|$ has a maximum at $H_{ex} \approx H_K$, associated with that for the rotational transverse permeability $\tilde{\mu}_{yy}$. Therefore, the introduced parameter gives the maximum impedance change. The magnetic and electric parameters taken for the calculations correspond to amorphous CoFeSiB/Cu/CoFeSiB sputtered multilayers [14–15]. For a wide film ($b > 100 \mu\text{m}$ for $d = 0.5 \mu\text{m}$ or $b > 10 \mu\text{m}$ for $d = 0.1 \mu\text{m}$), the results are very close to those obtained for an infinite in-plane film. With decreasing b , the impedance change ratio decreases substantially: for example, for $100 \mu\text{m}$ -wide films, $\Delta Z/Z$ reaches more than 300% at a frequency of 150 MHz, whereas its value is only about 70% for $b = 10 \mu\text{m}$ at this frequency. The decrease in MI is stronger at lower frequencies where the critical width b^* is larger. In the foregoing example, $\Delta Z/Z$ recovers up to 200% at a frequency of 900 MHz. As a result, the impedance plots do not show flat regions for small values of b .

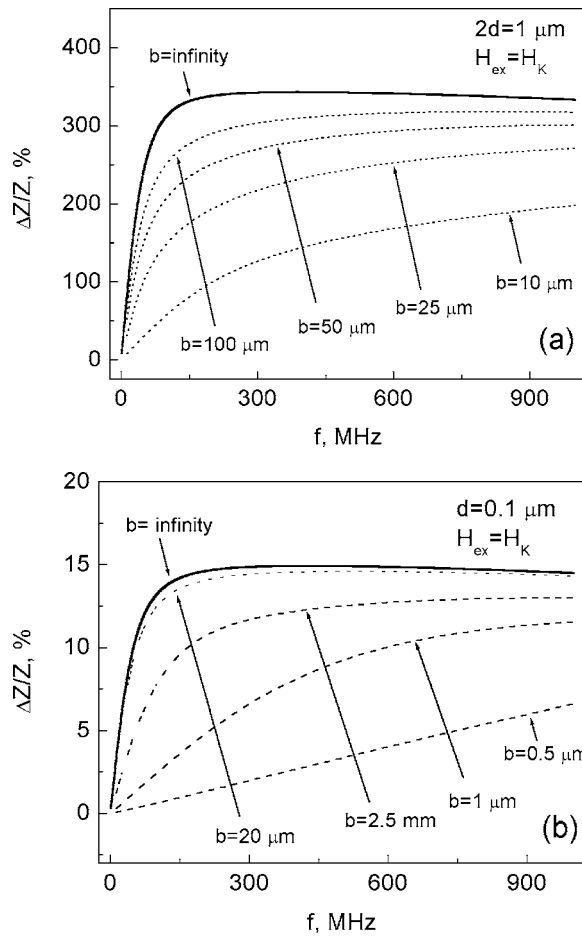


Figure 4 Maximum impedance change ratio $\Delta Z/Z = |[Z(H_K) - Z(0)]/Z(0)|$ in a sandwich structure with transverse anisotropy as a function of frequency, with the width b as a parameter. $d_1 = d_2 = 0.25 \mu\text{m}$ in (a) and $d_1 = d_2 = 0.05 \mu\text{m}$ in (b).

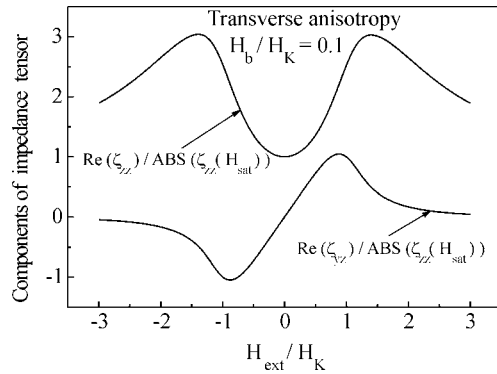
3 Asymmetric magnetoimpedance (AMI)

The field behavior of the surface impedance tensor (and voltage response) is associated with that of the permeability tensor. From (14) and (15) it follows that the permeability parameters $\tilde{\mu}_{yy}$ and $\tilde{\mu}_{zy}$ depend on the direction of the static magnetization $\mathbf{M}_0$. Therefore, in structures with an asymmetric arrangement of $\mathbf{M}_0$ (with respect to H_{ex}), the impedance also shows an asymmetric behavior (AMI) [18]-[20], [27]. This case is realized in multilayer structures with cross-anisotropy in the presence of the dc bias current I_b . The other important conclusion is that the components of $\hat{\zeta}$ have a different symmetry with respect to the dc magnetization: the diagonal components do not change when the direction of $\mathbf{M}_0$ is reversed, whereas the off-diagonal components do. Then, the voltage response in (6) represented by the combination of ζ_{zz} and ζ_{zy} can exhibit an asymmetric behavior, which is not associated with that for $\mathbf{M}_0(H_{\text{ex}})$ [27]-[30].

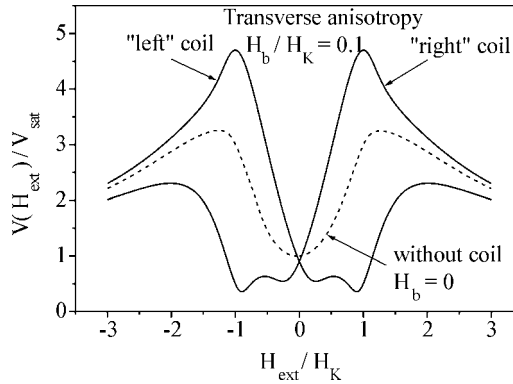
3.1 Dynamical AMI

This section is concerned with the role that the off-diagonal component ζ_{zy} plays in producing asymmetry in the total voltage response V_f . According to (6), the ζ_{zy} -contribution to V_f appears in the presence of the ac bias field h_{ex} produced by the coil current, which is the reason to name this type of asymmetry as dynamical AMI. Let us consider a film with a transverse anisotropy, since this case does not have asymmetry in $\mathbf{M}_0(H_{ex})$. The transverse domain structure is supposed to be eliminated by the application of a small dc current inducing the transverse bias field H_b . In this case, the voltage is asymmetrical because $\zeta_{zz}(H_{ex}) = \zeta_{zz}(-H_{ex})$ and $\zeta_{zy}(H_{ex}) = -\zeta_{zy}(-H_{ex})$.

Figure 5(a) shows plots of the real parts of ζ_{zz} and ζ_{zy} vs. H_{ex} , at a frequency of 120 MHz. The magnetic and electric parameters used correspond to amorphous CoSiB/Cu/CoSiB multilayers and $2d = 1 \mu\text{m}$ and $2d_1 = d_2$. The field behavior of the diagonal component ζ_{zz} is characterized by two symmetrical peaks when H_{ex} is equal to $1.3 H_K$. On the other hand, the off-diagonal component ζ_{zy} is an-



(a)



(b)

Figure 5 Surface impedance tensor (a) and voltage response (b) in a film attached to a coil (transverse anisotropy).

tisymmetric with respect to H_{ex} , showing nearly linear behavior in the field range of $\pm 0.8H_K$. At the end of this interval it shows either a maximum or minimum, and tends to zero as far as the magnetization angle θ becomes 0° and 180° which corresponds to $H_{ex} = \pm H_{sat}$, where H_{sat} is the field required to align the dc magnetization $\mathbf{M}_0$ along the z axis ($H_{sat} > H_K$ in the presence of H_b). The value of ζ_{zy} at the extremes is of the order of $\zeta_{zz}(H_{sat})$ and, consequently, according to (6) the application of the ac bias can considerably modify the voltage behavior.

Figure 5(b) shows plots of the voltage V_f as a function of H_{ex} , calculated in the presence of the ac external field h_{ex} induced by the coil current. It is seen that V_f exhibits a considerable asymmetry with respect to H_{ex} . For comparison, the voltage behavior without h_{ex} (or without the coil) is shown as well. In this case, the voltage has two symmetrical maximums in line with the ζ_{zz} vs. H_{ex} plot of Figure 5(a). Upon applying h_{ex} in a positive or negative sense (or using the “right” or “left” coil), one of these peaks is suppressed whereas the other is enhanced, and the voltage exhibits essentially asymmetric behavior with an increased sensitivity. A similar result was found in an amorphous wire with circumferential magnetization excited by a current applied to the film and the coil around it [29]. If then two oppositely biased films are connected with the differential amplifier, the output voltage becomes nearly linear in the field range of $\pm 0.8H_K$. It means that the linearity of the voltage output within the field interval of the order of the anisotropy field can be obtained by applying a much smaller ac field, the measure for which is the ac field induced by the ac current flowing in the film.

3.2 Static AMI in a film with cross-anisotropy

In this section, AMI due to an asymmetric magnetization reversal is discussed [18]-[20], [27]. The needed magnetic configuration can be realized in the multilayer structure by combining a cross-anisotropy and dc bias current producing a circumlatory bias field H_b . The rotational magnetization process is manifest in Figure 6

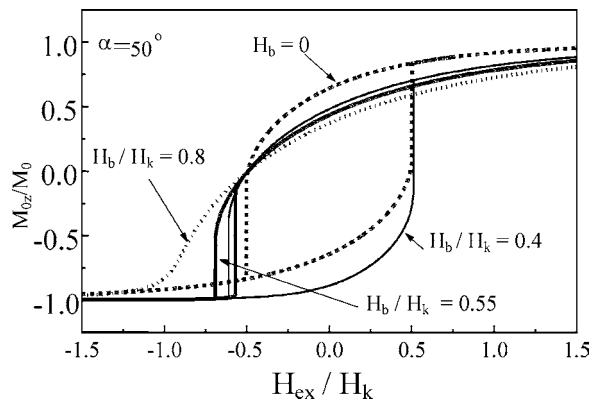


Figure 6 Magnetization curves $M_{0z}(H_{ex})$ in a wire with a helical anisotropy ($\alpha = 50^\circ$) for different magnitudes of dc bias field H_b .

where the magnetization plots for different values of H_b are given. Reversal of $\mathbf{M}_0$ by rotation is possible, since the magnetization vector during its rotation is held parallel to the surface, without going through high-energy demagnetization states. It is seen that the dc bias causes transition from a symmetric hysteresis curve to an asymmetric anhysteretic one at $H_b = H^* = H_K \cos \alpha$. The impedance characteristics reflect this behavior as shown in Figure 7, wherein the plots of ζ_{zz} vs. H_{ex} are given [24]. For H_b slightly larger than H^* , the field sensitivity of the impedance change is especially high, and the nominal change can be more than 100% for negative fields, when H_{ex} is changed by only $0.1 H_K$. The other components of the impedance tensor exhibit a similar behavior.

4 Experimental methods

A Hewlett–Packard Network Analyzer with a specially designed high-frequency measuring cell is used for impedance measurements [31–32]. The sketch of the measuring system is shown in Figure 8. The analyzer is configured in the two-port option (P1, P2) allowing the S_{21} -parameter to be measured. This parameter is the ratio of the forward transmission signal V_{out} to the excitation signal V_{in} . Depending on the excitation method, the voltage V_{in} (in port P1) determines either the ac current i in the film or current i_c in the coil. The output voltage V_{out} (in port P2) is taken either from the film ($V_{out} = V_f$) or from the attached coil ($V_{out} = V_c$). After choosing a certain excitation and measuring channels, all the components of $\hat{\zeta}$ can be determined. Since the analyzer has an unequal sweep parameter frequency, the field scanning has to be done by another device. The field H_{ex} is produced from the coil driven by a functional generator and power amplifier. The field is driven in both positive and negative directions to produce the hysteresis plots. The operating

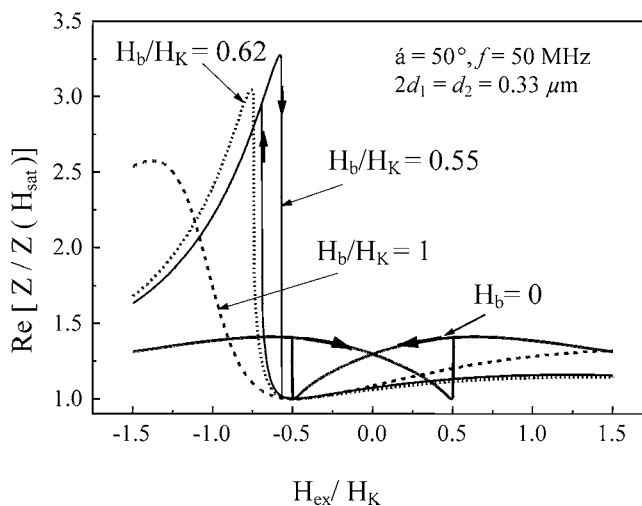


Figure 7 Modifications in the longitudinal impedance ζ_{zz} vs. H_{ex} under the effect of dc bias $0 \leq H_b / H_K \leq 1$, when $\alpha = 50^\circ$.

processes of the analyzer and functional generator are synchronized by a computer program. Thus, the functional generator is used as a programmable stepwise dc source to provide the field scans.

The measured sample is placed onto the open-type cell made of copper-coated fiberglass printed circuit board (PCB). All the connection stripes were made 2.8 mm in width to provide the wave impedance of the order of $50\ \Omega$ over a wide range of frequencies. The electrical circuits of the cells for measuring ζ_{zz} and ζ_{yz} are shown in Figure 9. A blocking capacitor (C) prevents the dc bias current I_b from entering the analyzer. The cell has input and output 3.5-mm connectors that are connected to the analyzer ports using 3.5-mm to type-N adapters. The microwave track including the cables and adapters was calibrated for the two-port measurements.

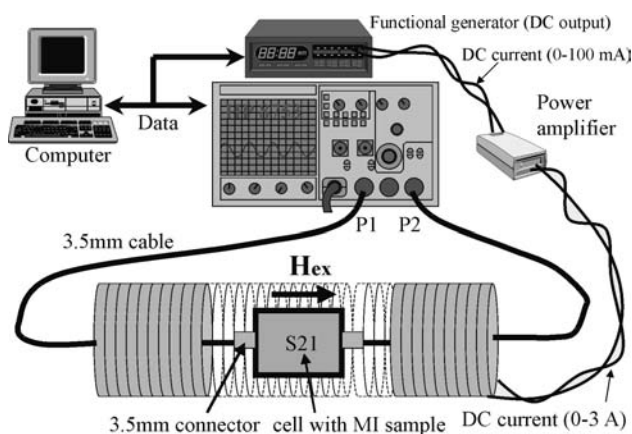


Figure 8 Sketch of the S_{21} -parameter measurement system.

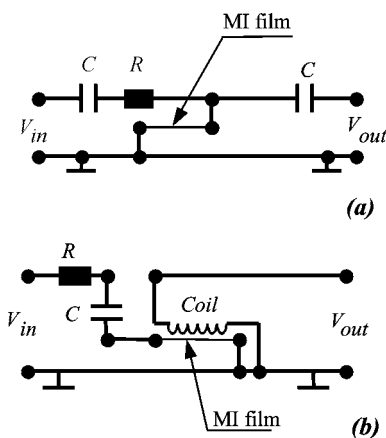


Figure 9 Electronic circuits of the cells to measure (a) diagonal and (b) off-diagonal components of the surface impedance tensor.

The longitudinal diagonal component ς_{zz} is determined by measuring the voltage V_f across the film excited by an ac current i . In this case, the cell of Figure 9(a) is used, for which $V_{out} = V_f$ and $i_c = 0$, so that the S_{21} parameter is directly proportional to ς_{zz} . It follows from (9) that $S_{21} = \varsigma_{zz}(H_{ex})(Al_i/V_{in})$. The circulatory diagonal component ς_{yy} corresponds to the voltage V_c in the secondary coil mounted on the film, which is excited by the primary coil (i_c). This case does not present substantial interest and is omitted here. The off-diagonal components $\varsigma_{zy} = \varsigma_{yz}$ can be determined by measuring the coil voltage V_c when the film is subjected to current i , as shown in Figure 9(b). In this case, $V_{out} = V_c$, $i_c = 0$, and (9) gives $S_{21} = \varsigma_{yz}(H_{ex})(2AbN_2i/V_{in})$. Alternatively, the off-diagonal components can be found by measuring the voltage V_f in the presence of i_c at the condition $i = 0$. If the amplitudes of the ac excitation current (in the film or in the external coil) are chosen to be less than 1 mA, nonlinear ac magnetization processes such as irreversible domain-wall displacements are not possible. The experimental studies are carried out with a dc current which effectively governs the static magnetic structure.

5 Film preparation and experimental results

Let us now consider experimental results on MI in thin-film structures. The films are prepared by rf or dc magnetron sputtering methods. The in-plane anisotropy can be established by applying the magnetic field during the deposition. In certain cases, further annealing is required. The induced anisotropy field is in the range of 5–20 Oe. After deposition, the layers can be patterned by conventional photolithographic methods and chemical etching in the needed geometry. Typically, a copper layer is used as the nonmagnetic lead. However, it was found that gold may have advantages because it is more reliable and produces much better edge definition when chemically etched.

5.1 MI in CoFeSiB/Cu/CoFeSiB multilayers

In the case of CoFeSiB/M/CoFeSiB multilayers ($M = \text{Cu, Ag, Au}$) several microns thick, a very large impedance change (MI ratio) of about 100% can be obtained at a frequency of 1 MHz, whereas the change would not be noticeable in a similar single layer [14-15]. For small fields $H_{ex} \sim 0$, the impedance magnitude is as low as 0.32–0.36 Ω , which indicates that it is dominated by the resistance of the M-layer (more than 50 times smaller than that of amorphous layers). The following results were obtained using a copper lead. The MI ratio has two sharp peaks of 145% at $H_{ex} = \pm 9$ Oe for CoSiB multilayers, and of 125% at $H_{ex} = \pm 18$ Oe for CoFeSiB multilayers. With increasing frequency, the MI ratio increases, reaching a maximum at 10 MHz (340% for CoSiB films). These results are obtained for films having a width of the conductive lead ($b = 0.5$ mm) more than 10 times larger than the critical length b^* (taking $\tilde{\mu}_{yy} = 500$, $d_1 = 1.5$ μm and $d_2 = 2$ μm ,

the parameter $b^* \approx \lambda = 39 \mu\text{m}$), in good agreement with the theoretical results obtained for infinite structures (see Figure 4). As the frequency is further increased, the MI ratio drops rapidly, which can be associated with the experimental technique determining the input impedance of the total system, including that of the lead wires.

A considerable enhancement of the MI effect in multilayers can be achieved by insulator separation between the conductive lead and the magnetic films [15], which further decreases the dc resistance. A CoSiB/SiO₂/Cu/SiO₂/CoSiB multilayer exhibited the MI ratio of 620% for 11 Oe.

Morikawa *et al.* [15] investigated the MI effect in CoSiB structures having different inner leads made of silver, copper or titanium. The insertion of a Ag layer (having the smallest resistivity of $1.62 \mu\Omega \text{cm}$) resulted in the largest MI ratio of 440%, where as the structures having a Ti inner layer with the resistivity of $47.8 \mu\Omega \text{cm}$ (only 2.7 times smaller than that of CoSiB) did not exhibit noticeable improvements in MI behavior (as compared with that of a CoSiB layer): the MI ratio was only 50% at a frequency of 100 MHz.

5.2 MI in NiFe/M/NiFe multilayers

Permalloy (Ni₈₀Fe₂₀), having almost no magnetostriction, is an excellent soft magnetic material. The resistivity of a NiFe layer is $20 \mu\Omega \text{cm}$, which is about 12 times larger than that of Cu, Ag or Au. It implies that this material is also suitable for multilayer MI, as was proposed by Hika *et al.* [16]. Here we present new results on MI in NiFe/Au/NiFe multilayers obtained for various geometries and frequency ranges.

The films were deposited on a glass substrate in the presence of the dc field of about 100 Oe. The film geometry has been proven to affect greatly the dc magnetic structure. Taking l in the range of 2–5 mm and with thickness $d < 2 \mu\text{m}$, we found that longitudinal anisotropy is induced if the film width is smaller than $200 \mu\text{m}$. After deposition, annealing was needed to establish the anisotropy axes in the transverse direction. This effect is due to the tension existing at glass/metal interface. With a proper choice of glass substrate, the tension can be released by annealing, so that the anisotropy returns to the direction established by the magnetic field applied during the deposition process.

Figure 10 shows plots of the real part of Z as a function of H_{ex} for films with longitudinal anisotropy. For films with larger widths ($b > 100 \mu\text{m}$), the MI ratio reaches about 50% at frequencies of 50–70 MHz. In narrower films ($b < 50 \mu\text{m}$), higher frequencies are needed to obtain large MI ratios, as shown in Figure 11. This agrees well with theory (compare with Figure 4).

Figure 12 shows plots of the real part of Z as a function of H_{ex} for films with transverse anisotropy. These films were deposited in the presence of a dc magnetic field. After deposition, they were annealed to release the stress and establish transverse anisotropy. On comparing with the data obtained for the case of longitudinal anisotropy, it is seen that the transverse anisotropy results in much higher sensitivity. The field characteristics alter as well, displaying two symmetrical maximums

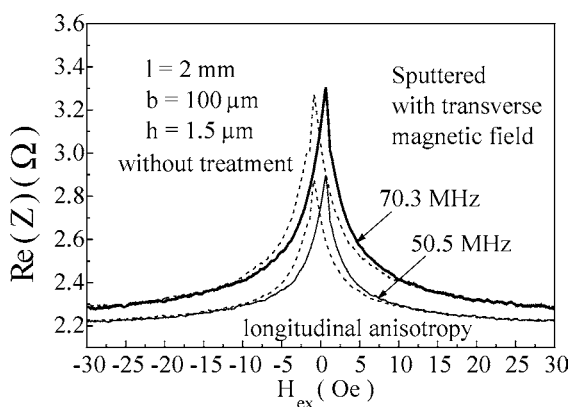


Figure 10 Real part of the impedance vs. field in FeNe/Au/FeNe films with longitudinal anisotropy for $b = 100 \mu\text{m}$. Dashed curves correspond to H_{ex} reversed from negative to positive direction.

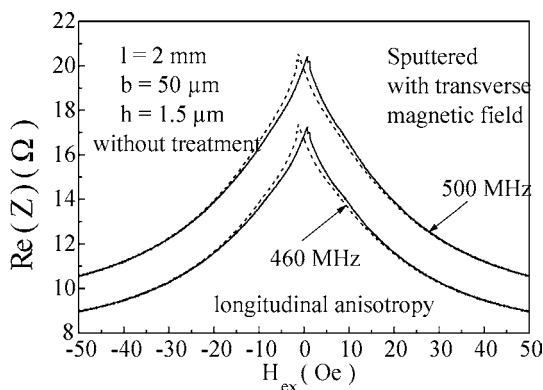


Figure 11 Real part of Z vs. H_{ex} for FeNe/Au/FeNe films with longitudinal anisotropy; $b = 50 \mu\text{m}$, and $f = 400\text{--}500 \text{ MHz}$. Dashed curves correspond to H_{ex} reversed from negative to positive direction.

when the external field equals the anisotropy field (about 10 Oe). The MI ratio and sensitivity increase with frequency reaching 240% and 24%/Oe, respectively, at frequency of 500 MHz.

5.3 Asymmetric MI (AMI)

We have considered miniature multilayer structures exhibiting the MI sensitivity in the range of 25% per oersted, which makes them very attractive for sensor application. One more important parameter of sensor operation is linearity. On the other hand, the MI characteristics seen in Figures 10–12 are not only nonlinear, but also shaped in a way that the operation near zero-field point can present serious problems. Generally, a dc bias field can be used to set properly the operating point on the MI characteristics. In fact, biasing produces asymmetry with respect to the

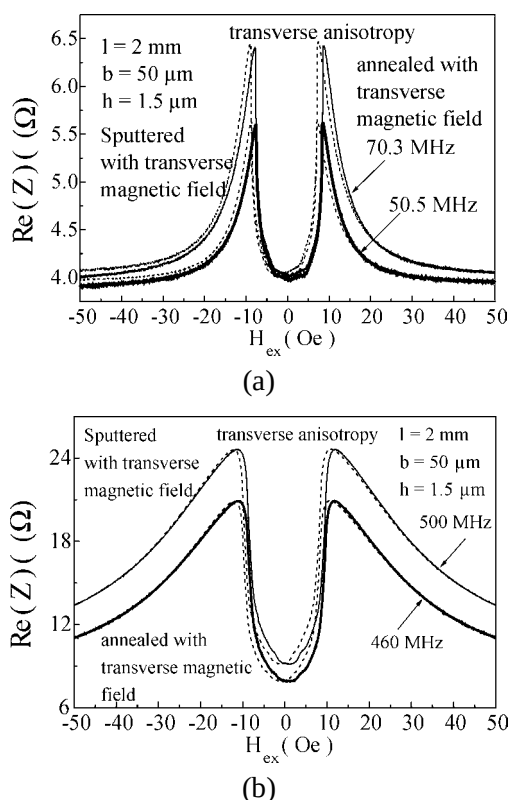


Figure 12 Real part of Z vs. H_{ex} for FeNe/Au/FeNe films with transverse anisotropy. (a) $f = 50\text{--}70$ MHz; (b) $f = 450\text{--}500$ MHz. Dashed curves correspond to H_{ex} reversed from negative to positive direction.

sensed field H_{ex} . To simplify the sensor design, it would be preferable to utilize intrinsic AMI characteristics as analyzed in Section 3. Two types of AMI effects can be realized in multilayers. The first one requires asymmetrical static magnetic configuration as that of cross-anisotropy multilayers [33]. The second type, the so-called dynamical AMI, is realized in film-coil structures.

First we consider the results obtained in cross-anisotropy films of the composition $\text{Co}_{70.2}\text{Fe}_{7.8}\text{B}_{22}/\text{Cu}/\text{Co}_{70.2}\text{Fe}_{7.8}\text{B}_{22}$. The films were made on a glass substrate by dc sputtering. The sputter rates were 17 nm/min and 50 nm/min, while the background pressure and Ar gas pressure were 3×10^{-5} Pa and 0.1 Pa, respectively. During deposition, the temperature of the substrate was 35°C , and a constant magnetic field of 200 Oe was applied in the transverse direction to the MI element in order to add uniaxial anisotropy. Finally, crossed anisotropy was induced in the sample by current annealing (30 mA) in a longitudinal field of 11.8 Oe at a temperature of 215°C . The sample in-plane view and magnetic configuration is shown in Figure 13. The thickness of all the layers is $0.5 \mu\text{m}$. The anisotropy axes n_K in the upper and bottom layers are at approximately 67° to the z axis, which was estimated from the dc magnetization loop measurements.

20% per oersted. With increasing frequency, the field characteristics change, as is evident in Figure 15, since the rotational processes become more significant. Figure 16 shows the differential characteristic from two oppositely biased MI films, which has a near-linear region in the field interval of ± 5 Oe. This demonstrates the achieved linearity without using a dc bias field.

Very promising results on AMI have been obtained using NiFe/Au/NiFe multilayers deposited in the presence of a dc magnetic field. After the deposition of the bottom layer, the substrate was rotated by 90° ; therefore, the two magnetic layers were deposited in cross magnetic fields. After releasing stress by annealing, a cross-anisotropy film was finally produced. The AMI characteristics obtained in the presence of a dc bias current I_b are given in Figure 17. In this case, a well-pronounced asymmetry is obtained at high frequencies (in the 30–70 MHz range) is due to the rotational ac dynamics. Although the field sensitivity in the presence

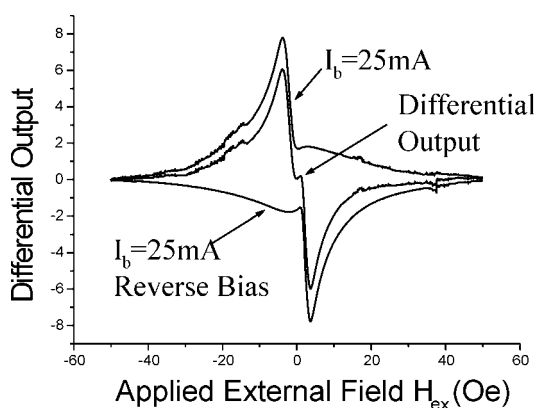


Figure 16 Differential voltage response at 10-MHz frequency, obtained using two oppositely biased films with the MI characteristics given in Figure 14.

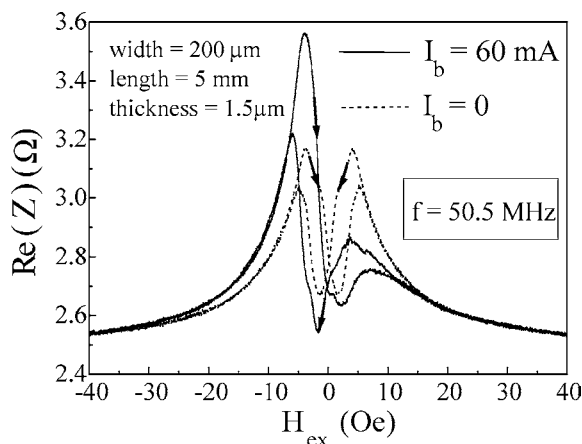


Figure 17 Real part of impedance $\text{Re}(Z)$ vs. H_{ex} in NiFe/Au/NiFe multilayers with cross-anisotropy at frequency 50.5 MHz for two values of bias current I_b : 0 and 60 mA.

of I_b can be very high, there is a substantial hysteresis. The obtained results on cross-anisotropy films display a very good compromise on size, sensitivity, and linearity.

6 Practical MI sensor design

Typically, MI sensors are based on diagonal impedance characteristics. A complicated biasing technique is then needed to properly set the operational regime for linear sensing. If an AMI effect is used, then two film elements are needed to obtain a differential response. Several sensor circuits have been proposed by Mohri and coauthors [5], [34], using either sinusoidal or pulsed excitation of an amorphous wire MI element. These circuits can be implemented also with thin films. Further development of MI sensors have shifted to the use of off-diagonal impedance components [35], when the output is taken as a coil voltage. In this case, circuit design can be simplified and sensor parameters (such as linearity and temperature stability) are improved. In this section, we focus on principles for using the off-diagonal impedance for sensor applications.

Let us consider a circuit where the off-diagonal response (proportional to ζ_{yz}) is taken from the pickup coil, and a single MI element is used. The principal electronic circuit is shown in Figure 18. The C-MOS IC multivibrator (with inverters Q) in this circuit produces smoothed rectangular pulses at a frequency of a few hundred kilohertz. After passing through the differential circuit ($C1-R1$), the pulse is sharpened and high-frequency harmonics exist in its spectrum. Pulse excitation is preferred over sinusoidal excitation for several reasons, such as simplicity of electronic design, low-cost components and high stability. In addition, the pulse spectrum contains a dc offset (zero harmonic) that provides the dc bias required for the existence of the off-diagonal response. Thus, the pulse circuit produces both the high-frequency excitation and dc bias. The ac off-diagonal pulsed response V_c is

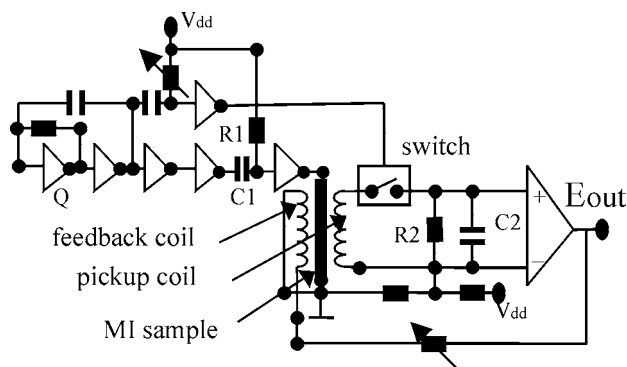


Figure 18 Electronic circuit of a linear sensor using the off-diagonal impedance for a linear sensor. It comprises a C-MOS IC multivibrator with inverters (Q), a differential circuit ($C1-R1$), analog synchronised switch, rectifier ($R2-C2$), a differential amplifier, and negative feedback.

taken from the pick-up coil using the analog synchronized switch, and converted by the rectifier (R2–C2). The synchronous rectification suppresses the noise that appears during the rectification process and produces a quasi-dc voltage of a magnitude proportional to the original pulse amplitude. Thereafter, the rectified voltage is amplified. This amplified signal is characterized by an amplitude and a sign, both of which are sensitive to the dc external magnetic field H_{ex} since the off-diagonal response is used. Thus, a near-linear output voltage signal is obtained without use of negative feedback, which, however, can be added to further improve linearity. The circuit proposed by Kawajiri *et al.* [35] uses two MI elements and two analog switches to create a balanced circuit with very high temperature stability.

Another method for obtaining a linear response is provided by so-called mixed excitation [27]–[30], wherein an MI film with transverse anisotropy (or wire with the circumferential anisotropy) is excited by both the ac longitudinal current i and the ac longitudinal magnetic field h_{ex} . The excitation can be pulsed or sinusoidal. For sinusoidal excitation, an additional dc bias current is required to generate off-diagonal surface impedance. The field h_{ex} can be induced by a coil mounted around the sample and connected in series. The output signal V_f is measured across the MI sample. As can be deduced from (6), the ac response is formed by the sum of the diagonal $\zeta_{zz}(H_{ex})$ and off-diagonal $\zeta_{zy}(H_{ex})$ impedance components with corresponding coefficients. Since $\zeta_{zz}(H_{ex})$ and $\zeta_{zy}(H_{ex})$ are symmetric and antisymmetric, respectively, the field dependence of the output voltage magnitude $|V_f|$ on H_{ex} is asymmetric. The linear sensor circuit is shown in Figure 19. The signals $V1(H_{ex})$ and $V2(H_{ex})$, taken from the two MI sensors with reverse asymmetries, are detected in a differential sense to produce the total linear field response $V_{dif}(H_{ex})$, as shown in Figure 20. Therefore, no additional dc bias field is required in this case to produce a linear output response as well.

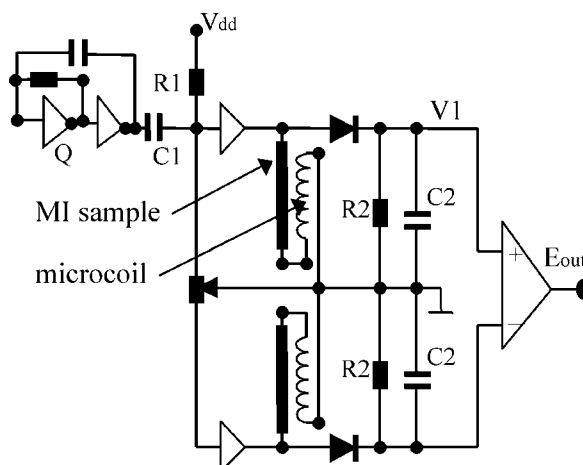


Figure 19 Electronic circuit of a linear sensor using mixed excitation. The MI samples are excited by both the ac longitudinal current i_z and the ac longitudinal magnetic field h_{ex} . The output signals are measured across the MI samples.

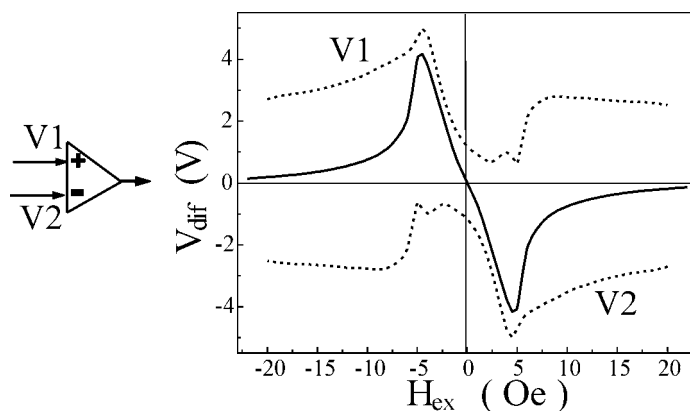


Figure 20 Principle underlying the near-linear field response of a differential circuit. Two voltage responses V_1 and V_2 produced with opposite ac bias are added together in the differential amplifier.

7 Conclusions

The following conclusions emerge from the developments presented in this chapter:

1. The MI effect in multilayers, which has two soft ferromagnetic films (F) sandwiching a highly conductive lead (M), has significant advantages over that of a single layer. First, the MI ratio is several times greater, ranging within 100–600% for films of 1–10 μm thickness. Second, the impedance shows large changes (50–100%) at relatively low frequencies (~ 1 MHz in this case) when the skin effect is not substantial.
2. Theoretical analysis has been developed for a three-layer structure (F/M/F) with magnetic films having an in-plane anisotropy of a “spiral” type. For a considerable conductivity difference between the layers F and M, the inductance of the magnetic films is the main contributor to the system impedance at relatively low frequencies. The voltage response is linearly proportional to the permeability tensor, which can be extremely sensitive to the longitudinal sensed field. The theory accounts for the influence of the in-plane dimensions on the MI ratio, yielding a quantitative agreement with the experiment.
3. Consideration of asymmetric MI requires an approach involving a surface impedance tensor. Regarding the conceptual aspects, it has been demonstrated that there are two principally different mechanisms of AMI. One is related to the asymmetric static magnetization reversal, and the other is caused by the dynamical cross-magnetization process represented by the off-diagonal impedance.
4. The application of MI multilayers for new, sensitive and quick-response micromagnetic sensors is at an early stage of its development. However,

it can be expected that significant technical achievements in this field will be made in the foreseeable future. The tensor character of the surface impedance opens new principles for linear sensors designed to take advantage of antisymmetric off-diagonal impedance.

References

- [1] D. Robbes, C. Dolabdjian, S. Saez, Y. Monfort, G. Kaiser and P. Ciureanu, Highly sensitive uncooled magnetometers: state of art. Superconductive magnetic hybrid magnetometers, an alternative to SQUIDS? *IEEE Trans Appl Superconductivity* **11** (2001), 629–634.
- [2] D.J. Mapps, Magnetoresistive sensors, *Sens Actuat A: Phys* **59** (1997), 9–19.
- [3] F. Primdahl, The fluxgate mechanism, part I: The gating curves of parallel and orthogonal fluxgates, *IEEE Trans Magn* **6** (1970), 376–383.
- [4] K. Mohri, L.V. Panina, T. Uchiyama, K. Bushida and M. Noda, Sensitive and quick response micro-magnetic sensor utilizing magneto-impedance in co-rich amorphous wires, *IEEE Trans Magn* **31** (1995), 1266–1275.
- [5] K. Mohri, T. Uchiyama and L.V. Panina, Recent advances of micro magnetic sensors and sensing application, *Sens Actuat A: Phys* **59** (1997), 1–8.
- [6] M. Vazquez and A. Hernando, A soft magnetic wire for sensor applications, *J Phys D: Appl Phys* **29** (1996), 939–949.
- [7] L.V. Panina and K. Mohri, Magneto-impedance effect in amorphous wires, *Appl Phys Lett* **65** (1994), 1189–1191.
- [8] L.V. Panina, K. Mohri, K. Bushida and M. Noda, Giant magneto-impedance and magneto-inductive effects in amorphous alloys, *J Appl Phys* **76** (1994), 6198–6203.
- [9] R.S. Beach and A.E. Berkowitz, Giant magnetic-field dependent impedance of amorphous FeCoSiB wire, *Appl Phys Lett* **64** (1994), 3652–3654.
- [10] R.S. Beach and A.E. Berkowitz, Sensitive field-dependent and frequency-dependent impedance spectra of amorphous FeCoSiB wire and ribbon, *J Appl Phys* **76** (1994), 6209–6213.
- [11] J. Valazquez, M. Vazquez, D.X. Chen and A. Hernando, Giant magnetoimpedance in nonmagnetostrictive amorphous wires, *Phys Rev B* **50** (1994), 16737–16740.
- [12] L.V. Panina, K. Mohri, T. Uchiyama, K. Bushida and M. Noda, Giant magneto-impedance in co-rich amorphous wires and films, *IEEE Trans Magn* **31** (1995), 1249–1260.
- [13] D.P. Makhnovskiy, L.V. Panina and D.J. Mapps, Field-dependent surface impedance tensor in amorphous wires with two types of magnetic anisotropy: helical and circumferential, *Phys Rev B* **63** (2001), 144424–144441.
- [14] T. Morikawa, Y. Nishibe, H. Yamadera, Y. Nonomura, M. Takeuchi and Y. Taga, Giant magneto-impedance effect in layered thin films, *IEEE Trans Magn* **33** (1997), 4367–4369.

- [15] T. Morikawa, Y. Nishibe, H. Yamadera, Y. Nonomura, M. Takeuchi, J. Sakata and Y. Taga, Enhancement of giant magneto-impedance in layered film by insulator separation, *IEEE Trans Magn* **32** (1996), 4965–4967.
- [16] K. Hika, L.V. Panina and K. Mohri, Magneto-impedance in sandwich film for magnetic sensor heads, *IEEE Trans Magn* **32** (1996), 4594–4596.
- [17] L.V. Panina and K. Mohri, Magneto-impedance in multilayer films, *Sens Actuat A: Phys* **81** (2000), 71–77.
- [18] D.P. Makhnovskiy, A.S. Antonov, A.N. Lagar'kov and L.V. Panina, Field-dependent surface impedance of a bilayer film with an antisymmetric bias magnetization, *J Appl Phys* **84** (1998), 5698–5702.
- [19] L.V. Panina, D.P. Makhnovskiy and K. Mohri, Analysis of magneto-impedance in multilayers with cross-anisotropy, *J Mag Soc Jpn* **23** (1999), 925–931.
- [20] D.P. Makhnovskiy, L.V. Panina, A.N. Lagar'kov and K. Mohri, Effect of antisymmetric bias field on magneto-impedance in multilayers with crossed anisotropy, *Sens Actuat A: Phys* **81** (2000), 106–110.
- [21] L.V. Panina, D. Zarechnuk, D.P. Makhnovskiy and D.J. Mapps, Two-dimensional analysis of magnetoimpedance in magnetic/metallic multilayers, *J Appl Phys* **89** (2001), 7221–7223.
- [22] V. Korenivski, GHz magnetic film inductors, *J Magn Magn Mater* **215–216** (2000), 800–806.
- [23] A. Sukstanskii, V. Korenivski and A. Gromov, Impedance of a ferromagnetic sandwich strip, *J Appl Phys* **89** (2001), 775–786.
- [24] D.P. Makhnovskiy and L.V. Panina, Size effect on magneto-impedance in layered films, *Sens Actuat A: Phys* **81** (2000), 91–94.
- [25] A. Paton, Analysis of the efficiency of thin-film magnetic recording heads, *J Appl Phys* **42** (1971), 5868–5870.
- [26] R.W. Cross, S.E. Rassek, S.C. Sanders, M.R. Parker, J.A. Barnard and S.A. Hossain, Size and self-field effects in giant magnetoresistive thin-film devices, *IEEE Trans Magn* **30** (1994), 3825–3827.
- [27] L.V. Panina, D.P. Makhnovskiy and K. Mohri, Mechanism of asymmetrical magnetoimpedance in amorphous wires, *J Appl Phys* **85** (1999), 5444–5446.
- [28] D.P. Makhnovskiy, L.V. Panina and D.J. Mapps, Asymmetrical magneto-impedance in a sandwich film with a transverse anisotropy using an AC bias, *J Magn Magn Mater* **215** (2000), 629–633.
- [29] K. Gunji, L.V. Panina and K. Mohri, Asymmetrical magneto-impedance in amorphous wires with helical current excitation, *J Mag Soc Jpn* **21** (1997), 793–797.
- [30] D.P. Makhnovskiy, L.V. Panina and D.J. Mapps, Asymmetrical magnetoimpedance in as-cast CoFeSiB amorphous wires due to ac bias, *Appl Phys Lett* **77** (2000), 121–123.
- [31] D.P. Makhnovskiy, L.V. Panina and D.J. Mapps, Measurement of field-dependent surface impedance tensor in amorphous wires with circumferential anisotropy, *J Appl Phys* **87** (2000), 4804–4806.

- [32] D.P. Makhnovskiy, L.V. Panina and D.J. Mapps, Surface impedance tensor in amorphous wires with helical anisotropy: Magnetic hysteresis and asymmetry, *J Appl Phys* **89** (2001), 7224–7226.
- [33] K. Ueno, H. Hiramoto, K. Mohri, T. Uchiyama, and L.V. Panina, Sensitive asymmetrical MI effect in crossed anisotropy sputtered films, *IEEE Trans Magn* **36** (2000), 3448–3450.
- [34] K. Mohri, T. Uchiyama, L.P. Shen, C.M. Cai and L.V. Panina, Sensitive micro magnetic sensor family utilizing magneto-impedance (MI) and stress-impedance (SI) effects for intelligent measurements and controls, *Sens Actuat A: Phys* **91** (2001), 85–90.
- [35] N. Kawajiri, M. Nakabayashi, C.M. Cai, K. Mohri and T. Uchiyama, Highly stable micro sensor using C-MOS IC multivibrator with synchronous rectification, *IEEE Trans Magn* **35** (1999), 3667–3669.



Larissa V. Panina received her Ph.D. degree in 1985 from Moscow State University. Her specialist research experiences are in the fields of magnetism, magnetic materials, and electrodynamics of magnetic/metallic composites. During the period 1990–1998, she was with Nagoya University (Japan) where she played a leading role in the discovery of the magnetoeimpedance effect. From 1999, she has been with the University of Plymouth (UK), leading the research on electromagnetic sensors and microwave materials. To date, she has published 76 journal papers and has delivered 17 invited lectures at international conferences.



Dmitriy P. Makhnovskiy received a M.Sc. degree in 1993 from Moscow State University. His specializes in the areas of magnetic materials, thin-film technology, and electrodynamics of magnetic/metallic composites. From 1994 to 1998 he worked at the Centre for Theoretical and Applied Electrodynamics of the Institute for High Temperature, Russian Academy of Science (Moscow, Russia). He joined the University of Plymouth (UK) in 1999, where he is playing a key role in research on the magnetoeimpedance effect and microwave materials. To date, he has published 20 journal papers.

Metamaterials: An Introduction

Rodger M. Walser

Abstract

Metamaterials, so named to recognize and emphasize their purpose, aim to achieve material performance “beyond” the limitations of conventional composites. The holy grail of conventional composite design is to achieve composite properties that are a volumetric average of those of its constituents. One of the goals of this chapter is to outline the strategy of metamaterial design, its hallmark being the exploitation of low dimensionality to extend composite performance. Another goal is to discuss the generality of the metamaterial strategy, and to illustrate its successful implementation in diverse application areas. Finally, the use of metamaterial synthesis to achieve nearly ideal electromagnetic composite responses is discussed. The rules for this approach are illustrated by outlining the design of a laminated electromagnetic composite that possesses some remarkable and previously unobserved properties.

1 Introduction

To meet societal needs, the *cause-and-effect* orientation that has dominated materials research for the past four decades is being rapidly integrated with a systems approach that has a *goal-and-means* orientation [1]. In the cause-and-effect process, the complexity of materials is reduced to a minimum so that deductive reasoning can be focused on (usually) understanding a single material–property relationship. In contrast, performance-driven materials synthesis must embrace the *essential complexity* of material structure and a multitude of constraints, including those imposed by socioeconomic factors such as cost, environment, and reliability. Even in the absence of the socioeconomic constraints, the true atomic-scale synthesis of a macroscopic material having an optimized combination of multiple properties that satisfy performance goals, but which does not violate constraints imposed by physics, requires the solution of an N -body problem so large and complex that it is out of man’s reach, even if Moore’s law [2] were extended for the next millennium.

In fact, it is fair to say that, to this point in time, nature has rescued us from this chore by providing an abundance of materials that are assembled by the non-deterministic fluctuations that occur at first-order phase transitions; i.e., we cannot know the rules but do not need to. The insatiable appetite of technology, however, has forced us to emulate nature, and synthesize sophisticated composites with nondeterministic constituents to relieve the physics constraints and obtain further increases in material performance.

There are two fundamentally different strategies for synthesizing ordered macroscopic composites from different repertoires of nondeterministic constituents. In the first strategy (relatively new to mankind), ordered macroscopic materials are synthesized from the bottom up by assembling macromolecules exemplified by fullerenes and carbon nanotubes. I suggest that this type of new materials be called *mesomaterials* to recognize that their architecture is derived from mesoscale supramolecules [3]. In the second, more mature strategy, ordered macroscopic composites are synthesized from the top down, by first disassembling and then recombining macroscopic constituents (Figure 1). This chapter is concerned with the second type, which I have suggested be called *metamaterials* [4] to recognize that their architecture is aimed at achieving performance beyond that of conventional macroscopic composites.

This chapter is organized as follows: Section 2 reminds us of the formal rules governing the synthesis of conventional macroscopic composites. The motivation for metamaterials discussed in Section 2.1 is derived from the need to extend the performance limits of conventional composites as an application area matures. The key feature of metamaterial synthesis described in Section 2.2 is the use of an artificial periodic cellular architecture with cells containing low-dimensional meta-particles whose properties are specifically designed to achieve an optimized combination of responses beyond that achievable with macroscopic constituents. Metamaterials can, in principle, be synthesized for any application area, and the general

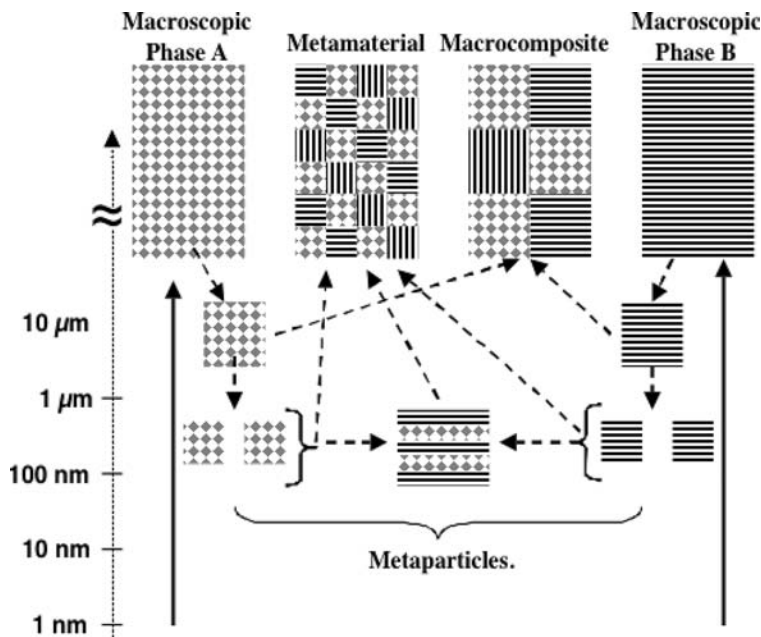


Figure 1 Contrast in synthesis of conventional macroscopic composites and metamaterials. Metaparticles are constituents in which new excitations or relaxations achieve significance when a principal dimension is made comparable with a characteristic internal length scale. Some characteristic length scales are listed in Section 2.2.

motivation for doing so is discussed in Section 3. Two examples of metamaterial synthesis are discussed in Sections 3.1 and 3.2. Electromagnetic metamaterials, a special focus of this chapter, are discussed in Section 4.

2 Conventional macroscopic composites

The design of macroscopic composites is a highly developed branch of materials science for developing useful materials for every field of engineering. At present, the synthesis of composites in nearly every area of technology is primarily limited to that of *macroscopic composites* in which *macroscopic material constituents* are assembled to realize *new macroscopic materials*. Their design rules have one principal consideration: to ensure that the composite properties are derived from the law of mixtures [5] which, for a two-phase composite with volume fractions f_A and f_B , of macroscopic constituents A and B , allows the composite property to be expressed as

$$P_C = \langle P \rangle_V = f_A P_A + f_B P_B, \quad (f_A + f_B = 1). \quad (1)$$

Equation (1) reduces the complexity of the design, allowing the properties of the composite to be computed from a volumetric average of the properties of its

macroscopic constituents. In general, to satisfy (1) the constituents must be randomly distributed or have either a series or parallel anisotropy. With this motivation, the design rules for *conventional macroscopic composites* [5] can be precisely stated as shown on the left side of Table 1.

For conventional composites, these rules are regarded as *necessary and sufficient* to guarantee that their properties can be obtained from the law of mixtures, i.e., (1). Rule 1C eliminates natural composites (such as wood and granite) because their structure formation was not under human control. Hence, the properties of the constituents are not known, and the validity of Rules 3C and 4C cannot be confirmed. A similar argument eliminates composites with kinetic structures, e.g., spinodal decompositions [6-7]. Rules 2C and 2M ensure that the properties of the composite are valid averages. Rule 3C is the primary objective for the synthesis of macroscopic composites, and its validity is guaranteed by (1). Rule 4C rejects composites in which nonlocal or reactive coupling, which invalidates (1), occurs between constituents after assembly. This rule was originally expressed for mechanical composites, but the italicized addition generalizes its application to electromagnetic and other types of composites.

2.1 Motivation for metamaterials

The range of composite functions allowed for a two-phase composite by the law of mixtures (1) is given by the straight line connecting *A* and *B* in Figure 2. For most applications, a performance goal is established by an *application figure of merit* (FOM). When socioeconomic and other factors are removed, the performance goals for the material are expressed as a material FOM which, along with constraints on the minimum and maximum acceptable values of the properties, defines an acceptable range of composite properties for that application. The example shown in Figure 2 uses a commonly occurring FOM expressed as a condition on a product of the required composite properties in the form

$$P_{1C} \cdot P_{2C} \geq c, \tag{2}$$

Table 1 Comparison of the design rules for conventional (C) and metamaterial (M) composites.

Conventional (C) composites must:		Metamaterial (M) composites must:	
1C	Be manmade	1M	Be manmade
2C	Have constituents combined three-dimensionally	2M	Have constituents combined three-dimensionally
3C	Have new properties <i>not achieved by any of its constituents acting alone</i>	3M	<i>Either</i> have new or extended properties <i>due to its metaparticles</i>
4C	Have two chemically distinct materials with distinct interfaces (<i>i.e., have only local interactions</i>)	4M	<i>and/or</i> have local <i>and/or</i> non-local interactions between constituents

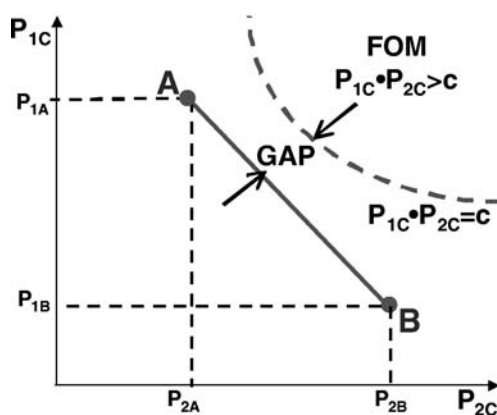


Figure 2 In this example, there is a gap between the range of properties that can be realized in a ideal macroscopic composite (on the straight line between A and B) and those required by the material FOM (above the hyperbola defined by $P_{1C}P_{2C} = c$) for a particular application.

where numerical subscripts indicate property 1 and 2, and subscripts A, B and C indicate materials A, B and their composite C.

As an application area matures, it becomes increasingly difficult to satisfy the material FOM by synthesizing new nondeterministic macroscopic materials. This is expected since the relationships between material properties required to maximize the FOM are invariably in conflict with the relationships between the properties that are imposed by physics.

For example, consider the case of a magnetomechanical converter (second entry in Table 4). The design of a fiber optic magnetometer requires a magnetic material that will respond to small magnetic fields and transmit a strain to an optical fiber via its magnetostrictive response [8]. The FOM of the material for this application is the product of λ_s , the saturation magnetostriction, and the reciprocal of H_k , the magnetocrystalline anisotropy. Since both λ_s and H_k have their origin in the spin-orbit coupling, it is not surprising that there is a definite relationship between these quantities in all types of magnetic materials as shown in Figure 3.

There is an obvious conflict between the FOM requirement that λ_s/H_k be maximized, and the constraint that $\lambda_s \propto H_k$ revealed by Figure 3. To circumvent this problem, a metamaterial multilayer was fabricated with alternating magnetically soft (NiFe) and magnetically hard (CoFe) layers, as discussed by Rengarajan *et al.* [9]. The low-dimensionality concept used in this metamaterial design involved reducing the multilayer thicknesses so that alternating films were exchange-coupled. In the multilayer, this coupling causes the anisotropy to decrease faster than the magnetostriction. This *exchange spring* effect has been used to design a wide variety of novel magnetic multilayer metamaterials [10]–[12].

The law of mixtures can be used to generalize these conflicts for macroscopic composites. If two material properties P_{1C} and P_{2C} of the two-phase composite

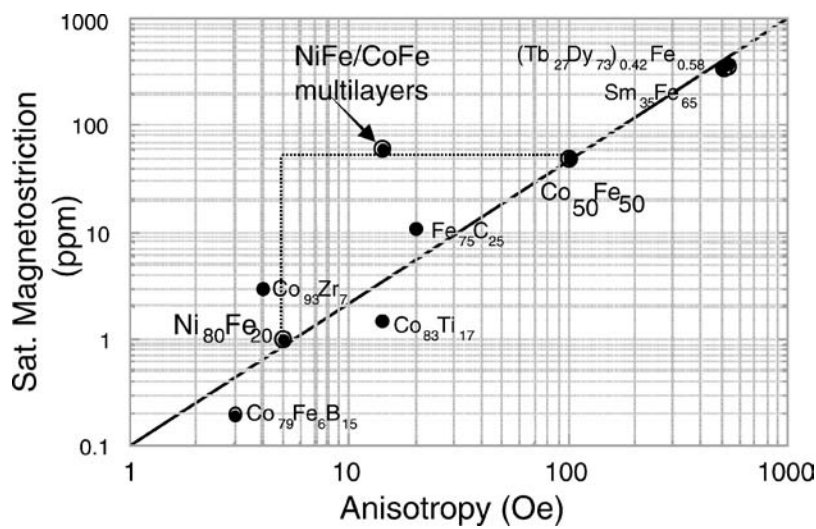


Figure 3 Plot of the saturation magnetostriction versus the magnetocrystalline anisotropy for a wide variety of magnetically soft and magnetically hard materials. The results show an almost linear relationship between these quantities, confirming thereby that the spin-orbit coupling in the material plays the determining role for both quantities. This constraint was modified by fabricating a multilayer metamaterial consisting of alternating soft and hard magnetic films [9].

discussed in connection with Figure 2 are required to satisfy a FOM of form

$$\frac{P_{1C}}{P_{2C}} \geq c, \tag{3}$$

and the law of mixtures (1), then

$$\frac{P_{1C}}{P_{2C}} = \frac{f_A P_{1A} + f_B P_{1B}}{f_A P_{2A} + f_B P_{2B}} \geq c. \tag{4}$$

If, furthermore, physics requires that

$$P_{1A} = \alpha P_{2A}, P_{1B} = \alpha P_{2B}, \tag{5}$$

then the FOM can be written as

$$\text{FOM} = \frac{P_{1A}}{P_{2A}} \frac{f_A + f_B (P_{1B}/P_{1A})}{f_A + f_B (P_{2B}/P_{2A})} = \alpha \geq c. \tag{6}$$

As an application area matures, therefore, it will eventually require that $c > \alpha$; and it will no longer be possible to satisfy the FOM with a conventional macroscopic composite material. As discussed in the following, metamaterial composites can then be synthesized to extend the capabilities of conventional composites to meet these further requirements.

at which the response of interest can be modified. If R and R' are the responses of the macroscopic material and low-dimensional metaparticle, respectively, to a local excitation E , then the low-dimensional property of the metaparticle P' can be found by noting that

$$R = P_{\infty} E, \quad R' = P' E, \quad (7)$$

and by taking

$$R' = R f(L/l_c) \quad (8)$$

to obtain

$$P' = P_{\infty} f(L/l_c). \quad (9)$$

The function $f(L/l_c)$ is the response of the metaparticle with principal dimension L in the range $L/l_c \leq 1$ relative to that of the macroscopic material P_{∞} , as a function of the characteristic length l_c of the excited states of the material responding to the excitation E . For example, when the thickness t of a metallic thin film is small compared to the electron mean-free path l_c , it can be shown [16] that the increase in its resistivity ρ due to the increased electron scattering at the boundaries is given by

$$\rho = \rho_{\infty} \cdot 3k \log(1 - l_c/t), \quad t/l \ll 1, \quad (10)$$

Table 3 Characteristic dimensions of a material at which modifications of static and dynamic macroscopic properties are observed (* indicates interface).

Excitation (E)	Response (R)	Properties (P)	Static Length (l_c)		
			Macro- scopic	Meso- scopic	Atomic*
Electric Field	Current Density	Electrical Conductivity		Free path Fermi λ	Tunnel Distance
Magnetic Field	Magnetic Flux Density	Magnetic Permeability	Induced Dipoles Domain Width	Domain Wall Thickness	Exchange Coupling
Electric Field	Electric Flux Density	Dielectric Permittivity	Induced Dipoles Domain Width	Domain Wall Width	Dipole Coupling
Excitation	Response	Properties	Dynamic	(Coherence)	Length (l_c)
Incident Wave	Transmit Wave	Transmission Coefficient	Internal E&M Wavelength		
Heat Flux	Thermal Gradient	Thermal Diffusivity	Thermal Diffusion Length		

where ρ_∞ is the resistivity of the macroscopic material. Although by no means exhaustive, Table 3 lists some characteristic material dimensions at which some modified static and dynamic responses will be observed.

3 Examples of metamaterials

The application of a material as a sensor, transducer, source, etc., involves reversible or irreversible processes; and the material FOM can always be interpreted in terms of thermodynamic requirements. On the other hand, material symmetries always impose constraints on the relationships of the same properties as in, for example, the Onsager reciprocal relations [17-18]. The property application requirements and physics constraints are always in opposition. For example, the material FOM (ratio of magnetostriction to magnetic anisotropy) considered in Section 2.1 is a measure of the ratio of the mechanical work done by the magnetomechanical converter to the magnetic energy supplied to it. As pointed out, the key properties involved in these processes are related through the spin-orbit coupling.

Table 4 lists some material FOMs for some application areas of current interest. Two examples from this Table are discussed to illustrate the use of low dimensionality in metamaterials design and to highlight some recent successes.

Table 4 Material figures of merit (FOMs) for various engineering applications.

Application: quantity	Figure of merit (FOM) = $P_1 * P_2$
Mechanical Support: Strength to Weight Ratio	$(\text{Bulk Modulus}) * (\text{Density})^{-1}$ $= B * \rho^{-1}$
Peltier Refrigerator: Coefficient of Performance	$(\text{Thermopower})^2 * (\text{Electrical / Thermal Conductivity})$ $= S^2 * (\sigma / \kappa)$
Microwave Magnetics: Snoek's Relation	$\text{Permeability}) * (\text{Ferromagnetic Resonance})$ $= \mu * f_{\text{FMR}}$
Permanent Magnet: Energy Product	$(\text{Magnetization}) * (\text{Coercive Force})$ $= 4\pi M_s * H_c$
Magnetic Sensor: Magnetoresistance	$(\text{Sensitivity}) * (\text{Conductivity})$ $= (\delta\rho / \delta H) * \sigma$
Recording Head: Magnetic Permeance	$(\text{Permeability}) * (\text{Skin Depth})$ $= \mu * \delta$
Tunable device: Low-Loss Bandwidth	$(\text{Q Factor}) * (\text{Tunability})$ $= (\mu' / \mu'') * (\delta\mu / \delta H)$
Phase Shifter: Insertion Phase	$(\text{Permeability}) * (\text{Permittivity})$ $= \mu * \epsilon$
Microwave Match: Wave Impedance	$(\text{Permeability}) * (\text{Permittivity})^{-1}$ $= \mu * \epsilon^{-1}$
Microwave devices: Surface Resistance	$(\text{Conductivity}) * (\text{Skin depth})$ $= \sigma * \delta$

3.1 Thermoelectric metamaterials

To make matters more specific, consider the material properties required for state-of-the-art thermoelectric refrigerators. The efficiency (or coefficient of performance) of these devices is determined by the product of its Carnot efficiency and a material parameter that is maximized through a dimensionless quantity called the ZT product [19]–[21]. In the definition

$$ZT = \frac{S^2 \sigma_e}{\Lambda} T = \frac{(S/\sqrt{L_o})^2}{1 + \left[\frac{\Lambda_L}{L_o \sigma_e} T \right]}, \quad (11)$$

S is the thermopower, σ_e is the electrical conductivity, $\Lambda = \Lambda_L + \Lambda_E$ is the sum of lattice and electronic thermal conductivities, and $L_o = 2.443 \times 10^{-8} \text{ (V/K)}^2$ is the Lorentz number. In the 1950s and 1960s, the search for materials with higher ZT values was focused on increasing the electrical parameter $S^2 \sigma_e$, while decreasing the ratio of Λ_L/Λ_E (constrained in metals by the Wiedeman-Franz law). These efforts produced several materials (e.g., PbTe, BiTe, and SiGe) with $ZT \sim 1$. However, after the initial advances, it was discovered that further increases in S were invariably accompanied by decreases in σ_e . No significant increases in ZT were made for the next three decades until 1995, when Slack [19] suggested that materials with $ZT \approx 4$ could be produced if the quantity σ_e/Λ_L was maximized in a so-called PGEC, a material that acted like a phonon glass (poor thermal conductor) and electron crystal (good electron conductor). Although Slack did not know how such a material could be made, he was clearly describing a metamaterial! With funding provided by U.S. Department of Defense research initiatives [22], Slack's insights were ingeniously pursued by several investigators, whose efforts over the next two to three years demonstrated the prospects for tripling the FOM of materials for use in Peltier refrigerators; see Figure 5.

Superlattices offer the most dramatic prospects for improved thermoelectric materials. These structures consist of thin (10–50 Å) layers of alternating quantum wells (QWs) and barrier layers (BLs), as shown in Figure 6. Two strategies for utilizing low dimensionality are contrasted in Figure 7:

- (i) increasing S by enhancing the density of states near the Fermi level [23] $d_{\text{QW}} \approx \lambda_f = \text{Fermi wavelength}$, and
- (ii) decreasing the ratio Λ_L/σ_e by enhancing phonon scattering relative to electron scattering at the QW/BL interfaces [24] ($d_{\text{QW}} \approx \lambda_{\text{mfp}} = \text{phonon mean-free path}$).

3.2 High-frequency magnetic metamaterials

The performance FOM of a sensor material is commonly expressed as a constraint on the response-bandwidth product. For magnetic materials this FOM is referred

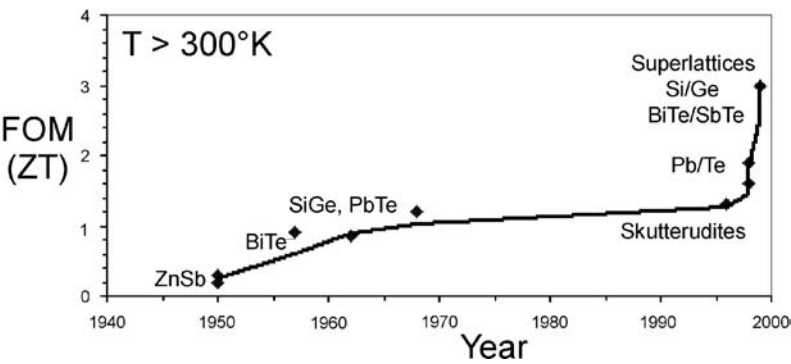


Figure 5 Increases in thermoelectric FOMs over the past 50 years [22]. After remaining constant for about 30 years, prospects for tripling the FOM have been demonstrated over the past few years, principally by exploiting low dimensionality in superlattices.

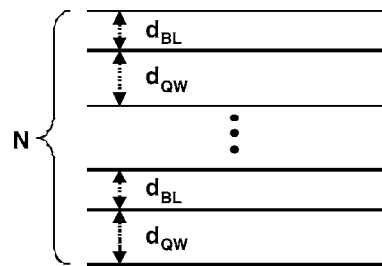


Figure 6 Thermoelectric superlattices with N pairs of alternating quantum well (QW) electron confinement layers and barrier layers (BLs). Electron confinement requires that $d_{QW} \approx \lambda_f$, and phonon blocking occurs when $d_{QW} \approx \lambda_{mfp}$.

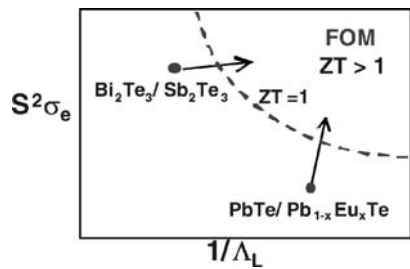


Figure 7 Increases in FOMs of PbTe/PbEuTe superlattices [14] produced primarily by reducing the lattice conductivity Λ_L and in BiTe/SbTe superlattices [24] produced primarily by increasing the electronic power factor ratio $S^2\sigma_e$.

as Snoek’s relation [25]

$$\mu_{dc} \cdot f_{res}^n > c, \tag{12}$$

where μ_{dc} is the static magnetic permeability, f_{res} is the ferromagnetic resonance (FMR) frequency, and n is an exponent that varies with the form of magnetic

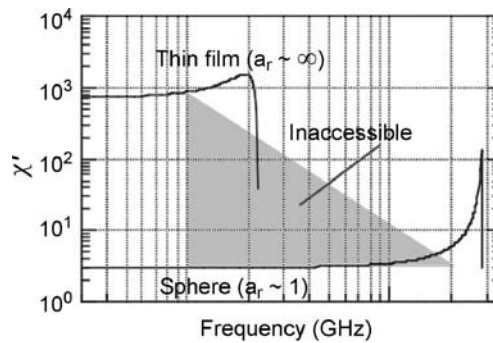


Figure 8 Frequency range in which the optimum magnetic susceptibility-bandwidth product cannot be achieved with either thin films or spheres. This FOM can be achieved below ~ 1 GHz with the intrinsic anisotropy of thin films, but requires a shape anisotropy of a metaparticle at higher frequencies.

anisotropy. An important class of magnetic metamaterials, discussed in Section 4, is based on the microwave permeabilities of thin films with uniform planar saturation magnetization $4\pi M_s$, and uniaxial anisotropy H_k , for which $n = 2$. Compared with ferrites, these films have many outstanding characteristics for applications in the microwave regime, including high magnetizations in the range of 10 kG–20 kG. However, their small anisotropy fields of less than ~ 50 Oe limit their maximum FMR to frequencies below ~ 1 GHz. As a consequence, they are prevented from achieving an optimum FOM (from Snoek's relation) in the 1 to 20 GHz frequency range, as shown in Figure 8 [26].

A metaparticle with a large shape anisotropy was designed to circumvent this problem; see Figure 9. The optimum shape was calculated using the Stoner-Wohlfarth model for the magnetic susceptibility of a single domain thin film, and Kittel's resonance equation with shape demagnetization calculated from Osborn's equation [27]; see Ref. 26 for computational details. The experimental results of Figure 10 show that these metaparticles exhibit the theoretically maximum FOM at all frequencies in the targeted range.

In principle, the range of FMR frequencies that are realizable by lithographed shape anisotropy is comparable with that observed in crystalline materials. An important difference is that the ferromagnetic films can be economically mass processed at temperatures that are integration-compatible with that of most materials.

4 Electromagnetic metamaterials

4.1 Conventional electromagnetic composites

Although the metamaterial design rules can be used to synthesize all types of electromagnetic composites, only linear mediums are considered here so that a clear exposition of the basic concepts and their usefulness can be made without unnecessary complexity. At present, effective medium theory (EMT) is used to guide

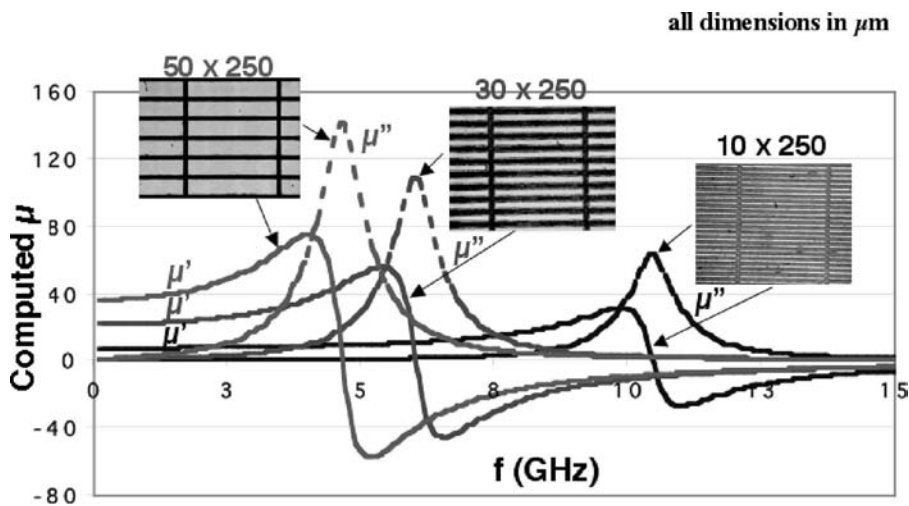


Figure 9 Lithographed thin magnetic film metaparticles and their calculated FMR spectrums [28]. The films were $1.4 \mu\text{m}$ thick and had a $4\pi M_s \approx 10 \text{ kG}$. The easy magnetic axis dimension for all the metaparticles was $250 \mu\text{m}$, while the hard axis dimensions were 10, 30 and $50 \mu\text{m}$ as shown.

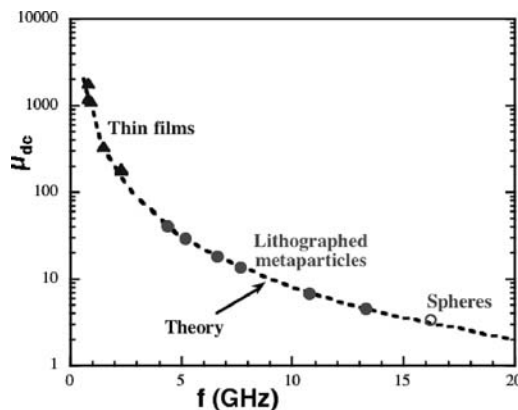


Figure 10 FMR frequency f_r and μ_{dc} computed from saturation magnetization and anisotropy for thin films (filled triangles), lithographed particles (filled circles), and spheres (hollow circle). In all cases the saturation magnetization is 10 kG . The lithographed films were 1 and $5 \mu\text{m}$ thick with easy axis = $250 \mu\text{m}$. The hard axis dimensions were 10, 20, 30, 40, 50 and $60 \mu\text{m}$. Theoretically maximum FOM (dotted line) is that for a single domain, thin magnetic film. Lithographed metaparticles provide the optimum FOM for all frequencies by merely changing the geometry of the lithographic mask.

the synthesis of most electromagnetic composites. The assumptions of EMT are compatible with the rules of macroscopic composites (see Table 1), except that the *local-field* approximation used in the EMT violates Rule 4A. This approximation is an ingenious idea to eliminate the necessity of making a self-consistent many-body calculation of the effective dielectric constant of a lattice of dipoles [29]. The

most general form of the EMT is the Lorentz-Lorenz equation

$$\frac{P_C - P_M}{P_C + 2P_M} = f_A \frac{P_A - P_M}{P_A + 2P_M} + f_B \frac{P_B - P_M}{P_B + 2P_M}, \quad (13)$$

which relates the composite constitutive parameter P_C ($P_C = \mu$ or ε) to the properties and volume fractions of the spherical particle constituents A and B, and the contiguous matrix M. Other commonly used forms of the EMT, such as the ones named after Bruggeman and Maxwell Garnett, can be derived from (13) as special cases [30]. Numerous studies of EMT have been made to determine the bounds on the range of achievable composite properties compared to those of the constituents. However, for reasons noted earlier, the EMT composite properties can never exceed those established by the law of mixtures.

This is illustrated by solving the Bruggeman form of the EMT [obtained from (13) by considering the matrix to be the composite itself] for the composite property. The physically real solution can be written in the form

$$P_C = \frac{1}{4} \left[(2P_+ - P_-) + \sqrt{(2P_+ - P_-)^2 + 8P_A P_B} \right], \quad (14)$$

where $P_+ = f_A P_A + f_B P_B$ and $P_- = f_A P_B + f_B P_A$. For the cases of interest, the term $8P_A P_B$ is small compared to the term $(2P_+ - P_-)^2$ under the radical in (14). Thus, the composite property is given approximately by

$$P_C \approx \frac{1}{2} (2P_+ - P_-) < P_+ = f_A P_A + f_B P_B. \quad (15)$$

Equation (15) shows that unless there are strong interactions between constituents, the composite property calculated from the Bruggeman EMT is always smaller than the volumetric average property. Similar conclusions can be reached by analyzing the other forms of EMT. The negative deviation from the law of mixtures can be very large when shape depolarizing or demagnetizing factors [27] are included.

4.2 The need for electromagnetic metamaterials

Figure 11 contrasts the materials available (above ~ 100 MHz) to microwave designers interested in a variety of problems. Displayed in this way, this figure shows how limited the material selection is, and how important electromagnetic composites should be to this field. However, the discussion in Section 4.1 suggests that conventional electromagnetic composites may not prove very useful in extending the range of available material properties. The actual situation is really much worse, as the following example will show.

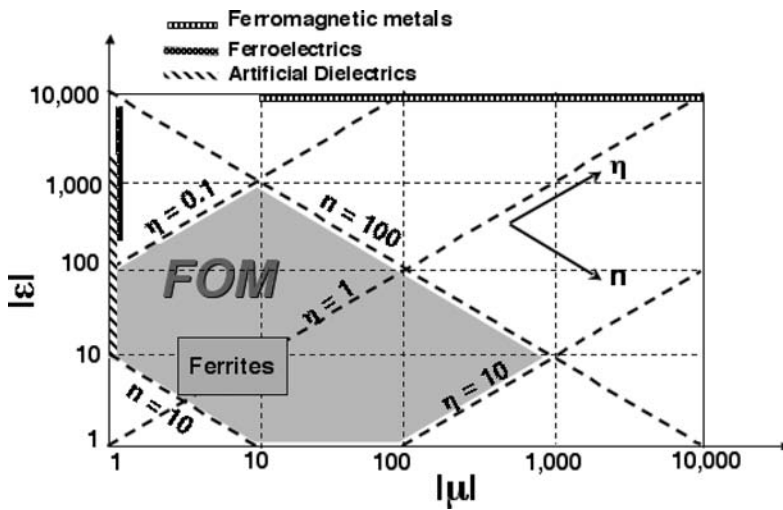


Figure 11 Constitutive parameter chart with overlays of lines of constant refractive index $n = (\mu\epsilon)^{1/2}$ and constant wave impedance $\eta = (\mu/\epsilon)^{1/2}$. All quantities are normalized with respect to free space (or vacuum). The inlays show the approximate range of conventional microwave materials that are currently available. The box labeled FOM straddles the $\eta = 1$ line and indicates the range of parameters needed for advances in many application areas.

Because of their large index of refraction $n = (\mu\epsilon)^{1/2}$,¹ there is considerable interest in fabricating tunable phase shifters with ferroelectric materials to reduce the microwave device size [31-32]. However, to match the input impedance of these devices over a practical bandwidth is quite difficult and requires a rather heroic design effort. Ref. 23 discusses the difficulty involved. The specifications for the material would require the wave impedance to be an acceptable fraction of that of free space, and the index of refraction to be as large as possible. The FOM for this device lies in the region so labeled in Figure 11 and, for this example application, it would require the addition of a tunability condition.

Initially it appears that properties satisfying this FOM could be realized by fabricating a composite with ferroelectric and ferrite particles. A research effort to accomplish this was recently reported [33] in which various composites with varying volume fractions of barium titanate ceramic powders ($\epsilon = 1300$ and $\mu = 1$ at 10 MHz) were copressed with nickel zinc ferrite powders ($\epsilon = 15$ and $\mu = 200$ at 10 MHz). The measurement results are compared in Figure 11 to a theoretical calculation made using the Maxwell Garnett EMT equation

$$P_C = P_M \frac{(P_P + 2P_M) + 2f_P(P_P - P_M)}{(P_P + 2P_M) - f_P(P_P - P_M)}, \quad (16)$$

where $P = \epsilon_c$ (or μ_c) and the subscripts C , M , and P designate, respectively, the composite, matrix and particulate phases. In the Maxwell Garnett EMT, either the barium titanate (BT) or ferrite (NZ) could be designated as the matrix or particle

¹ μ and ϵ are the permeability and the permittivity relative to vacuum's.

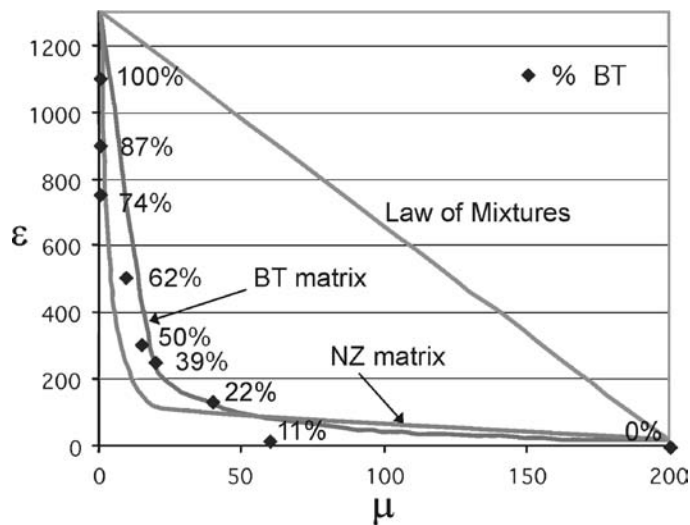


Figure 12 Comparison of measured (at 10 MHz) and computed constitutive parameters (μ , ε) for a ferroelectric, barium titanate (BT)—ferrite, nickel-zinc, (NZ) composite [33]. Computations were made using the Maxwell Garnett EMT for the cases in which the matrix and particle phases were interchanged.

phases. Both cases were computed and shown against the experimental data in Figure 12.

Figure 11 shows that despite the reasonably good agreement with the EMT, the composite falls far short of a volumetric average. The results also show that the failure is intrinsic to this type of electromagnetic composite when the matrix and particle phases have a large disparity in the values of their properties. In such cases, the microstructure is critical and, due to reciprocal weighting effects, a small volume fraction of the phase with the lowest value of μ or ε can produce a large change in the composite property. This effect is easily observed in Figure 11 as the large drops from the initial maximum values on the axes. This phenomenon severely limits the design possibilities for electromagnetic composites.

The failure of researchers working in the area of electromagnetic composites to note the enormous deviation of the performance of composites from the law of mixtures can be attributed to the fact that they have been mainly concerned with verifying the agreement of the individual properties with EMT and with bounds on parameter variation.

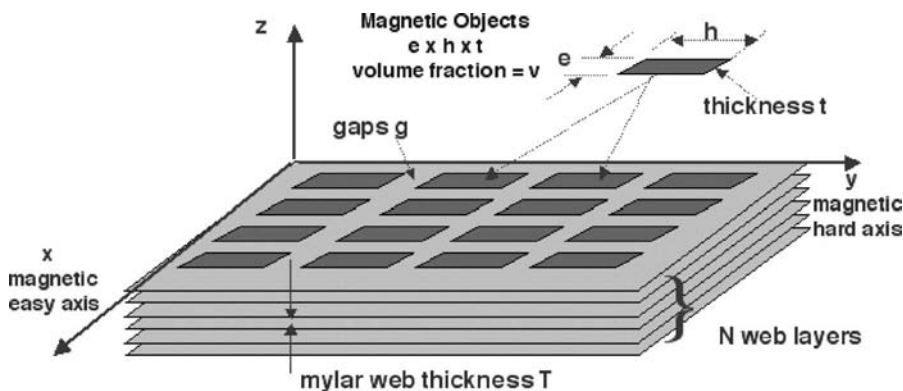
4.3 Metamaterial electromagnetic composites

There are no materials occurring in nature that have the constitutive parameters required by the FOM in Figure 11. Furthermore, the discussions in Sections 4.1 and 4.2 indicate that it is unlikely that this deficiency can be overcome using conventional macroscopic electromagnetic composites.

The design of a metamaterial composite for this purpose begins with the cellular architecture.

- For the example in Section 4.2, the FOM could be achieved for plane wave propagation in a specific direction, with cubic cells oriented with axes along the electric field $\mathbf{E}$, the magnetic field $\mathbf{H}$ and the wave propagation vector $\mathbf{k}$.
- The geometry, orientation and properties of the metaparticles in each cell are selected to allow each of the properties required by the FOM to be independently varied.
- The metaparticle geometry and orientation are selected to give the maximum response with the minimum depolarization and demagnetization. This was clearly the goal of the thin-film metaparticle design in Section 4.2.
- To eliminate retardation, all cellular dimensions are much smaller than the wavelength internal to the composite, and the metaparticle dimensions in the direction of propagation are small compared to the wavelength internal to the metaparticle.
- The metaparticle density and geometry are selected to maximize metal particle interaction to increase the magnitude of each property.

These constraints generally require a lowering of the connectivity [34] of the conventional electromagnetic effective medium to a so-called 3-0 metamaterial composite, wherein the metaparticles are isolated in a three-dimensionally connected matrix. An architecture and cell structure satisfying these requirements is shown in Figure 13. The dielectric and magnetic properties, dimensions, and geometry of each of the objects in the metaparticle have been selected to maximize the FOM. In practice, ferromagnetic thin films have proved to be more useful than ferrites in the synthesis of electromagnetic metamaterials [28]. These advantages are derived from their larger magnetizations, large and adjustable uniaxial mag-



$$\mu = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \mu_y & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad \varepsilon = \varepsilon_{\text{mylar}} \begin{bmatrix} \varepsilon_x & 0 & 0 \\ 0 & \varepsilon_y & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Figure 13 Electromagnetic metamaterial: laminated composite consisting of N mylar substrates containing magnetic $\sim 10\text{--}20\%$ volume fraction of thin-film metaparticles.

netic anisotropies (see Section 4.2), single magnetic domain configuration, small magnetic damping, and ease of processing. The calculated FMR spectrums for a $\sim 1.4\text{-}\mu\text{m}$ -thick ferromagnetic film are shown in Figure 8. These spectrums are generally in very good agreement with measured data. (Ref. 19 gives the details on the processing, composition, and microwave characteristics of a ferromagnetic film metaparticle.) Since the magnetic film is metallic, the exclusion of the electric field to its surrounding volume enhances the cell capacitance [35], making it unnecessary to include a separate dielectric object in the metaparticle cell.

The original ideas for these metamaterials were advanced over three decades ago [36-37], but only recently made possible by advances in processing techniques. Metamaterial composites of the type shown in Figure 13 have been fabricated by laminating polymer webs metallized with patterned, magnetic alloy thin films. These composites have several levels of anisotropy, including that of the cellular architecture, metaparticle geometry and the uniaxial anisotropy of the thin films.

The architectural rules for synthesizing this metamaterial are essentially identical to fabricating a discrete transmission line, and can be easily extended to include multiple, orthogonal TEM, TE and/or TM waves [38]. For analysis, the metaparticles are replaced by their discrete equivalent circuit elements. The synthesis can, in principle, be extended to include nonlocal excitations; in fact, transmission line modeling is routinely used to analyze photonic crystals. The synthesis of nonelectromagnetic metamaterials can be accommodated by replacing the governing equations with their circuit analogs and proceeding as above. Obviously, this approach will require an in-depth understanding of the actual material properties and those of their discrete electrical analogs.

The discrete transmission equivalent circuit for the metamaterial with the cell in Figure 13 is shown in Figure 14. Using the geometry and volume fraction shown, effective μ and ε can be calculated for the composite. The results are shown in Figure 15 along with assumed material and geometry parameters. Figure 15 shows that there is a wide frequency range in which the μ and ε of the composite can be

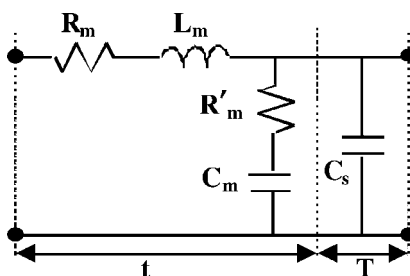


Figure 14 Discrete transmission line equivalent circuit of metaparticle cell shown in Figure 13. All the parameters are per unit length. For the magnetic film: $L_m = a_f \cdot \text{Re}[\mu_m]$, $R_m = 2\pi f \cdot a_f \cdot \text{Im}[\mu_m]$, and $R'_m = \rho_m e/h$. For the substrate: $C_s = \varepsilon_s$ and $C_m \approx \varepsilon_{\text{eff}} \cdot [e/[g + (T/\pi) \ln 4]]$, where t , T = film and substrate thickness; e , g = easy axis and gap dimensions; ε_{eff} , ε_s = permittivity of gap and substrate; a_f , v_f = film area and volume fractions; and μ_m , ρ_m = film permeability and resistivity.

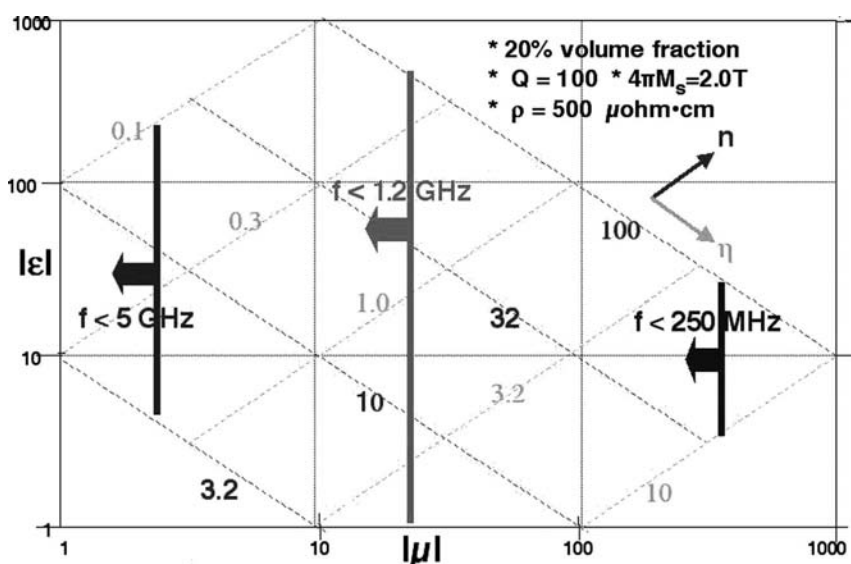


Figure 15 Calculated constitutive parameters for the metamaterial shown in Figure 12. The vertical lines corresponding to specific frequencies indicate that it is theoretically possible to design a metamaterial to have any combination of parameters to the left of the bar. The decrease in this space at higher frequencies is due to the onset of eddy current losses that would increase the material quality factor Q below the assumed value of 100 (loss tangent above 0.01). The inset gives the magnetic film properties and volume fraction used in the calculations.

independently varied by varying the geometry. Metamaterials synthesized by this approach have properties that are more than adequate to satisfy the FOM.

5 Conclusions

The usefulness of the metamaterial composites has now been demonstrated and holds the promise of making valuable contributions to extending and improving the performance of materials engineering in many application areas. This chapter represents the fledgling attempt to structure this emerging discipline and focus attention on the power of this approach in order to make dramatic advances in the performance of composites for almost any application area. The metaparticle concept that was much of the focus of this chapter integrates the advances made in many areas of device engineering and nanotechnology, not all of which can be utilized practically for synthesizing macroscopic materials. The scope was intentionally limited to a discussion of fairly simple, linear metamaterials that could be designed without exploiting nonlocal interactions. Metamaterials that incorporate nonlocal interactions already have a significant role in the design of electromagnetic composites and are destined for a greatly expanded one.

Since we are currently blessed with a plethora of novel candidate materials and a wealth of understanding of low-dimensional physics, it can be expected that the success of metamaterials will be paced in the immediate future by innovative

design as well as by advances in processing technology. The electromagnetic metamaterials discussed in Section 4.3 have some remarkably interesting and previously unobserved properties including: (i) the largest index of refraction of a low-loss material ever observed in nature, (ii) the possibility of synthesizing compatible isorefractive and isoimpedant multilayers, (iii) a large range of shape anisotropy fields and FMR frequencies, (iv) lower magnetic damping than polycrystalline ferrites, (v) low temperature integration compatibility, and (vi) a density approximately one third that of ferrites.

Acknowledgments

With much gratitude the author acknowledges many useful conversations with A.P. Valanju and P.M. Valanju, and the funding provided by State of Texas Advanced Technology Development and Transfer Program Grant #003658-0849-1999, DOD-ARPA Grant #MDA-972-98-1-0009 and the Industry Associates of the Center for Electromagnetic Materials and Devices.

References

- [1] G.B. Olson, Designing a new material world, *Science* **288** (2000), 993–998.
- [2] G.E. Moore, Cramming more components onto integrated circuits, *Electronics* **38** (8) (1965), 114–117 (April 19 issue).
- [3] P.A. Pincus, Chair, J. Bock, N. Clark, P. Eisenberger, E.A. Evans, L.W. Jelinski, C.M. Knobler, R. Nuzzo, J.M. Schnur, E. Thomas and D. Tirrell, *Biomolecular self-assembling materials: Scientific and technological frontiers*, Panel on Biomolecular Materials, National Research Council, National Academy of Sciences Press, Washington, DC, USA, 1996, 24–50.
- [4] R.M. Walser, Metamaterials: What are they? What are they good for? *Bull Am Phys Soc* **45** (2000), 1005.
- [5] L.J. Broutman and R.K. Krock, *Modern composites*, Addison-Wesley, Reading, MA, USA, 1967, 7–13.
- [6] J.W. Cahn, Spinodal decomposition, *Trans Metall Soc AIME* **242** (1968), 166–180.
- [7] J.E. Hilliard, Spinodal decomposition, *Phase transformations* (H.I. Aaronson and M.H. Cohen, eds), American Society for Metals, Metals Park, OH, USA, 1968, 497–560.
- [8] S. Rengarajan and R.M. Walser, High-speed fiber-optic sensor for magnetic-field mapping, *J Appl Phys* **81** (1997), 4278–4280.
- [9] S. Rengarajan, E.J. Yun, W.S. Kang and R.M. Walser, Effect of intermixing on the magnetic properties of $\text{Co}_{50}\text{Fe}_{50}/\text{Ni}_{80}\text{Fe}_{20}$ multilayers, *J Appl Phys* **81** (1997), 4761–4763.
- [10] E. Quandt and A. Ludwig, Giant magnetostrictive multilayers, *J Appl Phys* **85** (1999), 6232–6237.

- [11] E.F. Kneller and R. Hawig, The exchange-spring magnet: A new material principle for permanent magnets, *IEEE Trans Magn* **27** (1991), 3558–3600.
- [12] J.S. Jiang, E.E. Fullerton, C.H. Sowers, S.D. Bader, A.J. Shapiro, R.D. Shull, V.S. Gornakov and V.I. Nikitenko, Spring magnet films, *IEEE Trans Magn* **35** (1999), 3229–3234.
- [13] J.D. Joannopoulos, R.D. Meade and J.N. Winn, *Photonic crystals*, Princeton University Press, Princeton, NJ, USA, 1995.
- [14] J. Davies, *The physics of low dimensional semiconductors: An introduction*, Cambridge University Press, Cambridge, UK, 1997.
- [15] M.J. Kelley, *Low dimensional semiconductors: Materials, technology, devices, physics*, Oxford University Press, Oxford, UK, 1997.
- [16] K. Fuchs, The conductivity of thin metallic films according to the theory of metals, *Proc Camb Phil Soc* **34** (1938), 100–108.
- [17] J.F. Nye, *Physical properties of crystals*, Oxford University Press, Oxford, UK, 1967, 207–212.
- [18] H.B. Callen, *Thermodynamics*, Wiley, New York, NY, USA, 1960, Chap. 16.
- [19] G.A. Slack, New materials and performance limits for thermoelectric cooling, *CRC handbook of thermoelectrics* (D.M. Rowe, ed), CRC Press, Boca Raton, FL, USA, 1994.
- [20] G. Mahan, B. Sales and J. Sharp, Thermoelectric materials: New approaches to an old problem, *Phys Today* **50** (3) (1997), 42–47 (March issue).
- [21] L.D. Hicks and M.S. Dresselhaus, Effect of quantum wells on the thermoelectric figure of merit, *Phys Rev B* **47** (1993), 12727–12731.
- [22] L.E. DuBois, An introduction to the DARPA program in advanced thermoelectric materials and devices, *Proc 18th Int Conf Thermoelectrics*, Baltimore, MD, USA, 1999 (2000), 1–4.
- [23] M.S. Dresselhaus, Y.M. Lin, G. Dresselhaus, X. Sun, S.B. Cronin, T. Koga and J.Y. Ying, Advances in 1D and 2D thermoelectric materials, *Proc 18th Int Conf Thermoelectrics*, Baltimore, MD, USA, 1999 (2000), 92–99.
- [24] R. Venkatasubramanian, E. Sivola, T. Colpitts and B. O’Quinn, Phonon-blocking electron-transmitting structures, *Proc 18th Int Conf Thermoelectrics* 30, Baltimore, MD, USA, 1999 (2000), 100–103.
- [25] J.L. Snoek, Dispersion and absorption on magnetic ferrites at frequencies above 1 Mc, *Physica* **14** (1948), 204–207.
- [26] R.M. Walser, W. Win and P.M. Valanju, Shape-optimized ferromagnetic particles with maximum theoretical microwave susceptibility, *IEEE Trans Magn* **34** (1998), 1390–1392.
- [27] J.A. Osborn, Demagnetization factors of the general ellipsoid, *Phys Rev* **67** (1945), 351–370.
- [28] R.M. Walser, A.P. Valanju and W. Win, Materials issues in the application of ferromagnetic thin films in tunable high Q VHF-UHF devices, *Mater Res Soc Symp Proc* **63** (2000), 83–94.
- [29] C. Kittel, *Introduction to solid state physics*, 4th ed, Wiley, New York, NY, USA, 1971, 459.

- [30] D.E. Aspnes, Local field effects and effective-medium theory: a microscopic perspective, *Am J Phys* **50** (1982), 704–709.
- [31] R.R. Romanofsky, F.W. Van Keuls, J.D. Warner, C.H. Mueller, S.A. Alterovitz, F.A. Miranda and A.H. Qureshi, Analysis and optimization of thin film ferroelectric phase shifters, *Mater Res Soc Symp Proc* **63** (2000), 3–14.
- [32] J.B.L. Rao, D.P. Patel and V. Krichevsky, Voltage-controlled ferroelectric lens phased array, *IEEE Trans Antennas Propagat* **47** (1999), 458–468.
- [33] J.V. Mantese, A.L. Micheli, D.F. Dungan, R.G. Geyer, J. Baker-Jarvis and J. Grosvenor, Applicability of effective medium theory to ferroelectric/ferrimagnetic composites with composition and frequency-dependent complex permittivities and permeabilities, *J Appl Phys* **79** (1996), 1655–1660.
- [34] R.E. Newnham, D.P. Skinner and L.E. Cross, Connectivity and piezoelectric-pyroelectric composites, *Mat Res Bull* **13** (1978), 525–536. [The scheme denoting the connectivity of two-phase composites discussed in this paper was used to designate the metamaterial in Figure 12 as a 3-0 composite.]
- [35] R.E. Collin, *Field theory of guided waves*, McGraw-Hill, New York, NY, USA, 1960, 515–516, 538–540.
- [36] R.M. Walser, *A study of thin film magnetodielectrics*, Ph.D. Dissertation, University of Michigan, Ann Arbor, MI, USA, 1967.
- [37] R.M. Walser and R.J. Hach, *Thin film magnetodielectric materials*, U.S. Patent 3,540,047 Nov. 10 (1970).
- [38] C. Christopoulos, *The transmission line modeling method TLM*, IEEE Press, Piscataway, NJ, USA, 1995.



Rodger M. Walser is the Bowen Endowed Professor in Engineering, Professor of Electrical and Computer Engineering, and Director of the Center for Electromagnetic Materials and Devices at the University of Texas at Austin. His current research interests are in the areas of thin-film magnetism and micromagnetic sensors and devices; physics, processing, and characterization of electromagnetic materials and devices; and microwave magnetic materials and devices. He is the inventor of metamaterials, a new materials science paradigm that is now being extensively used by numerous groups throughout

the scientific community to guide the synthesis of high-performance materials and devices. He is the founder, President and CEO of MetaMaterials LLC, Austin, TX. He was previously founder, President and Chief Scientist at the Ardex Corp., Austin, TX; founder and Research Engineer at the KMS Industries and Conduction Inc., both in Ann Arbor, MI. He has 11 years of industrial experience; has published more than 150 archival research papers, more than 160 technical papers and reports; and has made more than 240 research presentations.

Homogenization of Linear and Nonlinear Complex Composite Materials

Tom G. Mackay

Abstract

A composite medium may be characterized as an effectively homogeneous medium, provided electromagnetic wavelengths are sufficiently long as compared with length scales of the component phase inhomogeneities. The estimation of the constitutive parameters of homogenized composite mediums is a matter of considerable scientific and technological significance. A survey of recent developments in the homogenization of complex composite mediums is presented. Specifically, the implementations of the Maxwell Garnett formalism and the Bruggeman formalism—both well established for simple composites—are presented for bianisotropic composite mediums. Recent incremental and differential refinements to the Maxwell Garnett formalism are discussed. The strong-property-fluctuation theory (SPFT) provides an approach to homogenization that—in contrast to the conventional approaches of Maxwell Garnett and Bruggeman—takes more detailed account of the distributional statistics of the component phases. The SPFT for linear bianisotropic mediums and weakly nonlinear isotropic chiral mediums is presented. The issues of SPFT convergence and nonlinearity enhancement are discussed.

1 Introduction

The electromagnetic response of composite material is a topic of long-standing and continuing scientific interest [1]–[3]. Indeed, since we inhabit an atomistic world, this issue has universal importance. In general, a comprehensive electromagnetic description of complex composite materials lies beyond the reaches of present-day theory. However, considerable advances may be achieved for linear and weakly nonlinear composites, provided the length scales of the inhomogeneities are sufficiently short compared with the electromagnetic wavelengths. In this long-wavelength regime, the theory proceeds through characterizing the non-homogeneous composite as an effectively homogeneous medium [4]. Estimation of the constitutive properties of such a homogenized composite medium (HCM)—by means of *homogenization formalisms*—is the subject of this chapter.

Homogenization has attracted the attention of scientists since the earliest days of electromagnetic theory. For example, Arago and Biot introduced a “mixture formula” for inert gases in the early nineteenth century [1], [5]. The establishment of the Maxwell equations enabled more rigorous approaches to homogenization to be followed. Two formalisms in particular, the Maxwell Garnett formalism and the Bruggeman formalism, were developed in the early twentieth century and have since been widely applied [1], [6]. While homogenization formalisms applicable to relatively simple HCMs have been extensively studied, recent advances in the materials sciences relating to complex and unconventional composite mediums have highlighted the necessity for more advanced formalisms [2].

Within the past few years, the scope of homogenization formalisms has been significantly advanced on three fronts: Firstly, the Maxwell Garnett formalism and the Bruggeman formalism have been established for the most general class of linear HCMs, namely bianisotropic HCMs [7–8]. A key mathematical component, exploited in both formalisms, is the *depolarization dyadic*; i.e., the singularity of the dyadic Green function [9]. The construction of the depolarization dyadic for bianisotropic mediums by Michel and Weiglhofer [10] paved the way for the development of the corresponding Maxwell Garnett and Bruggeman formalisms. Secondly, the incremental [11] and differential [12] refinements to the Maxwell Garnett formalism, which overcome the dilute composite limitation, have been constructed for bianisotropic HCMs. Thirdly, an approach to homogenization that takes detailed account of the distributional statistics of the component phases has been developed. In conventional approaches to homogenization, as typified by the Maxwell Garnett and Bruggeman formalisms, the distributional statistics of the component phases are described solely via their respective volume fractions. A more sophisticated approach is provided by the *strong-property-fluctuation theory* (SPFT), in which spatial correlation functions—of arbitrarily high order—may be accommodated [13]. The second-order SPFT for linear bianisotropic HCMs is well established [14]–[16]. Furthermore, convergence at the level of the second-order SPFT has been demonstrated via numerical studies [17]. The SPFT has also been recently extended into the weakly nonlinear regime [18]–[20].

A survey of recent developments in the homogenization of complex composite mediums is presented in this chapter. The emphasis is placed on providing essential mathematical formulas and illustrative numerical results; for detailed derivations, the reader is referred to original scientific papers. A description of the general homogenization framework to be considered is provided in Section 2. The Maxwell Garnett and Bruggeman approaches to homogenization are outlined in Section 3; the incremental and differential Maxwell Garnett formalisms are also presented and illustrated by means of a representative numerical example. In Section 4 the SPFT for linear bianisotropic HCMs is described; the second- and third-order theories, as well as the issue of convergence, are discussed. Finally, the SPFT for weakly nonlinear HCMs is considered in Section 5, wherein the phenomenon of nonlinearity enhancement is also pursued.

The following notation is adopted: 6-vectors (3-vectors) are in bold (normal) face and underlined, whereas 6×6 (3×3) dyadics are in bold (normal) face and underlined twice. The adjoint, determinant, inverse, trace, transpose and $\alpha\beta$ th entry of a dyadic $\underline{\underline{\mathbf{Q}}}$ are represented by $\text{adj}(\underline{\underline{\mathbf{Q}}})$, $\det \underline{\underline{\mathbf{Q}}}$, $\underline{\underline{\mathbf{Q}}}^{-1}$, $\text{tr} \underline{\underline{\mathbf{Q}}}$, $\underline{\underline{\mathbf{Q}}}^T$ and $[\underline{\underline{\mathbf{Q}}}]_{\alpha\beta}$, respectively. For dyadics and vectors, the dot product denotes contraction of indexes. The ensemble average of a quantity N is written as $\langle N \rangle$. The 6×6 (3×3) identity dyadic is denoted by $\underline{\underline{\mathbf{I}}}$ ($\underline{\underline{\mathbf{I}}}$). The permittivity and permeability of free space (i.e., vacuum) are given as ϵ_0 and μ_0 , respectively; $k_0 = \omega\sqrt{\epsilon_0\mu_0}$ is the free-space wavenumber, while ω is the angular frequency and an $\exp(-i\omega t)$ time dependence is implicit.

2 Preliminaries

2.1 Component phases

The focus in this chapter is on HCMs arising from two component phases. The component phases are designated as a and b . All space is assumed to be partitioned into the disjoint regions V_a and V_b , which contain the phases a and b , respectively. We consider component phases characterized by linear bianisotropic constitutive relations, given in the frequency domain as [21]

$$\underline{\underline{\mathbf{C}}}(\underline{\underline{\mathbf{r}}}) = \underline{\underline{\mathbf{K}}}_\ell \cdot \underline{\underline{\mathbf{F}}}(\underline{\underline{\mathbf{r}}}), \quad \underline{\underline{\mathbf{r}}} \in V_\ell, \quad (\ell = a, b), \quad (1)$$

with the 6-vector electromagnetic fields being

$$\underline{\underline{\mathbf{C}}}(\underline{\underline{\mathbf{r}}}) = \begin{bmatrix} \underline{\underline{\mathbf{D}}}(\underline{\underline{\mathbf{r}}}) \\ \underline{\underline{\mathbf{B}}}(\underline{\underline{\mathbf{r}}}) \end{bmatrix}, \quad \underline{\underline{\mathbf{F}}}(\underline{\underline{\mathbf{r}}}) = \begin{bmatrix} \underline{\underline{\mathbf{E}}}(\underline{\underline{\mathbf{r}}}) \\ \underline{\underline{\mathbf{H}}}(\underline{\underline{\mathbf{r}}}) \end{bmatrix}. \quad (2)$$

The complex-valued 6×6 constitutive dyadic

$$\underline{\underline{\mathbf{K}}}_\ell = \begin{bmatrix} \underline{\underline{\epsilon}}_\ell & \underline{\underline{\xi}}_\ell \\ \underline{\underline{\zeta}}_\ell & \underline{\underline{\mu}}_\ell \end{bmatrix}, \quad (\ell = a, b), \quad (3)$$

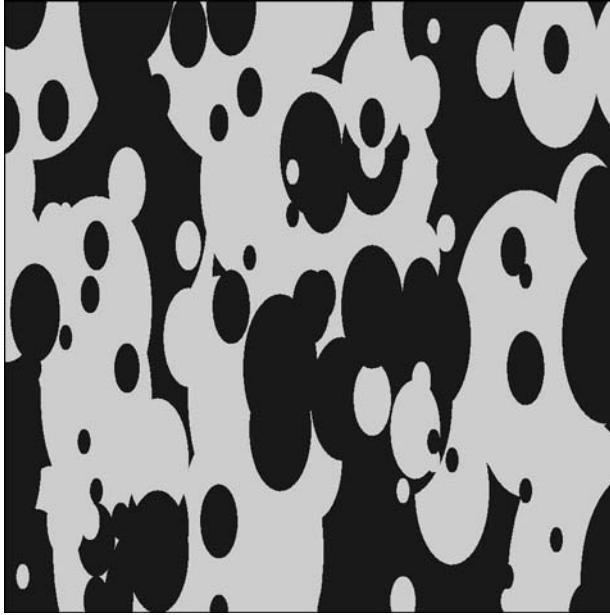


Figure 1 A schematic representation of a random two-phase composite medium based on conformal, identically oriented ellipsoidal particles.

comprises the 3×3 permittivity and permeability dyadics $\underline{\underline{\epsilon}}_\ell$ and $\underline{\underline{\mu}}_\ell$, respectively, as well as the magnetoelectric dyadics $\underline{\underline{\xi}}_\ell$ and $\underline{\underline{\zeta}}_\ell$.

The component phases are envisioned as random particulate dispersals. For each component phase, the constituent particles are taken to be identically oriented, conformal ellipsoids, as schematically illustrated in Figure 1. The shape dyadic

$$\underline{\underline{U}} = \frac{1}{\sqrt[3]{U_x U_y U_z}} \text{diag}(U_x, U_y, U_z), \quad (U_x, U_y, U_z > 0), \quad (4)$$

parameterizes the particulate surfaces as

$$\underline{r}_e(\theta, \phi) = \eta \underline{\underline{U}} \cdot \hat{\underline{r}}(\theta, \phi), \quad (5)$$

where $\hat{\underline{r}}(\theta, \phi)$ is the radial unit vector specified by the spherical polar coordinates θ and ϕ . The linear ellipsoidal dimensions, as determined by η , are assumed to be sufficiently small that the electromagnetic long-wavelength regime pertains.

The component phase distributions are specified in terms of the characteristic functions

$$\Phi_\ell(\underline{r}) = \begin{cases} 1, & \underline{r} \in V_\ell, \\ 0, & \underline{r} \notin V_\ell, \end{cases} \quad (\ell = a, b). \quad (6)$$

Specifically, statistical moments of the characteristic functions are employed: the n th moment is the expectation value $\langle \Phi_\ell(\underline{r}_1) \cdots \Phi_\ell(\underline{r}_n) \rangle$ and represents the probability of $\underline{r}_1, \dots, \underline{r}_n \in V_\ell$ ($\ell = a, b$). The volume fraction of phase ℓ , namely f_ℓ , is given by the first statistical moment of Φ_ℓ ; i.e., $\langle \Phi_\ell(\underline{r}) \rangle = f_\ell$. Clearly, $f_a + f_b = 1$. In conventional approaches to homogenization, such as the Maxwell Garnett and Bruggeman approaches, the component phase distributions are characterized in terms of only the first moment of $\Phi_\ell(\underline{r})$. A more sophisticated approach is adopted in the strong-property-fluctuation theory (SPFT) [13-14], wherein statistical correlation functions of arbitrarily high order may be utilized to describe the component phase distributions.

2.2 Depolarization and polarizability dyadics

A mathematical construct of key importance in homogenization formalisms is the singularity of the unbounded dyadic Green function, i.e., the *depolarization dyadic* $\underline{\underline{D}}^{U/k}$ [8]. It provides the electromagnetic response of a $\underline{U}$ -shaped exclusion region, embedded in a homogeneous background medium with 6×6 constitutive dyadic $\underline{\underline{K}}_k$, in the limit $\eta \rightarrow 0$. Note that the background medium may be either of the component phases (i.e., $k = a$ or b) or an effectively homogeneous medium. For a $\underline{U}$ -shaped exclusion volume in the bianisotropic medium with 3×3 constitutive dyadics $\underline{\underline{\epsilon}}_k, \underline{\underline{\xi}}_k, \underline{\underline{\zeta}}_k$ and $\underline{\underline{\mu}}_k$, the corresponding depolarization dyadic is given as [10]

$$\underline{\underline{D}}^{U/k} = \begin{bmatrix} \underline{\underline{D}}_{ee}^{U/k} & \underline{\underline{D}}_{em}^{U/k} \\ \underline{\underline{D}}_{me}^{U/k} & \underline{\underline{D}}_{mm}^{U/k} \end{bmatrix}, \quad (7)$$

where

$$\underline{\underline{D}}_{\lambda\lambda'}^{U/k} = \underline{\underline{U}}^{-1} \cdot \underline{\underline{D}}_{\lambda\lambda'}^{U/k} \cdot \underline{\underline{U}}^{-1}, \quad (\lambda, \lambda' = e, m), \quad (8)$$

and

$$\underline{\underline{D}}_{\lambda\lambda'}^{U/k} = \frac{1}{4\pi i\omega} \int_{\phi=0}^{2\pi} \int_{\theta=0}^{\pi} \frac{(\hat{\underline{q}} \cdot \underline{\underline{\tau}}_{\lambda\lambda'} \cdot \hat{\underline{q}}) \hat{\underline{q}} \hat{\underline{q}} \sin \theta d\theta d\phi}{(\hat{\underline{q}} \cdot \underline{\underline{\epsilon}}_k \cdot \hat{\underline{q}})(\hat{\underline{q}} \cdot \underline{\underline{\mu}}_k \cdot \hat{\underline{q}}) - (\hat{\underline{q}} \cdot \underline{\underline{\xi}}_k \cdot \hat{\underline{q}})(\hat{\underline{q}} \cdot \underline{\underline{\zeta}}_k \cdot \hat{\underline{q}})}, \quad (9)$$

with $\hat{\underline{q}} = (\sin \theta \cos \phi, \sin \theta \sin \phi, \cos \theta)$ and

$$\begin{aligned} \underline{\underline{v}}_k &= \underline{\underline{U}}^{-1} \cdot \underline{\underline{v}}_k \cdot \underline{\underline{U}}^{-1}, \quad (v = \epsilon, \xi, \zeta, \mu), \\ \underline{\underline{\tau}}_{ee} &= \underline{\underline{\mu}}_k, \quad \underline{\underline{\tau}}_{em} = -\underline{\underline{\xi}}_k, \quad \underline{\underline{\tau}}_{me} = -\underline{\underline{\zeta}}_k, \quad \underline{\underline{\tau}}_{mm} = \underline{\underline{\epsilon}}_k. \end{aligned} \quad (10)$$

Explicit representations for $\underline{\underline{D}}^{U/k}$ can be found in cases where the constitutive dyadics possess certain symmetries [9-10], [22], but the double integrals in (9) gen-

erally require numerical evaluation. Parenthetically, it is remarked that depolarization dyadics associated with cylindrical exclusion volumes in anisotropic mediums have been established recently [23-24]. These may be used to develop homogenization formalisms appropriate to composites based on needle- or disk-shaped particulate geometries; however, this matter will not be considered here.

Consider an ellipsoidal particle with shape dyadic $\underline{\underline{U}}^p$, made of a medium characterized by the constitutive dyadic $\underline{\underline{K}}_p$. Let the particle be embedded in the medium with constitutive dyadic $\underline{\underline{K}}_q$. The corresponding generalized *polarizability* density dyadic

$$\underline{\underline{\alpha}}^{p/q} = (\underline{\underline{K}}_p - \underline{\underline{K}}_q) \cdot [\underline{\underline{I}} + i\omega \underline{\underline{D}}^{U^p/q} \cdot (\underline{\underline{K}}_p - \underline{\underline{K}}_q)]^{-1}, \quad (11)$$

is a convenient quantity in homogenization formalisms.

3 Conventional approaches to homogenization

3.1 Maxwell Garnett formalism

The Maxwell Garnett homogenization formalism has been used extensively, despite its applicability being limited to dilute composites [1]. The Maxwell Garnett estimate of the HCM constitutive dyadic is [7]

$$\underline{\underline{K}}_{MG} = \underline{\underline{K}}_b + f_a \underline{\underline{\alpha}}^{a/b} \cdot (\underline{\underline{I}} - i\omega f_a \underline{\underline{D}}^{I/b} \cdot \underline{\underline{\alpha}}^{a/b})^{-1}, \quad (12)$$

where $\underline{\underline{D}}^{I/b}$ is the depolarization dyadic associated with a spherical exclusion volume immersed in the component phase b . The estimate (12) is valid for $f_a < 0.3$ only. Note that if $\underline{\underline{D}}^{I/b}$ in (12) were to be replaced by $\underline{\underline{D}}^{U^a/b}$, then the Bragg-Pippard formalism would arise [25].

3.2 Bruggeman formalism

The Bruggeman homogenization formalism has the advantage over the Maxwell Garnett formalism of being applicable for all volume fractions $f_a \in (0, 1)$ [6]. A characteristic feature of this formalism is that the component phases a and b are treated symmetrically. For bianisotropic HCMs, arising from particulate component phases a and b with ellipsoidal topology characterized by the shape dyadics $\underline{\underline{U}}^a$ and $\underline{\underline{U}}^b$, respectively, the Bruggeman estimate of the HCM constitutive dyadic $\underline{\underline{K}}_{Br}$ is obtained by solving the nonlinear equation [7]

$$f_a \underline{\underline{\alpha}}^{a/Br} + f_b \underline{\underline{\alpha}}^{b/Br} = \underline{\underline{0}}. \quad (13)$$

The usual method of dealing with (13) is to apply the simple Jacobi technique [8], by which an iterative solution is developed as

$$\underline{\underline{\mathbf{K}}}_{Br}[n] = \mathcal{T} \left\{ \underline{\underline{\mathbf{K}}}_{Br}[n-1] \right\}, \quad (n = 1, 2, \dots), \quad (14)$$

with the initial value $\underline{\underline{\mathbf{K}}}_{Br}[0] = \underline{\underline{\mathbf{K}}}_{MG}$. The action of the operator $\mathcal{T}$ is defined by

$$\begin{aligned} \mathcal{T} \left\{ \underline{\underline{\mathbf{K}}}_{Br} \right\} = & \left\{ f_a \underline{\underline{\mathbf{K}}}_a \cdot \left[\underline{\underline{\mathbf{I}}} + i\omega \underline{\underline{\mathbf{D}}}^{U^a/Br} \cdot \left(\underline{\underline{\mathbf{K}}}_a - \underline{\underline{\mathbf{K}}}_{Br} \right) \right]^{-1} \right. \\ & \left. + f_b \underline{\underline{\mathbf{K}}}_b \cdot \left[\underline{\underline{\mathbf{I}}} + i\omega \underline{\underline{\mathbf{D}}}^{U^b/Br} \cdot \left(\underline{\underline{\mathbf{K}}}_b - \underline{\underline{\mathbf{K}}}_{Br} \right) \right]^{-1} \right\} \\ & \cdot \left\{ f_a \left[\underline{\underline{\mathbf{I}}} + i\omega \underline{\underline{\mathbf{D}}}^{U^a/Br} \cdot \left(\underline{\underline{\mathbf{K}}}_a - \underline{\underline{\mathbf{K}}}_{Br} \right) \right]^{-1} \right. \\ & \left. + f_b \left[\underline{\underline{\mathbf{I}}} + i\omega \underline{\underline{\mathbf{D}}}^{U^b/Br} \cdot \left(\underline{\underline{\mathbf{K}}}_b - \underline{\underline{\mathbf{K}}}_{Br} \right) \right]^{-1} \right\}^{-1}. \end{aligned} \quad (15)$$

3.3 Recent developments: incremental and differential Maxwell Garnett formalisms

The *incremental* Maxwell Garnett homogenization formalism is a recent refinement of the Maxwell Garnett formalism in which the HCM is built incrementally, by adding the component phase a to the component phase b not all at once but in a fixed number (N) of stages [11-12]. After each increment, the composite is homogenized using the Maxwell Garnett formalism. Thereby, the incremental Maxwell Garnett formalism cures one of the major limitations of the Maxwell Garnett formalism, namely the restriction of its applicability to dilute composites. We have the iteration scheme

$$\underline{\underline{\mathbf{K}}}[n+1] = \underline{\underline{\mathbf{K}}}[n] + \varrho \underline{\underline{\alpha}}^{a/n} \cdot \left(\underline{\underline{\mathbf{I}}} - i\omega \varrho \underline{\underline{\mathbf{D}}}^{I/n} \cdot \underline{\underline{\alpha}}^{a/n} \right)^{-1} \quad (16)$$

for $n = 0, 1, \dots, N-1$, where we set $\varrho = 1 - f_b^{1/N}$ and $\underline{\underline{\mathbf{K}}}[0] = \underline{\underline{\mathbf{K}}}_b$. In (16), $\underline{\underline{\alpha}}^{a/n}$ is the polarizability dyadic of a component phase a particle, of shape specified by $\underline{\underline{U}}$, relative to a medium with constitutive dyadic $\underline{\underline{\mathbf{K}}}[n]$; i.e.,

$$\underline{\underline{\alpha}}^{a/n} = \left(\underline{\underline{\mathbf{K}}}_a - \underline{\underline{\mathbf{K}}}[n] \right) \cdot \left[\underline{\underline{\mathbf{I}}} + i\omega \underline{\underline{\mathbf{D}}}^{U/n} \cdot \left(\underline{\underline{\mathbf{K}}}_a - \underline{\underline{\mathbf{K}}}[n] \right) \right]^{-1}, \quad (17)$$

while $\underline{\underline{\mathbf{D}}}^{I/n}$ ($\underline{\underline{\mathbf{D}}}^{U/n}$) is the depolarization dyadic of a spherical ($\underline{\underline{U}}$ -shaped) exclusion region in the medium with constitutive dyadic $\underline{\underline{\mathbf{K}}}[n]$. The incremental Maxwell Garnett estimate of the HCM constitutive dyadic,

$$\underline{\underline{\mathbf{K}}}_{IMG} = \underline{\underline{\mathbf{K}}}[N], \quad (18)$$

follows after N iterations.

The *differential* Maxwell Garnett homogenization formalism arises from the corresponding incremental formalism in the limit $N \rightarrow \infty$ [8], [12]. That is, the differential Maxwell Garnett estimate of the HCM constitutive dyadic $\underline{\underline{\mathbf{K}}}_{DMG}$ is given as the solution of the ordinary differential equation

$$\frac{\partial}{\partial \nu} \underline{\underline{\mathbf{K}}}(\nu) = \frac{1}{1 - \nu} \underline{\underline{\alpha}}^{a/\nu}, \quad (19)$$

with initial value $\underline{\underline{\mathbf{K}}}(0) = \underline{\underline{\mathbf{K}}}_b$. Here, ν is a continuous variable representing the volume fraction of component phase a material added to component phase b , and $\underline{\underline{\alpha}}^{a/\nu}$ is the polarizability dyadic of a $\underline{\underline{U}}$ -shaped particle composed of component phase a relative to a medium with constitutive dyadic $\underline{\underline{\mathbf{K}}}(\nu)$. The particular solution of (19) satisfying $\nu = f_a$ is the differential Maxwell Garnett estimate; i.e.,

$$\underline{\underline{\mathbf{K}}}_{DMG} = \underline{\underline{\mathbf{K}}}(f_a). \quad (20)$$

3.4 A numerical example

A representative numerical example is presented to illustrate the homogenization formalisms of Sections 3.1–3.3: Consider the homogenization scenario wherein component phase b is a uniaxial dielectric medium with 3×3 permittivity dyadic $\underline{\underline{\epsilon}}_b = \epsilon_0 \text{diag}(\epsilon^b, \epsilon^b, \epsilon_z^b)$, while component phase a is an isotropic dielectric medium with relative permittivity ϵ^a . Spherical component phase topology is assumed. The permittivity dyadic of the corresponding HCM has the uniaxial form $\epsilon_{HCM} = \epsilon_0 \text{diag}(\epsilon^{HCM}, \epsilon^{HCM}, \epsilon_z^{HCM})$. The permittivity of free space is denoted by ϵ_0 .

For the parameter values $\epsilon^b = 1.2$, $\epsilon_z^b = 4$ and $\epsilon^a = 9$, the HCM constitutive parameter values, calculated using the Maxwell Garnett (including incremental and differential variants) and the Bruggeman formalisms, are plotted in Figure 2 as a function of the volume fraction f_a . The constitutive parameters calculated using the four different formalisms are in close agreement at low values of f_a (i.e., for $f_a < 0.25$, say). However, as the volume fraction of component phase a increases, the constitutive parameter estimates of the Maxwell Garnett and Bruggeman formalisms are seen to diverge. Furthermore, the discrepancies between the Maxwell Garnett estimates and the incremental and differential Maxwell Garnett estimates are observed to increase as f_a increases. These observations are consistent with expectation as the Maxwell Garnett formalism is valid only for dilute HCMs. The incremental Maxwell Garnett estimates presented in Figure 2 were computed with $N = 5$; the observed close agreement between the incremental and differential Maxwell Garnett constitutive parameter values indicate that the incremental Maxwell Garnett iteration scheme converges rapidly. For all values of volume fraction $f_a \in (0, 1)$, the incremental and differential Maxwell Garnett estimates are bounded by the the Maxwell Garnett and Bruggeman estimates of the HCM constitutive parameters.

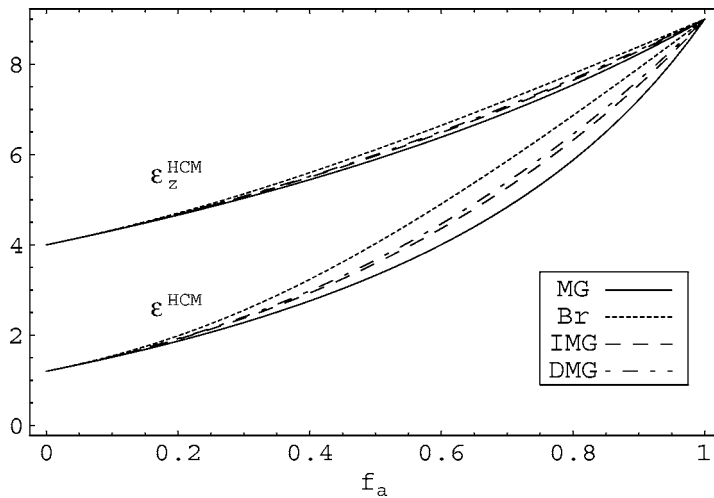


Figure 2 The HCM relative permittivity parameters, calculated using the Maxwell Garnett (MG), Bruggeman (Br), incremental MG (IMG) and differential MG (DMG) formalisms, plotted against the volume fraction f_a . The IMG algorithm was implemented with $N = 5$.

The characteristics of the homogenization formalisms of Sections 3.1–3.3 reflected in Figure 2 are typical of those observed for more complex HCMs; further details may be found elsewhere [8], [11–12].

4 SPFT homogenization

4.1 Generalities

The strong-property-fluctuation theory (SPFT) provides an approach to homogenization based on the iterative refinement of a *comparison medium* [14]; i.e., a homogeneous medium specified by the constitutive dyadic

$$\underline{\underline{\mathbf{K}}}_{cm} = \begin{bmatrix} \underline{\underline{\epsilon}}_{cm} & \underline{\underline{\xi}}_{cm} \\ \underline{\underline{\zeta}}_{cm} & \underline{\underline{\mu}}_{cm} \end{bmatrix}. \quad (21)$$

Unlike the conventional approaches to homogenization outlined in Section 3, the SPFT accommodates a comprehensive description of the distributional statistics of the component phases [26]. Thereby, coherent scattering losses may be accounted for. Successive SPFT iterates recruit successively higher-order spatial correlation functions. The topologies of both component phases a and b are assumed to be characterized by the shape dyadic $\underline{\underline{U}}$.

The n th-order SPFT estimate of the constitutive dyadic of the HCM is obtained as [17]

$$\underline{\underline{\mathbf{K}}}_{Dy}^{[n]} = \underline{\underline{\mathbf{K}}}_{cm} - \frac{1}{i\omega} \left(\underline{\underline{\mathbf{I}}} + \underline{\underline{\Sigma}}^{[n]} \cdot \underline{\underline{\mathbf{D}}}^{U/cm} \right)^{-1} \cdot \underline{\underline{\Sigma}}^{[n]}, \quad (22)$$

wherein the subscript Dy on $\underline{\underline{\mathbf{K}}}_{Dy}^{[n]}$ derives from the central equation in the SPFT, namely the *Dyson equation* [27]. The depolarization dyadic $\underline{\underline{\mathbf{D}}}^{U/cm}$ relates to a U -shaped exclusion particle, embedded in the comparison medium. The *mass operator* term $\underline{\underline{\Sigma}}^{[n]}$ —which has an infinite series representation—is defined in terms of the dyadic Green function $\underline{\underline{\mathbf{G}}}_{cm}(\underline{\underline{R}})$ of the comparison medium, where $\underline{\underline{R}} = \underline{\underline{r}} - \underline{\underline{r'}}$, together with the normalized polarizability density dyadic $\underline{\underline{\chi}}_{\ell} = -i\omega \underline{\underline{\alpha}}_{\ell/cm}$ ($\ell = a, b$). In the derivation of (22), the long-wavelength approximation is invoked.

4.2 Degenerate cases

The lowest-order truncation of the mass operator series, i.e., $\underline{\underline{\Sigma}}^{[0]} = \underline{\underline{\mathbf{0}}}$, gives the trivial result $\underline{\underline{\mathbf{K}}}_{Dy}^{[0]} = \underline{\underline{\mathbf{K}}}_{cm}$. The first-order mass operator approximation is conventionally chosen by demanding that [14], [28]

$$\left\langle \underline{\underline{\chi}}_a \Phi_a(\underline{\underline{r}}) + \underline{\underline{\chi}}_b \Phi_b(\underline{\underline{r}}) \right\rangle = \underline{\underline{\mathbf{0}}}, \quad (23)$$

in order to ensure the elimination of certain secular terms. Correspondingly, $\underline{\underline{\Sigma}}^{[1]} = \underline{\underline{\mathbf{0}}}$ and $\underline{\underline{\mathbf{K}}}_{Dy}^{[1]} = \underline{\underline{\mathbf{K}}}_{cm}$. Since (23) is equivalent to (13), we have $\underline{\underline{\mathbf{K}}}_{cm} \equiv \underline{\underline{\mathbf{K}}}_{Br}$; i.e., the lowest-order SPFT and the Bruggeman formalism provide identical estimates of the HCM constitutive dyadic.

4.3 Bilocal approximation

The second order SPFT—otherwise known as the bilocally approximated SPFT—has been widely implemented [14]–[16], [26], [28]. Under the bilocal approximation, a two-point covariance function

$$\begin{aligned} \Gamma(\underline{\underline{R}}) &= \langle \Phi_a(\underline{\underline{r}}) \Phi_a(\underline{\underline{r'}}) \rangle - \langle \Phi_a(\underline{\underline{r}}) \rangle \langle \Phi_a(\underline{\underline{r'}}) \rangle \\ &= \langle \Phi_b(\underline{\underline{r}}) \Phi_b(\underline{\underline{r'}}) \rangle - \langle \Phi_b(\underline{\underline{r}}) \rangle \langle \Phi_b(\underline{\underline{r'}}) \rangle, \end{aligned} \quad (24)$$

along with its associated correlation length L , characterize the distributional statistics of the component phases a and b . In particular, within a region of linear dimensions given by L , and of shape determined by the covariance function, the correlated responses of scattering centers result in an attenuation of the macroscopic coherent field. On the other hand, the responses of scattering centers separated by distances much greater than L are statistically independent.

The bilocally approximated mass operator term is given by [14]

$$\underline{\underline{\Sigma}}^{[2]} = (\underline{\underline{\chi}}_a - \underline{\underline{\chi}}_b) \cdot \underline{\underline{\mathbf{P}}}^{\Gamma} \cdot (\underline{\underline{\chi}}_a - \underline{\underline{\chi}}_b), \quad (25)$$

where

$$\underline{\underline{\mathbf{P}}}^\Gamma = \mathcal{P} \int \Gamma(\underline{\underline{\mathbf{R}}}) \underline{\underline{\mathbf{G}}}_{cm}(\underline{\underline{\mathbf{R}}}) d^3 \underline{\underline{\mathbf{R}}}. \quad (26)$$

The principal value integral (26) has been investigated for the following two-point covariance functions [16]

$$\Gamma_0(\underline{\underline{\mathbf{R}}}) = f_a f_b \left[1 - \mathcal{H}(|\underline{\underline{\mathbf{U}}}^{-1} \cdot \underline{\underline{\mathbf{R}}}| - \sigma) \right], \quad (27)$$

$$\Gamma_1(\underline{\underline{\mathbf{R}}}) = f_a f_b \exp(-\sigma |\underline{\underline{\mathbf{U}}}^{-1} \cdot \underline{\underline{\mathbf{R}}}|), \quad (28)$$

$$\Gamma_2(\underline{\underline{\mathbf{R}}}) = f_a f_b \exp(-\sigma |\underline{\underline{\mathbf{U}}}^{-1} \cdot \underline{\underline{\mathbf{R}}}|^2), \quad (29)$$

where $\mathcal{H}$ is the Heaviside function (i.e., $\mathcal{H}(x) = \int_{-\infty}^x \delta(s) ds$, with $\delta(s)$ being the Dirac delta function) and $\sigma > 0$. Each of the covariance functions (27)–(29) reflects the ellipsoidal topology of the component phases insofar as each is spherically symmetric with respect to the coordinate frame of $\underline{\underline{\mathbf{U}}}^{-1} \cdot \underline{\underline{\mathbf{R}}}$. Evaluations of (26) with the covariances (27)–(29) are presented in Appendix 1 for the special case where the comparison medium belongs to the class of reciprocal biaxial bianisotropic mediums.

Numerical investigations have demonstrated that the choice of covariance function has only a secondary influence on the SPFT estimates of the HCM constitutive dyadic [16]. The Heaviside function covariance (27)—which may be physically motivated [29]—has been used extensively in recent studies [14]–[17].

4.4 Trilocal approximation

The issue of SPFT convergence is addressed through considering the third-order approximation to the mass operator. The calculation of the *trilocally approximated* mass operator term $\underline{\underline{\Sigma}}^{[3]}$ requires a three-point covariance function [17], [30–31].

Consider the following natural generalization of the two-point Heaviside function covariance (27) to

$$\langle \theta_a(\underline{\underline{\mathbf{r}}}) \theta_a(\underline{\underline{\mathbf{r}}}') \theta_a(\underline{\underline{\mathbf{r}}}'') \rangle = \begin{cases} f_a^3 & \min\{L_{12}, L_{13}, L_{23}\} > L \\ f_a & \max\{L_{12}, L_{13}, L_{23}\} \leq L \\ \frac{1}{3}(f_a + 2f_a^3) & \text{one of } L_{12}, L_{13}, L_{23} \leq L \\ \frac{1}{3}(2f_a + f_a^3) & \text{two of } L_{12}, L_{13}, L_{23} \leq L \end{cases}, \quad (30)$$

where

$$L_{12} = |\underline{\underline{\mathbf{r}}} - \underline{\underline{\mathbf{r}}}'|, \quad L_{13} = |\underline{\underline{\mathbf{r}}} - \underline{\underline{\mathbf{r}}}''|, \quad L_{23} = |\underline{\underline{\mathbf{r}}}' - \underline{\underline{\mathbf{r}}}''|. \quad (31)$$

In addition, let us restrict attention to isotropic chiral comparison mediums [32–33] based on spherical topology. Then, the corresponding trilocally approximated mass operator term is given by [17]

$$\begin{aligned} \underline{\underline{\Sigma}}^{[3]} = \underline{\underline{\Sigma}}^{[2]} + \frac{f_a(1-2f_a)}{3(1-f_a)^2} \underline{\underline{\chi}}_a \cdot \left(\underline{\underline{\mathbf{V}}} \cdot \underline{\underline{\chi}}_a \cdot \underline{\underline{\mathbf{P}}}^{\Gamma_0} \right. \\ \left. + \underline{\underline{\mathbf{P}}}^{\Gamma_0} \cdot \underline{\underline{\chi}}_a \cdot \underline{\underline{\mathbf{V}}} + \underline{\underline{\mathbf{P}}}^{\Gamma_0} \cdot \underline{\underline{\chi}}_a \cdot \underline{\underline{\mathbf{P}}}^{\Gamma_0} \right) \cdot \underline{\underline{\chi}}_a, \end{aligned} \quad (32)$$

where $\underline{\underline{\Sigma}}^{[2]}$ is the bilocally approximated mass operator term associated with the two-point covariance function (27), and

$$\underline{\underline{\mathbf{V}}} = \frac{1}{i\omega} \underline{\underline{\mathbf{K}}}^{-1}_{cm} - \underline{\underline{\mathbf{D}}}^{I/cm}. \quad (33)$$

4.5 Numerical results

We illuminate the zeroth-, second- and third-order SPFT approximations presented in Sections 4.2–4.4 by means of a representative numerical example. Consider the homogenization of an isotropic chiral [32–33] component phase a , based on spherical topology, with a component phase b which is taken to be free space. The constitutive relations for component phase a are

$$\underline{\underline{\epsilon}}_a = \epsilon_0 \epsilon^a \underline{\underline{\mathbf{I}}}, \quad \underline{\underline{\xi}}_a = -\underline{\underline{\zeta}}_a = i\sqrt{\epsilon_0 \mu_0} \xi^a \underline{\underline{\mathbf{I}}}, \quad \underline{\underline{\mu}}_a = \mu_0 \mu^a \underline{\underline{\mathbf{I}}}; \quad (34)$$

the parameter values $\epsilon^a = \delta(4 + i2)$, $\xi^a = \delta(2 + i1.5)$, and $\mu^a = \delta(2.5 + i)$ are selected. The parameter $\delta = 10, 20$ and 30 provides the means to vary the constitutive contrast between the component phases. The volume fraction $f_a = 0.3$ and the angular frequency $\omega = 2\pi \times 10^{10} \text{ rad s}^{-1}$ are fixed. The permeability of free space is denoted by μ_0 .

The corresponding HCM is an isotropic chiral medium with constitutive relations

$$\underline{\underline{\epsilon}}_{Dy} = \epsilon_0 \epsilon^{Dy} \underline{\underline{\mathbf{I}}}, \quad \underline{\underline{\xi}}_{Dy} = -\underline{\underline{\zeta}}_{Dy} = i\sqrt{\epsilon_0 \mu_0} \xi^{Dy} \underline{\underline{\mathbf{I}}}, \quad \underline{\underline{\mu}}_{Dy} = \mu_0 \mu^{Dy} \underline{\underline{\mathbf{I}}}. \quad (35)$$

The constitutive parameters ϵ^{Dy} , ξ^{Dy} and μ^{Dy} , calculated using the zeroth-, second- and third-order SPFTs, are plotted as functions of the correlation length L in Figure 3. The values from the zeroth-order theory—which are identical to those values calculated using the Bruggeman homogenization formalism—are independent of correlation length. Also, the calculated values for all orders of approximation coincide at $L = 0$. The second- and third-order estimates of the constitutive parameters are seen to provide a significant correction to the corresponding lowest-order estimates at non-zero correlation lengths.

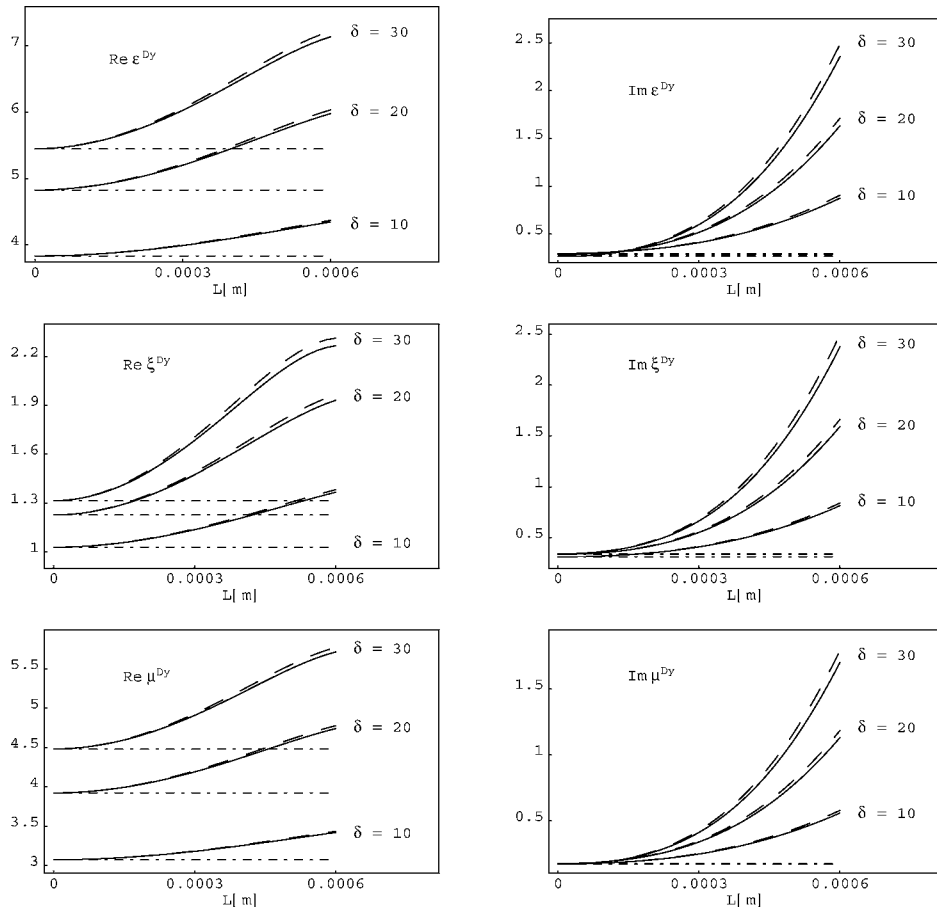


Figure 3 Real (left) and imaginary (right) parts of the HCM constitutive parameters, calculated using the zeroth-, second- and third-order mass operator approximations, plotted against correlation length L for $\delta = 10, 20, 30$. Key: broken dashed lines indicate zeroth-order values; solid lines indicate second-order values; dashed lines indicate third-order values.

Under the long-wavelength approximation, we require $Q \ll 1$, where

$$Q = \frac{\max\{|\gamma^+|, |\gamma^-|\}}{2\pi} L, \quad (36)$$

and $\gamma^\pm$ denote the left- and right-handed wavenumbers in the isotropic chiral comparison medium [33]. For the example illustrated in Figure 3, $Q = 0.1$ at $L = 5.1 \times 10^{-4}$ m for $\delta = 30$; at $L = 5.7 \times 10^{-4}$ m for $\delta = 20$; and at $L = 7.3 \times 10^{-4}$ m for $\delta = 10$. These limits establish the applicability ranges of the presented formalism. Thus, we may infer that in the long-wavelength regime, the SPFT converges at the level of the bilocal approximation for isotropic chiral HCMs. Further studies have demonstrated SPFT convergence at the level of the bilocal approximation for chiroferrite mediums that are both weakly uniaxial and weakly gyrotropic [17].

5 Weakly nonlinear regime

5.1 Generalities

The SPFT for weakly nonlinear isotropic chiral HCMs [20] is presented in this section. Consider the HCM arising from the following two component phases [34]: Component phase a is a cubically nonlinear, isotropic dielectric medium with constitutive dyadic

$$\underline{\underline{\mathbf{K}}}_a = \begin{bmatrix} \epsilon_0 (\epsilon^{a0} + \epsilon^{a1} |\underline{E}_a|^2) \underline{\underline{I}} & \underline{\underline{0}} \\ \underline{\underline{0}} & \mu_0 \underline{\underline{I}} \end{bmatrix} = \underline{\underline{\mathbf{K}}}_{a0} + \underline{\underline{\mathbf{K}}}_{a1} |\underline{E}_a|^2, \quad (37)$$

where ϵ^{a0} and ϵ^{a1} are the relative linear and nonlinear permittivities, respectively, and $\underline{E}_a$ is the electric field in phase a . Weak nonlinearity is assumed; i.e., $\epsilon_{a0} \gg \epsilon_{a1} |\underline{E}_a|^2$. Such electrostrictive mediums as characterized by (37) are capable of inducing Brillouin scattering, which is often a strong process [35]. Component phase b is a linear isotropic chiral medium specified by

$$\underline{\underline{\mathbf{K}}}_b = \begin{bmatrix} \epsilon_0 \epsilon^b \underline{\underline{I}} & i \sqrt{\epsilon_0 \mu_0} \xi^b \underline{\underline{I}} \\ -i \sqrt{\epsilon_0 \mu_0} \xi^b \underline{\underline{I}} & \mu_0 \mu^b \underline{\underline{I}} \end{bmatrix}. \quad (38)$$

Both component phases a and b are assumed to be based on spherical topology.

The corresponding comparison medium is a weakly nonlinear isotropic chiral medium characterized by [20]

$$\begin{aligned} \underline{\underline{\mathbf{K}}}_{cm} &= \underline{\underline{\mathbf{K}}}_{cm0} + \underline{\underline{\mathbf{K}}}_{cm1} |\underline{E}_{HCM}|^2 \\ &= \begin{bmatrix} \epsilon_0 \epsilon_{cm0} \underline{\underline{I}} & i \sqrt{\epsilon_0 \mu_0} \xi_{cm0} \underline{\underline{I}} \\ -i \sqrt{\epsilon_0 \mu_0} \xi_{cm0} \underline{\underline{I}} & \mu_0 \mu_{cm0} \underline{\underline{I}} \end{bmatrix} \\ &\quad + \begin{bmatrix} \epsilon_0 \epsilon_{cm1} \underline{\underline{I}} & i \sqrt{\epsilon_0 \mu_0} \xi_{cm1} \underline{\underline{I}} \\ -i \sqrt{\epsilon_0 \mu_0} \xi_{cm1} \underline{\underline{I}} & \mu_0 \mu_{cm1} \underline{\underline{I}} \end{bmatrix} |\underline{E}_{HCM}|^2, \end{aligned} \quad (39)$$

where $\underline{E}_{HCM}$ is the spatially averaged electric field in the HCM. The depolarization dyadic associated with a spherical exclusion volume embedded in the comparison medium is given by the two-term Maclaurin series [20]

$$\begin{aligned} \underline{\underline{\mathbf{D}}}^{I/cm} &= \underline{\underline{\mathbf{D}}}_0^{I/cm} + \underline{\underline{\mathbf{D}}}_1^{I/cm} |\underline{E}_{HCM}|^2 \\ &= \frac{1}{3i\omega\sigma_1} \begin{bmatrix} \mu_0 \mu_{cm0} \underline{\underline{I}} & -i \sqrt{\epsilon_0 \mu_0} \xi_{cm0} \underline{\underline{I}} \\ i \sqrt{\epsilon_0 \mu_0} \xi_{cm0} \underline{\underline{I}} & \epsilon_0 \epsilon_{cm0} \underline{\underline{I}} \end{bmatrix} \\ &\quad + \frac{1}{3i\omega\sigma_1} \begin{bmatrix} d_1^{ee} \underline{\underline{I}} & d_1^{em} \underline{\underline{I}} \\ d_1^{me} \underline{\underline{I}} & d_1^{mm} \underline{\underline{I}} \end{bmatrix} |\underline{E}_{HCM}|^2, \end{aligned} \quad (40)$$

where

$$d_1^{ee} = \mu_0 (\mu_{cm1} - \mu_{cm0} \sigma_2), \quad (41)$$

$$d_1^{em} = -i\sqrt{\epsilon_0\mu_0} (\xi_{cm1} - \xi_{cm0} \sigma_2) = -d_1^{me}, \quad (42)$$

$$d_1^{mm} = \epsilon_0 (\epsilon_{cm1} - \epsilon_{cm0} \sigma_2), \quad (43)$$

$$\sigma_1 = \epsilon_0\mu_0 \left(\epsilon_{cm0}\mu_{cm0} - \xi_{cm0}^2 \right), \quad (44)$$

$$\sigma_2 = \frac{\epsilon_{cm0}\mu_{cm1} + \epsilon_{cm1}\mu_{cm0} - 2\xi_{cm0}\xi_{cm1}}{\epsilon_{cm0}\mu_{cm0} - \xi_{cm0}^2}. \quad (45)$$

Similarly, under the assumption of weak nonlinearity, the generalized polarizabilities $\underline{\underline{\chi}}_{a,b}$ may be expressed as the sum of linear and nonlinear terms as

$$\underline{\underline{\chi}}_\ell = \underline{\underline{\chi}}_{\ell 0} + \underline{\underline{\chi}}_{\ell 1} |\underline{E}_{HCM}|^2, \quad (\ell = a, b), \quad (46)$$

where

$$\underline{\underline{\chi}}_{a0} = -i\omega \left(\underline{\underline{\mathbf{K}}}_{a0} - \underline{\underline{\mathbf{K}}}_{Br0} \right) \cdot \underline{\underline{\mathbf{Y}}}_{a0}^{-1}, \quad (47)$$

$$\underline{\underline{\chi}}_{b0} = -i\omega \left(\underline{\underline{\mathbf{K}}}_b - \underline{\underline{\mathbf{K}}}_{Br0} \right) \cdot \underline{\underline{\mathbf{Y}}}_{b0}^{-1}, \quad (48)$$

$$\underline{\underline{\chi}}_{a1} = -i\omega \left[\left(g\underline{\underline{\mathbf{K}}}_{a1} - \underline{\underline{\mathbf{K}}}_{Br1} \right) \cdot \underline{\underline{\mathbf{Y}}}_{a0}^{-1} + \left(\underline{\underline{\mathbf{K}}}_{a0} - \underline{\underline{\mathbf{K}}}_{Br0} \right) \cdot \underline{\underline{\mathbf{S}}} \left\{ \underline{\underline{\mathbf{Y}}}_a \right\} \right], \quad (49)$$

$$\underline{\underline{\chi}}_{b1} = -i\omega \left[-\underline{\underline{\mathbf{K}}}_{Br1} \cdot \underline{\underline{\mathbf{Y}}}_{b0}^{-1} + \left(\underline{\underline{\mathbf{K}}}_b - \underline{\underline{\mathbf{K}}}_{Br0} \right) \cdot \underline{\underline{\mathbf{S}}} \left\{ \underline{\underline{\mathbf{Y}}}_b \right\} \right]. \quad (50)$$

Here the dyadic quantities

$$\begin{aligned} \underline{\underline{\mathbf{Y}}}_a &= \underline{\underline{\mathbf{Y}}}_{a0} + \underline{\underline{\mathbf{Y}}}_{a1} |\underline{E}_{HCM}|^2 \\ &= \left[\underline{\underline{\mathbf{I}}} + i\omega \underline{\underline{\mathbf{D}}}_0^{I/cm} \cdot \left(\underline{\underline{\mathbf{K}}}_{a0} - \underline{\underline{\mathbf{K}}}_{Br0} \right) \right] + i\omega \left[\underline{\underline{\mathbf{D}}}_0^{I/cm} \cdot \left(g\underline{\underline{\mathbf{K}}}_{a1} - \underline{\underline{\mathbf{K}}}_{Br1} \right) \right. \\ &\quad \left. + \underline{\underline{\mathbf{D}}}_1^{I/cm} \cdot \left(\underline{\underline{\mathbf{K}}}_{a0} - \underline{\underline{\mathbf{K}}}_{Br0} \right) \right] |\underline{E}_{HCM}|^2 \end{aligned} \quad (51)$$

and

$$\begin{aligned} \underline{\underline{\mathbf{Y}}}_b &= \underline{\underline{\mathbf{Y}}}_{b0} + \underline{\underline{\mathbf{Y}}}_{b1} |\underline{E}_{HCM}|^2 \\ &= \left[\underline{\underline{\mathbf{I}}} + i\omega \underline{\underline{\mathbf{D}}}_0^{I/cm} \cdot \left(\underline{\underline{\mathbf{K}}}_b - \underline{\underline{\mathbf{K}}}_{Br0} \right) \right] \\ &\quad + i\omega \left[-\underline{\underline{\mathbf{D}}}_0^{I/cm} \cdot \underline{\underline{\mathbf{K}}}_{Br1} + \underline{\underline{\mathbf{D}}}_1^{I/cm} \cdot \left(\underline{\underline{\mathbf{K}}}_b - \underline{\underline{\mathbf{K}}}_{Br0} \right) \right] |\underline{E}_{HCM}|^2, \end{aligned} \quad (52)$$

have been introduced, along with the dyadic function

$$\underline{\underline{\mathbf{S}}} \{ \underline{\underline{\mathbf{Y}}}_\ell \} = \frac{1}{\det \underline{\underline{\mathbf{Y}}}_{\ell 0}} \left[\text{adj} \left(\underline{\underline{\mathbf{Y}}}_{\ell 1} \right) - {}^t \{ \underline{\underline{\mathbf{Y}}}_\ell \} \underline{\underline{\mathbf{Y}}}_{\ell 0}^{-1} \right], \quad (53)$$

with

$${}^t \{ \underline{\underline{\mathbf{Y}}}_\ell \} = (Y_{\ell 0}^{ee} Y_{\ell 1}^{mm} - Y_{\ell 0}^{em} Y_{\ell 1}^{me}) + (Y_{\ell 1}^{ee} Y_{\ell 0}^{mm} - Y_{\ell 1}^{em} Y_{\ell 0}^{me}) \quad (54)$$

and

$$\underline{\underline{\mathbf{Y}}}_{\ell 0} = \begin{bmatrix} Y_{\ell 0}^{ee} \underline{\underline{\mathbf{I}}} & Y_{\ell 0}^{em} \underline{\underline{\mathbf{I}}} \\ Y_{\ell 0}^{me} \underline{\underline{\mathbf{I}}} & Y_{\ell 0}^{mm} \underline{\underline{\mathbf{I}}} \end{bmatrix}, \quad \underline{\underline{\mathbf{Y}}}_{\ell 1} = \begin{bmatrix} Y_{\ell 1}^{ee} \underline{\underline{\mathbf{I}}} & Y_{\ell 1}^{em} \underline{\underline{\mathbf{I}}} \\ Y_{\ell 1}^{me} \underline{\underline{\mathbf{I}}} & Y_{\ell 1}^{mm} \underline{\underline{\mathbf{I}}} \end{bmatrix}. \quad (55)$$

In the representation (49) of $\underline{\underline{\chi}}_{a1}$, the approximation

$$|\underline{\underline{E}}_a|^2 \approx g |\underline{\underline{E}}_{HCM}|^2, \quad (56)$$

is implicit, wherein the local field factor $g = |[\underline{\underline{\mathbf{Y}}}_{a0}^{-1}]_{11}|^2$ [20].

Correct to the second order in $|\underline{\underline{E}}_{HCM}|$, the n th-order SPFT estimate of the HCM constitutive dyadic is given in terms of the two-term mass operator expression

$$\underline{\underline{\Sigma}}^{[n]} = \underline{\underline{\Sigma}}_0^{[n]} + \underline{\underline{\Sigma}}_1^{[n]} |\underline{\underline{E}}_{HCM}|^2, \quad (57)$$

as

$$\begin{aligned} \underline{\underline{\mathbf{K}}}_{Dy}^{[n]} &= \underline{\underline{\mathbf{K}}}_{Dy0}^{[n]} + \underline{\underline{\mathbf{K}}}_{Dy1}^{[n]} |\underline{\underline{E}}_{HCM}|^2 \\ &= \begin{bmatrix} \epsilon_0, \epsilon_{Dy0}^{[n]} \underline{\underline{\mathbf{I}}} & i\sqrt{\epsilon_0 \mu_0} \xi_{Dy0}^{[n]} \underline{\underline{\mathbf{I}}} \\ -i\sqrt{\epsilon_0 \mu_0} \xi_{Dy0}^{[n]} \underline{\underline{\mathbf{I}}} & \mu_0 \mu_{Dy0}^{[n]} \underline{\underline{\mathbf{I}}} \end{bmatrix} \\ &\quad + \begin{bmatrix} \epsilon_0 \epsilon_{Dy1}^{[n]} \underline{\underline{\mathbf{I}}} & i\sqrt{\epsilon_0 \mu_0} \xi_{Dy1}^{[n]} \underline{\underline{\mathbf{I}}} \\ -i\sqrt{\epsilon_0 \mu_0} \xi_{Dy1}^{[n]} \underline{\underline{\mathbf{I}}} & \mu_0 \mu_{Dy1}^{[n]} \underline{\underline{\mathbf{I}}} \end{bmatrix} |\underline{\underline{E}}_{HCM}|^2. \end{aligned} \quad (58)$$

5.2 Degenerate cases

As in the linear theory, the lowest-order SPFT estimates are $\underline{\underline{\mathbf{K}}}_{Dy}^{[0]} = \underline{\underline{\mathbf{K}}}_{Dy}^{[1]} = \underline{\underline{\mathbf{K}}}_{cm}$ and are consistent with the Bruggeman homogenization formalism. The Jacobi iteration scheme delivers the p th iterate of $\underline{\underline{\mathbf{K}}}_{Dy}^{[0]}$ as [8]

$$\underline{\underline{\mathbf{K}}}_{Dy,j}^{[0]}[p] = \mathcal{T}_j \{ \underline{\underline{\mathbf{K}}}_{Dy,j}^{[0]}[p-1] \}, \quad (j = 0, 1), \quad (59)$$

in terms of the $(p - 1)$ th iterate, where $p = 1, 2, \dots$. Suitable initial values are provided by

$$\underline{\underline{\mathbf{K}}}_{Dy0}^{[0]}[0] = f_a \underline{\underline{\mathbf{K}}}_{a0} + f_b \underline{\underline{\mathbf{K}}}_b, \quad \underline{\underline{\mathbf{K}}}_{Dy1}^{[0]}[0] = f_a \underline{\underline{\mathbf{K}}}_{a1}, \quad (60)$$

while the operators $\mathcal{T}_{0,1}$ are defined as

$$\mathcal{T}_0 \left\{ \underline{\underline{\mathbf{K}}}_{Dy0}^{[0]} \right\} = \left(f_a \underline{\underline{\mathbf{K}}}_{a0} \cdot \underline{\underline{\mathbf{Y}}}_{a0}^{-1} + f_b \underline{\underline{\mathbf{K}}}_b \cdot \underline{\underline{\mathbf{Y}}}_{b0}^{-1} \right) \cdot \left(f_a \underline{\underline{\mathbf{Y}}}_{a0}^{-1} + f_b \underline{\underline{\mathbf{Y}}}_{b0}^{-1} \right)^{-1} \quad (61)$$

and

$$\begin{aligned} \mathcal{T}_1 \left\{ \underline{\underline{\mathbf{K}}}_{Dy1}^{[0]} \right\} = & \left\{ f_a \left[g \underline{\underline{\mathbf{K}}}_{a1} \cdot \underline{\underline{\mathbf{Y}}}_{a0}^{-1} + \left(\underline{\underline{\mathbf{K}}}_{a0} - \underline{\underline{\mathbf{K}}}_{Dy0}^{[0]} \right) \cdot \underline{\underline{\mathbf{S}}} \left\{ \underline{\underline{\mathbf{Y}}}_a \right\} \right] \right. \\ & \left. + f_b \left(\underline{\underline{\mathbf{K}}}_b - \underline{\underline{\mathbf{K}}}_{Dy0}^{[0]} \right) \cdot \underline{\underline{\mathbf{S}}} \left\{ \underline{\underline{\mathbf{Y}}}_b \right\} \right\} \cdot \left(f_a \underline{\underline{\mathbf{Y}}}_{a0}^{-1} + f_b \underline{\underline{\mathbf{Y}}}_{b0}^{-1} \right)^{-1}. \end{aligned} \quad (62)$$

5.3 Bilocal approximation

The weakly nonlinear SPFT for isotropic chiral HCMs has been established under the bilocal approximation for the Heaviside function two-point covariance (27) [20]. As in the linear theory, the second-order mass operator term is provided by (25). The associated principal value integral $\underline{\underline{\mathbf{P}}}^{\Gamma_0}$ may be expressed as the sum of linear and nonlinear parts as

$$\underline{\underline{\mathbf{P}}}^{\Gamma_0} = \frac{if_a f_b \omega L^2}{3} \left(\underline{\underline{\mathbf{P}}}^{\Gamma_0}_0 + \underline{\underline{\mathbf{P}}}^{\Gamma_0}_1 |E_{HCM}|^2 \right); \quad (63)$$

evaluations of $\underline{\underline{\mathbf{P}}}^{\Gamma_0}_0$ and $\underline{\underline{\mathbf{P}}}^{\Gamma_0}_1$ are presented in Appendix 2. The linear and nonlinear contributions to the bilocally approximated mass operator term are, respectively [20],

$$\underline{\underline{\mathbf{P}}}^{[2]}_0 = \frac{if_a f_b \omega L^2}{3} \left(\underline{\underline{\mathbf{x}}}_{a0} - \underline{\underline{\mathbf{x}}}_{b0} \right) \cdot \underline{\underline{\mathbf{P}}} \cdot \left(\underline{\underline{\mathbf{x}}}_{a0} - \underline{\underline{\mathbf{x}}}_{b0} \right) \quad (64)$$

and

$$\begin{aligned} \underline{\underline{\mathbf{P}}}^{[2]}_1 = & \frac{if_a f_b \omega L^2}{3} \left[\left(\underline{\underline{\mathbf{x}}}_{a0} - \underline{\underline{\mathbf{x}}}_{b0} \right) \cdot \underline{\underline{\mathbf{P}}} \cdot \left(\underline{\underline{\mathbf{x}}}_{a1} - \underline{\underline{\mathbf{x}}}_{b1} \right) \right. \\ & + \left(\underline{\underline{\mathbf{x}}}_{a0} - \underline{\underline{\mathbf{x}}}_{b0} \right) \cdot \underline{\underline{\mathbf{P}}}_1 \cdot \left(\underline{\underline{\mathbf{x}}}_{a0} - \underline{\underline{\mathbf{x}}}_{b0} \right) \\ & \left. + \left(\underline{\underline{\mathbf{x}}}_{a1} - \underline{\underline{\mathbf{x}}}_{b1} \right) \cdot \underline{\underline{\mathbf{P}}}_0 \cdot \left(\underline{\underline{\mathbf{x}}}_{a0} - \underline{\underline{\mathbf{x}}}_{b0} \right) \right]. \end{aligned} \quad (65)$$

Thus, the respective linear and nonlinear components of $\underline{\underline{\mathbf{K}}}_{Dy}^{[2]}$ arise as

$$\underline{\underline{\mathbf{K}}}_{Dy0}^{[2]} = \underline{\underline{\mathbf{K}}}_{Br0} - \frac{1}{i\omega} \underline{\underline{\mathbf{Q}}}^{-1} \cdot \underline{\underline{\tilde{\Sigma}}}_0^{[2]}, \quad (66)$$

$$\underline{\underline{\mathbf{K}}}_{Dy1}^{[2]} = \underline{\underline{\mathbf{K}}}_{Br1} - \frac{1}{i\omega} \left(\underline{\underline{\mathbf{Q}}}^{-1} \cdot \underline{\underline{\tilde{\Sigma}}}_1^{[2]} + \underline{\underline{\mathbf{S}}} \{ \underline{\underline{\mathbf{Q}}} \} \cdot \underline{\underline{\tilde{\Sigma}}}_0^{[2]} \right), \quad (67)$$

wherein the dyadic

$$\underline{\underline{\mathbf{Q}}} = \underline{\underline{\mathbf{Q}}}_0 + \underline{\underline{\mathbf{Q}}}_1 |E_{HCM}|^2, \quad (68)$$

with components

$$\underline{\underline{\mathbf{Q}}}_0 = \underline{\underline{\mathbf{I}}} + \underline{\underline{\tilde{\Sigma}}}_0^{[2]} \cdot \underline{\underline{\mathbf{D}}}_0^{I/cm}, \quad \underline{\underline{\mathbf{Q}}}_1 = \underline{\underline{\tilde{\Sigma}}}_0^{[2]} \cdot \underline{\underline{\mathbf{D}}}_1^{I/cm} + \underline{\underline{\tilde{\Sigma}}}_1^{[2]} \cdot \underline{\underline{\mathbf{D}}}_0^{I/cm}, \quad (69)$$

and inverse

$$\underline{\underline{\mathbf{Q}}}^{-1} = \underline{\underline{\mathbf{Q}}}_0^{-1} + \underline{\underline{\mathbf{S}}} \{ \underline{\underline{\mathbf{Q}}} \} |E_{HCM}|^2, \quad (70)$$

has been introduced.

5.4 Trilocal approximation

The trilocally approximated mass operator associated with the second-order covariance term (27) and third-order covariance term (30) has respective linear and nonlinear parts [20]

$$\underline{\underline{\Sigma}}_0^{[3]} = \underline{\underline{\Sigma}}_0^{[2]} + \frac{f_a(1-2f_a)}{3(1-f_a)^2} \underline{\underline{\chi}}_{a0} \cdot \underline{\underline{\mathbf{L}}} \cdot \underline{\underline{\chi}}_{a0}, \quad (71)$$

$$\begin{aligned} \underline{\underline{\Sigma}}_1^{[3]} = \underline{\underline{\Sigma}}_1^{[2]} + \frac{f_a(1-2f_a)}{3(1-f_a)^2} \{ & \underline{\underline{\chi}}_{a0} \cdot (\underline{\underline{\mathbf{V}}}_0 \cdot \underline{\underline{\chi}}_{a1} \cdot \underline{\underline{\mathbf{W}}}_0 + \underline{\underline{\mathbf{V}}}_0 \cdot \underline{\underline{\chi}}_{a0} \cdot \underline{\underline{\mathbf{W}}}_1 \\ & + \underline{\underline{\mathbf{V}}}_1 \cdot \underline{\underline{\chi}}_{a0} \cdot \underline{\underline{\mathbf{W}}}_0 + \underline{\underline{\mathbf{W}}}_0 \cdot \underline{\underline{\chi}}_{a1} \cdot \underline{\underline{\mathbf{V}}}_0 + \underline{\underline{\mathbf{W}}}_0 \cdot \underline{\underline{\chi}}_{a0} \cdot \underline{\underline{\mathbf{V}}}_1 \\ & + \underline{\underline{\mathbf{W}}}_1 \cdot \underline{\underline{\chi}}_{a0} \cdot \underline{\underline{\mathbf{V}}}_0 + \underline{\underline{\mathbf{W}}}_0 \cdot \underline{\underline{\chi}}_{a1} \cdot \underline{\underline{\mathbf{W}}}_0 + \underline{\underline{\mathbf{W}}}_0 \cdot \underline{\underline{\chi}}_{a0} \cdot \underline{\underline{\mathbf{W}}}_1 \\ & + \underline{\underline{\mathbf{W}}}_1 \cdot \underline{\underline{\chi}}_{a0} \cdot \underline{\underline{\mathbf{W}}}_0) \cdot \underline{\underline{\chi}}_{a0} + \underline{\underline{\chi}}_{a0} \cdot \underline{\underline{\mathbf{L}}} \cdot \underline{\underline{\chi}}_{a1} + \underline{\underline{\chi}}_{a1} \cdot \underline{\underline{\mathbf{L}}} \cdot \underline{\underline{\chi}}_{a0} \}, \end{aligned} \quad (72)$$

wherein

$$\underline{\underline{\mathbf{L}}} = \underline{\underline{\mathbf{V}}}_0 \cdot \underline{\underline{\chi}}_{a0} \cdot \underline{\underline{\mathbf{W}}}_0 + \underline{\underline{\mathbf{W}}}_0 \cdot \underline{\underline{\chi}}_{a0} \cdot \underline{\underline{\mathbf{V}}}_0 + \underline{\underline{\mathbf{W}}}_0 \cdot \underline{\underline{\chi}}_{a0} \cdot \underline{\underline{\mathbf{W}}}_0, \quad (73)$$

and

$$\underline{\underline{\mathbf{W}}}_0 = i f_a f_b \omega L^2 \underline{\underline{\mathbf{P}}}_0^{F_0} / 3, \quad \underline{\underline{\mathbf{W}}}_1 = i f_a f_b \omega L^2 \underline{\underline{\mathbf{P}}}_1^{F_0} / 3, \quad (74)$$

$$\underline{\underline{\mathbf{V}}}_0 = \frac{1}{i\omega} \underline{\underline{\mathbf{K}}}_{cm0}^{-1} - \underline{\underline{\mathbf{D}}}_0^{I/cm}, \quad \underline{\underline{\mathbf{V}}}_1 = \frac{1}{i\omega} \underline{\underline{\mathbf{S}}} \left\{ \underline{\underline{\mathbf{K}}}_{cm} \right\} - \underline{\underline{\mathbf{D}}}_1^{I/cm}. \quad (75)$$

Thus, under the assumption of weak nonlinearity, the linear and nonlinear components of $\underline{\underline{\mathbf{K}}}_{Dy}^{[3]}$ emerge as

$$\underline{\underline{\mathbf{K}}}_{Dy0}^{[3]} = \underline{\underline{\mathbf{K}}}_{Br0} - \frac{1}{i\omega} \underline{\underline{\mathbf{M}}}_0^{-1} \cdot \underline{\underline{\tilde{\mathbf{S}}}}_0^{[3]}, \quad (76)$$

$$\underline{\underline{\mathbf{K}}}_{Dy1}^{[3]} = \underline{\underline{\mathbf{K}}}_{Br1} - \frac{1}{i\omega} \left(\underline{\underline{\mathbf{M}}}_0^{-1} \cdot \underline{\underline{\tilde{\mathbf{S}}}}_1^{[3]} + \underline{\underline{\mathbf{S}}} \left\{ \underline{\underline{\mathbf{M}}} \right\} \cdot \underline{\underline{\tilde{\mathbf{S}}}}_0^{[3]} \right), \quad (77)$$

respectively, where

$$\underline{\underline{\mathbf{M}}} = \underline{\underline{\mathbf{M}}}_0 + \underline{\underline{\mathbf{M}}}_1 |E_{HCM}|^2, \quad (78)$$

with

$$\underline{\underline{\mathbf{M}}}_0 = \underline{\underline{\mathbf{I}}} + \underline{\underline{\tilde{\mathbf{S}}}}_0^{[3]} \cdot \underline{\underline{\mathbf{D}}}_0^{I/cm}, \quad \underline{\underline{\mathbf{M}}}_1 = \underline{\underline{\tilde{\mathbf{S}}}}_0^{[3]} \cdot \underline{\underline{\mathbf{D}}}_1^{I/cm} + \underline{\underline{\tilde{\mathbf{S}}}}_1^{[3]} \cdot \underline{\underline{\mathbf{D}}}_0^{I/cm}. \quad (79)$$

5.5 Numerical results

Further light is shed on the weakly nonlinear homogenization scenario presented in Sections 5.2–5.4 by means of the following numerical example. Consider the HCM arising from component phases specified by the constitutive parameters: $\epsilon^{a0} = 2.4$, $\epsilon^{a1} = 9.07571 \times 10^{-12} \text{ m}^2 \text{ V}^{-2}$ ($\equiv 6.5 \times 10^{-4} \text{ esu}$); $\epsilon^b = \gamma \epsilon^{a0}$, $\xi^b = 1.2$ and $\mu^b = 1.4$. The parameter $\gamma \in [1, 6]$ allows the linear permittivity contrast between the two component phases to be varied. The angular frequency ω is fixed at $2\pi \times 10^{10} \text{ rad s}^{-1}$.

Consider the lowest-order SPFT, employing the iterative schemes (61) and (62), along with the initial values (60). The calculated HCM constitutive parameters $\epsilon_{Dy0}^{[0]}$ and $\epsilon_{Dy1}^{[0]} = \epsilon_{Dy1}^{[0]} / \epsilon^{a1}$ are plotted as functions of volume fraction f_a and γ in Figure 4. The graph of $\epsilon_{Dy0}^{[0]}$ is constrained by the values of ϵ^{a0} at $f_a = 1$ and the values of $\gamma \epsilon^{a0}$ at $f_a = 0$, and it follows an approximately linear progression for intermediate values of f_a . Of greater significance is the observation that the graph of $\epsilon_{Dy1}^{[0]}$ rises above unity for sufficiently large values of γ , in the neighborhood of $f_a \approx 0.6$. This phenomenon of *nonlinearity enhancement* has considerable potential for application in optical device technologies [36], and has been explored with the Bruggeman formalism [37–38]. An analogous nonlinearity enhancement occurs for electrostrictive dielectric HCMs [18–19]. In order to best consider the manifestation of nonlinearity enhancement, we fix $\gamma = 6$ for all remaining numerical calculations.

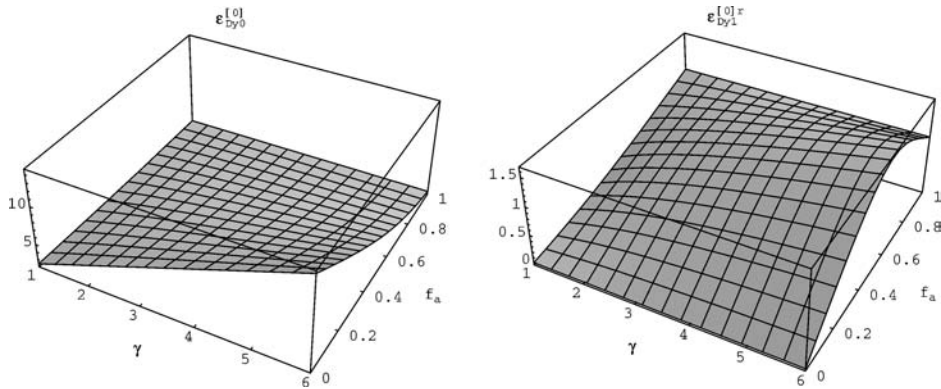


Figure 4 $\epsilon_{Dy0}^{[0]}$ and $\epsilon_{Dy1}^{[0]r} = \epsilon_{Dy1}^{[0]}/\epsilon^{a1}$ versus f_a and γ .

Attention is now turned to the bilocally approximated SPFT calculations. In Figure 5, the following quantities are plotted against f_a and k_0L :

$$\epsilon_{Dy0}^{[2-0]} = \epsilon_{Dy0}^{[2]} - \epsilon_{Dy0}^{[0]}, \quad \epsilon_{Dy1}^{[2-0]r} = \frac{1}{\epsilon^{a1}} \left(\epsilon_{Dy1}^{[2]} - \epsilon_{Dy1}^{[0]} \right). \quad (80)$$

Clearly, the second-order SPFT provides a significant correction to the lowest-order SPFT. Notice in particular the nonzero imaginary part of $\epsilon_{Dy0}^{[2-0]}$, which originates entirely from $\epsilon_{Dy0}^{[2]}$. As neither component phase a nor component phase b is dissipative, the imaginary part of $\epsilon_{Dy0}^{[2]}$ represents scattering losses. As regards specifically nonlinear aspects, we see from the graphs of $\epsilon_{Dy1}^{[2-0]}$ that the second-order SPFT is rather more conservative in predicting nonlinearity enhancement than is the zeroth-order SPFT. This is consistent with results from the SPFT for electrostrictive dielectric HCMs [18-19].

In order to address convergence issues, the constitutive parameters plotted in Figure 6 are the differences

$$\epsilon_{Dy0}^{[3-2]} = \epsilon_{Dy0}^{[3]} - \epsilon_{Dy0}^{[2]}, \quad \epsilon_{Dy1}^{[3-2]r} = \frac{1}{\epsilon^{a1}} \left(\epsilon_{Dy1}^{[3]} - \epsilon_{Dy1}^{[2]} \right). \quad (81)$$

From the displayed imaginary parts of $\epsilon_{Dy0}^{[3-2]}$, we infer that higher levels of coherent attenuation are associated with the third-order SPFT than are associated with the second-order SPFT. This reflects that fact that the third-order theory incorporates interactions between triplets of scattering centers, whereas the second-order theory makes allowance only for pairs of scatterers. Notice in particular the magnitudes of the parameters in Figure 6 as compared with the magnitudes plotted in Figures 4 and 5. Generally, for both linear and nonlinear parameters, an order-of-magnitude difference separates the third-order and the second-order values on the one hand, and the second-order and the zeroth-order values on the other. Thus, ob-

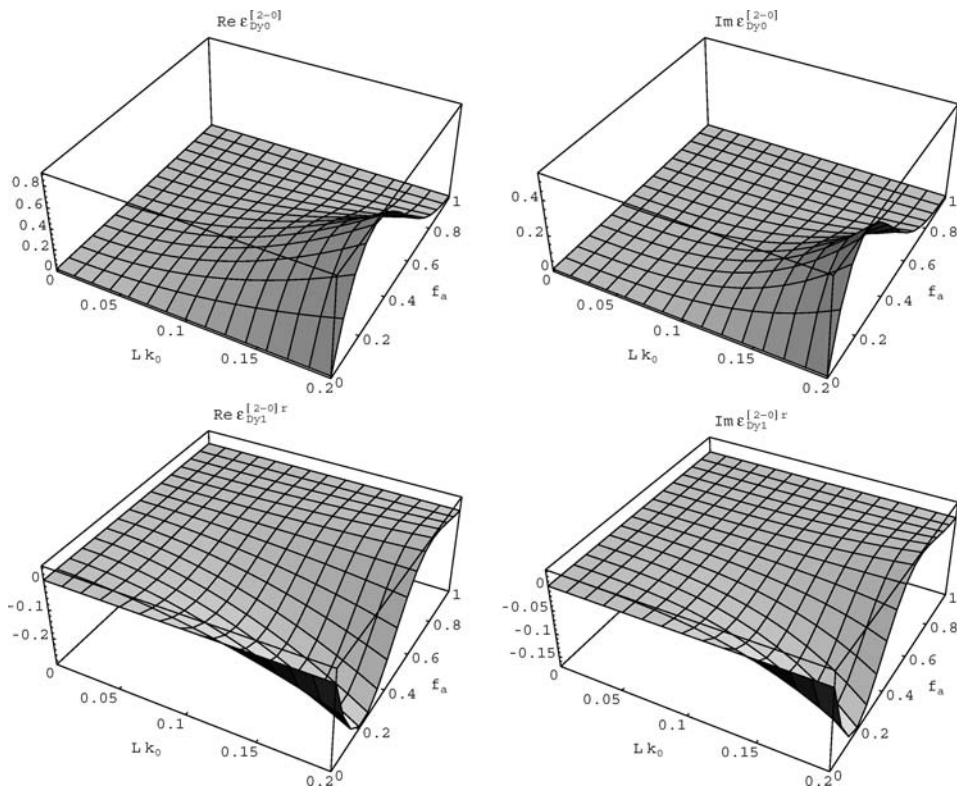


Figure 5 Real and imaginary parts of $\epsilon_{Dy0}^{[2-0]} = \epsilon_{Dy0}^{[2]} - \epsilon_{Dy0}^{[0]}$ and $\epsilon_{Dy1}^{[2-0]r} = (\epsilon_{Dy1}^{[2]} - \epsilon_{Dy1}^{[0]})/\epsilon^{a1}$ versus $k_0 L$ and f_a .

serve that SPFT convergence is achieved at the level of the bilocal approximation, for both linear and weakly nonlinear contributions [19].

6 Concluding remarks

Recent years have witnessed rapid developments in both the theory and practice of the materials sciences, particularly in the fields pertaining to complex composite materials [2]. In terms of homogenization, the well established Maxwell Garnett and Bruggeman formalisms have been extended to accommodate the most general class of linear HCMs, namely bianisotropic HCMs. Additionally, through the establishment of the incremental and differential Maxwell Garnett formalisms, the dilute HCM limitation—which restricts the applicability of the conventional Maxwell Garnett approach—has been overcome. The development of the SPFT for linear bianisotropic HCMs has enabled higher-order spatial correlations of the component phase distributions to be taken into account in estimations of HCM con-

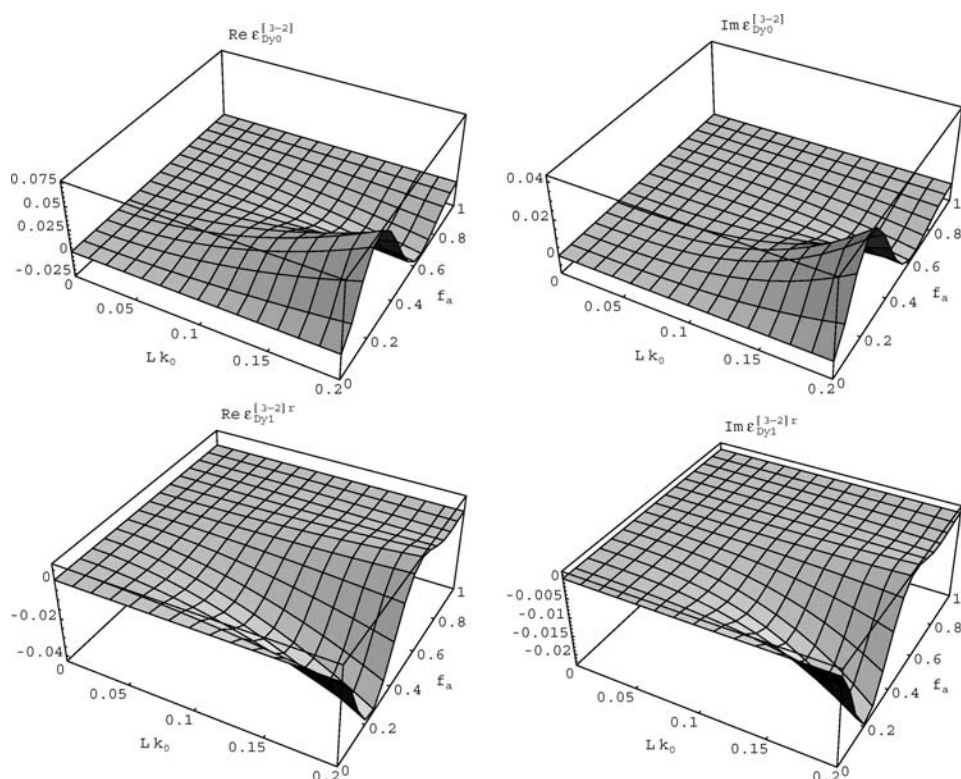


Figure 6 Real and imaginary parts of $\epsilon_{Dy0}^{[3-2]} = \epsilon_{Dy0}^{[3]} - \epsilon_{Dy0}^{[2]}$ and $\epsilon_{Dy1}^{[3-2]r} = (\epsilon_{Dy1}^{[3]} - \epsilon_{Dy1}^{[2]})/\epsilon^{a1}$ versus $k_0 L$ and f_a .

stitutive parameters. Since the convergence of the SPFT results at the second-order level, the relatively straightforward bilocally approximated SPFT may be applied in both linear and weakly nonlinear homogenization scenarios.

While perturbative nonlinearities may be accommodated in the SPFT, the challenge for the future lies in providing accurate theoretical descriptions of more generally nonlinear HCMs and diffusive HCMs. At present, it appears that the SPFT approach is appropriate only to systems which either (a) are wholly linear, or (b) exhibit sufficiently weak nonlinearities that the systems may be linearized via Maclaurin series expansions.

Appendix 1

Evaluations of the principal value integral term of $\underline{\underline{\mathbf{P}}}^r$ (26) are presented for the two-point covariance functions (27)–(29). The comparison medium is taken to be a reciprocal biaxial bianisotropic comparison medium; i.e., the 3×3 constitutive

dyadics $\underline{\underline{\epsilon}}_{cm}$, $\underline{\underline{\xi}}_{cm}$, $\underline{\underline{\zeta}}_{cm}$ and $\underline{\underline{\mu}}_{cm}$ are all diagonal and $\underline{\underline{\xi}}_{cm} = -\underline{\underline{\zeta}}_{cm}$. We have [16]

$$\begin{aligned} \underline{\underline{\mathbf{P}}}^{\Gamma_0} = & \frac{\omega^3 f_a f_b}{4\pi i} \int_{\phi=0}^{2\pi} d\phi \int_{\theta=0}^{\pi} d\theta \frac{1}{t_4(\underline{\underline{U}}^{-1} \cdot \underline{\underline{\hat{q}}})} \left\{ \frac{1}{\kappa_+ - \kappa_-} \right. \\ & \times \left[e^{i\sigma q} (1 - i\sigma q) \left(\frac{\underline{\underline{\mathbf{A}}}(\underline{\underline{U}}^{-1} \cdot \underline{\underline{\hat{q}}})}{\omega^2} + \frac{\underline{\underline{\mathbf{Y}}}(\underline{\underline{U}}^{-1} \cdot \underline{\underline{\hat{q}}})}{q^2} \right) \right]_{q=\sqrt{\kappa_-}}^{q=\sqrt{\kappa_+}} \\ & \left. + \frac{\underline{\underline{\mathbf{Y}}}(\underline{\underline{U}}^{-1} \cdot \underline{\underline{\hat{q}}})}{\kappa_+ \kappa_-} \right\}, \end{aligned} \quad (82)$$

$$\begin{aligned} \underline{\underline{\mathbf{P}}}^{\Gamma_1} = & \frac{\omega f_a f_b}{4\pi i} \int_{\phi=0}^{2\pi} d\phi \int_{\theta=0}^{\pi} d\theta \left\{ \frac{\sigma^2 \underline{\underline{\mathbf{A}}}(\underline{\underline{U}}^{-1} \cdot \underline{\underline{\hat{q}}}) + \omega^2 \underline{\underline{\mathbf{Y}}}(\underline{\underline{U}}^{-1} \cdot \underline{\underline{\hat{q}}})}{t_4(\underline{\underline{U}}^{-1} \cdot \underline{\underline{\hat{q}}}) (\sqrt{\kappa_+} + i\sigma)^2 (\sqrt{\kappa_-} + i\sigma)^2} \right. \\ & \left. - \frac{2i\sigma \left[\sqrt{\kappa_+} \sqrt{\kappa_-} \underline{\underline{\mathbf{A}}}(\underline{\underline{U}}^{-1} \cdot \underline{\underline{\hat{q}}}) - \omega^2 \underline{\underline{\mathbf{Y}}}(\underline{\underline{U}}^{-1} \cdot \underline{\underline{\hat{q}}}) \right] \left(\frac{1}{\sqrt{\kappa_+} + \sqrt{\kappa_-}} \right)}{t_4(\underline{\underline{U}}^{-1} \cdot \underline{\underline{\hat{q}}}) (\sqrt{\kappa_+} + i\sigma)^2 (\sqrt{\kappa_-} + i\sigma)^2} \right\}, \end{aligned} \quad (83)$$

$$\begin{aligned} \underline{\underline{\mathbf{P}}}^{\Gamma_2} = & \frac{f_a f_b \omega}{8i\pi\sigma} \int_{\phi=0}^{2\pi} d\phi \int_{\theta=0}^{\pi} d\theta \frac{1}{(\kappa_+ - \kappa_-) t_4(\underline{\underline{U}}^{-1} \cdot \underline{\underline{\hat{q}}})} \\ & \times \left\{ \left[m^2 \underline{\underline{\mathbf{A}}}(\underline{\underline{U}}^{-1} \cdot \underline{\underline{\hat{q}}}) + \omega^2 \underline{\underline{\mathbf{Y}}}(\underline{\underline{U}}^{-1} \cdot \underline{\underline{\hat{q}}}) \right] e^{-\frac{m^2}{8\sigma}} D_{-2} \left(\frac{-im}{\sqrt{2}\sigma} \right) \right\}_{m=\sqrt{\kappa_-}}^{m=\sqrt{\kappa_+}}, \end{aligned} \quad (84)$$

where

$$D_{-2}(z) = \exp\left(-\frac{z^2}{4}\right) - \left[\sqrt{\frac{\pi}{2}} z \exp\left(\frac{z^2}{4}\right) \operatorname{erfc}\left(\frac{z}{\sqrt{2}}\right) \right], \quad (85)$$

is the second-order parabolic cylinder function and erfc denotes the complementary error function [39]. In (82)–(84) the 6×6 dyadic quantities $\underline{\underline{\mathbf{A}}}$ and $\underline{\underline{\mathbf{Y}}}$ have been introduced; their components are given as

$$\left[\underline{\underline{\mathbf{A}}}(\underline{\underline{\hat{v}}}) \right]_{\alpha\beta} = \begin{cases} \left[\underline{\underline{\mathbf{T}}}_2(\underline{\underline{\hat{v}}}) \right]_{\alpha\beta} - \frac{t_2(\underline{\underline{\hat{v}}})}{t_4(\underline{\underline{\hat{v}}})} \left[\underline{\underline{\mathbf{T}}}_4(\underline{\underline{\hat{v}}}) \right]_{\alpha\beta}, & \alpha \equiv \beta \pmod{3}, \\ 0, & \alpha \not\equiv \beta \pmod{3}, \end{cases} \quad (86)$$

and

$$\left[\underline{\underline{\mathbf{Y}}}(\underline{\underline{\hat{v}}}) \right]_{\alpha\beta} = \begin{cases} \left[\underline{\underline{\mathbf{T}}}_0(\underline{\underline{\hat{v}}}) \right]_{\alpha\beta} - \frac{t_0}{t_4(\underline{\underline{\hat{v}}})} \left[\underline{\underline{\mathbf{T}}}_4(\underline{\underline{\hat{v}}}) \right]_{\alpha\beta}, & \alpha \equiv \beta \pmod{3}, \\ 0, & \alpha \not\equiv \beta \pmod{3}, \end{cases} \quad (87)$$

for $\alpha, \beta = 1, 2, \dots, 6$. The scalar quantities $t_4(\hat{\underline{\nu}})$, $t_2(\hat{\underline{\nu}})$, t_0 and $\kappa_{\pm}$ are defined as

$$t_4(\hat{\underline{\nu}}) = (\hat{\underline{\nu}} \cdot \underline{\underline{\epsilon}}_{cm} \cdot \hat{\underline{\nu}}) (\hat{\underline{\nu}} \cdot \underline{\underline{\mu}}_{cm} \cdot \hat{\underline{\nu}}) + (\hat{\underline{\nu}} \cdot \underline{\underline{\xi}}_{cm} \cdot \hat{\underline{\nu}}) (\hat{\underline{\nu}} \cdot \underline{\underline{\xi}}_{cm} \cdot \hat{\underline{\nu}}), \quad (88)$$

$$\begin{aligned} t_2(\hat{\underline{\nu}}) = & \text{tr} \left\{ \left[\hat{\underline{\nu}} \times \text{adj}(\underline{\underline{\epsilon}}_{cm}) \right] \cdot \left[\hat{\underline{\nu}} \times \text{adj}(\underline{\underline{\mu}}_{cm}) \right] \right. \\ & \left. - \left[\hat{\underline{\nu}} \times \text{adj}(\underline{\underline{\xi}}_{cm}) \right] \cdot \left[\hat{\underline{\nu}} \times \text{adj}(\underline{\underline{\xi}}_{cm}) \right] \right\} \\ & + \hat{\underline{\nu}} \cdot \left[\underline{\underline{\epsilon}}_{cm} \cdot \underline{\underline{F}}(\underline{\underline{\xi}}_{cm}, \underline{\underline{\xi}}_{cm}) \cdot \underline{\underline{\mu}}_{cm} - \underline{\underline{\xi}}_{cm} \cdot \underline{\underline{F}}(\underline{\underline{\epsilon}}_{cm}, \underline{\underline{\mu}}_{cm}) \cdot \underline{\underline{\xi}}_{cm} \right] \cdot \hat{\underline{\nu}}, \end{aligned} \quad (89)$$

$$t_0 = \det(\underline{\underline{\Pi}}), \quad (90)$$

$$\kappa_{\pm} = \omega^2 \left[\frac{-t_2(\hat{\underline{\nu}}) \pm \sqrt{t_2^2(\hat{\underline{\nu}}) - 4t_4(\hat{\underline{\nu}})t_0}}{2t_4(\hat{\underline{\nu}})} \right], \quad (91)$$

wherein

$$\underline{\underline{F}}(\underline{\underline{m}}, \underline{\underline{n}}) = \left[(\text{tr} \underline{\underline{m}}) \underline{\underline{I}} - \underline{\underline{m}} \right] \cdot \left[(\text{tr} \underline{\underline{n}}) \underline{\underline{I}} - \underline{\underline{n}} \right] - \left[(\text{tr} \underline{\underline{m}} \cdot \underline{\underline{n}}) \underline{\underline{I}} - \underline{\underline{m}} \cdot \underline{\underline{n}} \right], \quad (92)$$

$$\underline{\underline{\Pi}} = \underline{\underline{\epsilon}}_{cm} \cdot \underline{\underline{\mu}}_{cm} + \underline{\underline{\xi}}_{cm} \cdot \underline{\underline{\xi}}_{cm}, \quad (93)$$

with $\underline{\underline{m}}$ and $\underline{\underline{n}}$ being arbitrary 3×3 dyadics. The dyadics $\underline{\underline{\mathbf{T}}}_0(\hat{\underline{\nu}})$ and $\underline{\underline{\mathbf{T}}}_4(\hat{\underline{\nu}})$ are given in coordinate-free form as

$$\underline{\underline{\mathbf{T}}}_0(\hat{\underline{\nu}}) = \begin{bmatrix} \underline{\underline{\mu}}_{cm} \cdot \text{adj}(\underline{\underline{\Pi}}) & -\underline{\underline{\xi}}_{cm} \cdot \text{adj}(\underline{\underline{\Pi}}) \\ \underline{\underline{\xi}}_{cm} \cdot \text{adj}(\underline{\underline{\Pi}}) & \underline{\underline{\epsilon}}_{cm} \cdot \text{adj}(\underline{\underline{\Pi}}) \end{bmatrix}, \quad (94)$$

and

$$\underline{\underline{\mathbf{T}}}_4(\hat{\underline{\nu}}) = \begin{bmatrix} (\hat{\underline{\nu}} \cdot \underline{\underline{\mu}}_{cm} \cdot \hat{\underline{\nu}}) \hat{\underline{\nu}} \hat{\underline{\nu}} & -(\hat{\underline{\nu}} \cdot \underline{\underline{\xi}}_{cm} \cdot \hat{\underline{\nu}}) \hat{\underline{\nu}} \hat{\underline{\nu}} \\ (\hat{\underline{\nu}} \cdot \underline{\underline{\xi}}_{cm} \cdot \hat{\underline{\nu}}) \hat{\underline{\nu}} \hat{\underline{\nu}} & (\hat{\underline{\nu}} \cdot \underline{\underline{\epsilon}}_{cm} \cdot \hat{\underline{\nu}}) \hat{\underline{\nu}} \hat{\underline{\nu}} \end{bmatrix}. \quad (95)$$

Only the diagonal components of $\underline{\underline{\mathbf{T}}}_2(\hat{\underline{\nu}})$ are needed to evaluate (26):

$$\begin{aligned} [T_2^{ee}(\hat{\underline{\nu}})]_{\alpha\alpha} = & \left(\left\{ 2\underline{\underline{\epsilon}}_{cm} \cdot \text{adj}(\underline{\underline{\mu}}_{cm}) - \text{tr}[\underline{\underline{\epsilon}}_{cm} \cdot \text{adj}(\underline{\underline{\mu}}_{cm})] \underline{\underline{I}} \right. \right. \\ & \left. \left. + \underline{\underline{\mu}}_{cm} \cdot \underline{\underline{F}}(\underline{\underline{\xi}}_{cm}, \underline{\underline{\xi}}_{cm}) \right\} \cdot \hat{\underline{\nu}} \hat{\underline{\nu}} - \underline{\underline{F}}(\underline{\underline{\mu}}_{cm}, \underline{\underline{\xi}}_{cm} \cdot \underline{\underline{\xi}}_{cm} \cdot \hat{\underline{\nu}} \hat{\underline{\nu}}) \right. \\ & \left. - (\hat{\underline{\nu}} \cdot \underline{\underline{\epsilon}}_{cm} \cdot \hat{\underline{\nu}}) \text{adj}(\underline{\underline{\mu}}_{cm}) \right)_{\alpha\alpha}, \quad \alpha = 1, 2, 3, \end{aligned} \quad (96)$$

$$\begin{aligned} \left[\underline{\underline{T}}_2^{em}(\hat{\underline{\underline{v}}}) \right]_{\alpha\alpha} = & \left\{ \left[\underline{\underline{F}}(\underline{\underline{\epsilon}}_{cm}, \underline{\underline{\mu}}_{cm}) \cdot \underline{\underline{\xi}}_{cm} - \det(\underline{\underline{\xi}}_{cm}) \underline{\underline{I}} \right] \cdot \hat{\underline{\underline{v}}} \hat{\underline{\underline{v}}} \right. \\ & - \underline{\underline{F}}(\underline{\underline{\epsilon}}_{cm} \cdot \underline{\underline{\mu}}_{cm} \cdot \hat{\underline{\underline{v}}} \hat{\underline{\underline{v}}}, \underline{\underline{\xi}}_{cm}) \\ & \left. - \left(\hat{\underline{\underline{v}}} \cdot \underline{\underline{\xi}}_{cm} \cdot \hat{\underline{\underline{v}}} \right) \text{adj}(\underline{\underline{\xi}}_{cm}) \right\}_{\alpha\alpha}, \quad \alpha = 1, 2, 3, \end{aligned} \quad (97)$$

$$\begin{aligned} \left[\underline{\underline{T}}_2^{me}(\hat{\underline{\underline{v}}}) \right]_{\alpha\alpha} = & - \left\{ \left[\underline{\underline{F}}(\underline{\underline{\epsilon}}_{cm}, \underline{\underline{\mu}}_{cm}) \cdot \underline{\underline{\xi}}_{cm} - \det(\underline{\underline{\xi}}_{cm}) \underline{\underline{I}} \right] \cdot \hat{\underline{\underline{v}}} \hat{\underline{\underline{v}}} \right. \\ & - \underline{\underline{F}}(\underline{\underline{\epsilon}}_{cm} \cdot \underline{\underline{\mu}}_{cm} \cdot \hat{\underline{\underline{v}}} \hat{\underline{\underline{v}}}, \underline{\underline{\xi}}_{cm}) \\ & \left. - \left(\hat{\underline{\underline{v}}} \cdot \underline{\underline{\xi}}_{cm} \cdot \hat{\underline{\underline{v}}} \right) \text{adj}(\underline{\underline{\xi}}_{cm}) \right\}_{\alpha\alpha}, \quad \alpha = 1, 2, 3, \end{aligned} \quad (98)$$

$$\begin{aligned} \left[\underline{\underline{T}}_2^{mm}(\hat{\underline{\underline{v}}}) \right]_{\alpha\alpha} = & \left(\left\{ 2 \text{adj}(\underline{\underline{\epsilon}}_{cm}) \cdot \underline{\underline{\mu}}_{cm} - \text{tr} \left[\text{adj}(\underline{\underline{\epsilon}}_{cm}) \cdot \underline{\underline{\mu}}_{cm} \right] \underline{\underline{I}} \right. \right. \\ & \left. \left. + \underline{\underline{\epsilon}}_{cm} \cdot \underline{\underline{F}}(\underline{\underline{\xi}}_{cm}, \underline{\underline{\xi}}_{cm}) \right\} \cdot \hat{\underline{\underline{v}}} \hat{\underline{\underline{v}}} - \underline{\underline{F}}(\underline{\underline{\epsilon}}_{cm}, \underline{\underline{\xi}}_{cm} \cdot \underline{\underline{\xi}}_{cm} \cdot \hat{\underline{\underline{v}}} \hat{\underline{\underline{v}}}) \right. \\ & \left. - \left(\hat{\underline{\underline{v}}} \cdot \underline{\underline{\mu}}_{cm} \cdot \hat{\underline{\underline{v}}} \right) \text{adj}(\underline{\underline{\epsilon}}_{cm}) \right)_{\alpha\alpha}, \quad \alpha = 1, 2, 3. \end{aligned} \quad (99)$$

Values of σ are specified by setting

$$\int_{\underline{\underline{R}}} \Gamma_j(\underline{\underline{R}}) d^3 \underline{\underline{R}} = \frac{4\pi L^3}{3}, \quad (j = 0, 1, 2). \quad (100)$$

Hence, $\sigma = L$ for Γ_0 ; $\sigma = (\sqrt[3]{6})/L$ for Γ_1 ; and $\sigma = (\sqrt[3]{9\pi/16})/L^2$ for Γ_2 .

Appendix 2

Evaluations of the principal value integral terms $\underline{\underline{\mathbf{P}}}_0^{\Gamma_0}$ and $\underline{\underline{\mathbf{P}}}_1^{\Gamma_0}$ in (63) are presented. The comparison medium is taken to be a weakly nonlinear isotropic chiral medium with constitutive dyadic $\underline{\underline{\mathbf{K}}}_{cm}$ specified by (39).

The linear contribution to $\underline{\underline{\mathbf{P}}}^{\Gamma_0}$ is found to be [20]

$$\underline{\underline{\mathbf{P}}}_0 = \begin{bmatrix} \mu_0 \mu_{cm0} \left(1 + i \frac{2}{3} L \omega \Theta \right) \underline{\underline{I}} & i \sqrt{\epsilon_0 \mu_0} \xi_{cm0} \left(1 + i \frac{4}{3} L \omega \rho \right) \underline{\underline{I}} \\ -i \sqrt{\epsilon_0 \mu_0} \xi_{cm0} \left(1 + i \frac{4}{3} L \omega \rho \right) \underline{\underline{I}} & \epsilon_0 \epsilon_{cm0} \left(1 + i \frac{2}{3} L \omega \Theta \right) \underline{\underline{I}} \end{bmatrix}, \quad (101)$$

and the nonlinear contribution

$$\begin{aligned} \underline{\underline{\mathbf{P}}}_1 = & \begin{bmatrix} \mu_0 \mu_{cm1} \left(1 + i \frac{2}{3} L \omega \Theta\right) \underline{\underline{I}} & i \sqrt{\epsilon_0 \mu_0} \xi_{cm1} \left(1 + i \frac{4}{3} L \omega \rho\right) \underline{\underline{I}} \\ -i \sqrt{\epsilon_0 \mu_0} \xi_{cm1} \left(1 + i \frac{4}{3} L \omega \rho\right) \underline{\underline{I}} & \epsilon_0 \epsilon_{cm1} \left(1 + i \frac{2}{3} L \omega \Theta\right) \underline{\underline{I}} \end{bmatrix} \\ & + \begin{bmatrix} \mu_0 \mu_{cm0} i \frac{2}{3} L \omega \Pi \underline{\underline{I}} & -\sqrt{\epsilon_0 \mu_0} \xi_{cm0} \frac{4}{3} L \omega \lambda \underline{\underline{I}} \\ \sqrt{\epsilon_0 \mu_0} \xi_{cm0} \frac{4}{3} L \omega \lambda \underline{\underline{I}} & \epsilon_0 \epsilon_{cm0} i \frac{2}{3} L \omega \Pi \underline{\underline{I}} \end{bmatrix}, \end{aligned} \quad (102)$$

where

$$\Theta = \frac{\sqrt{\epsilon_0 \mu_0} (\epsilon_{cm0} \mu_{cm0} + \xi_{cm0}^2)}{\sqrt{\epsilon_{cm0} \mu_{cm0}}}, \quad (103)$$

$$\rho = \sqrt{\epsilon_0 \mu_0} \sqrt{\epsilon_{cm0} \mu_{cm0}}, \quad (104)$$

$$\Pi = \left(\frac{\epsilon_{cm0} \mu_{cm0} - \xi_{cm0}^2}{\epsilon_{cm0} \mu_{cm0}} \right) \lambda + \tau, \quad (105)$$

$$\lambda = \sqrt{\epsilon_0 \mu_0} \frac{\epsilon_{cm0} \mu_{cm1} + \epsilon_{cm1} \mu_{cm0}}{2 \sqrt{\epsilon_{cm0} \mu_{cm0}}}, \quad (106)$$

$$\tau = \frac{2 \sqrt{\epsilon_0 \mu_0} \xi_{cm0} \xi_{cm1}}{\sqrt{\epsilon_{cm0} \mu_{cm0}}}. \quad (107)$$

References

- [1] A. Lakhtakia (ed), *Selected papers on linear optical composite materials*, SPIE Press, Bellingham, WA, USA, 1996.
- [2] O.N. Singh and A. Lakhtakia (eds), *Electromagnetic fields in unconventional materials and structures*, Wiley, New York, NY, USA, 2000.
- [3] G.B. Smith, Nanostructured thin films, *Introduction to complex mediums for optics and electromagnetics* (W.S. Weiglhofer and A. Lakhtakia, eds), SPIE Press, Bellingham, WA, USA, 2003.
- [4] A. Lakhtakia, On direct and indirect scattering approaches for the homogenization of particulate composites, *Microw Opt Technol Lett* **25** (2000), 53–56.
- [5] J.-B. Biot and F. Arago, Mémoire sur les affinités des corps pour la lumière et particulièrement sur les forces réfringentes des différents gaz, *Mém Inst Fr* **7** (1806), 301–385.

- [6] L. Ward, *The optical constants of bulk materials and films*, Adam Hilger, Bristol, UK, 1988.
- [7] W.S. Weiglhofer, A. Lakhtakia and B. Michel, Maxwell Garnett and Bruggeman formalisms for a particulate composite with bianisotropic host medium, *Microw Opt Technol Lett* **15** (1997), 263–266; erratum: **22** (1999), 221.
- [8] B. Michel, Recent developments in the homogenization of linear bianisotropic composite materials, *Electromagnetic fields in unconventional materials and structures* (O.N. Singh and A. Lakhtakia, eds), Wiley, New York, NY, USA, 2000.
- [9] B. Michel, A Fourier space approach to the pointwise singularity of an anisotropic dielectric medium, *Int J Appl Electromagn Mech* **8** (1997), 219–227.
- [10] B. Michel and W.S. Weiglhofer, Pointwise singularity of dyadic Green function in a general bianisotropic medium, *Arch Elektron Übertrag* **51** (1997), 219–223; erratum: **52** (1998), 310.
- [11] A. Lakhtakia, Incremental Maxwell Garnett formalism for homogenizing particulate composite media, *Microw Opt Technol Lett* **17** (1998), 276–279.
- [12] B. Michel, A. Lakhtakia, W.S. Weiglhofer and T.G. Mackay, Incremental and differential Maxwell Garnett formalisms for bi-anisotropic composites, *Compos Sci Technol* **61** (2001), 13–18.
- [13] L. Tsang, J.A. Kong and R.T. Shin, *Theory of microwave remote sensing*, Wiley, New York, NY, USA, 1985.
- [14] T.G. Mackay, A. Lakhtakia and W.S. Weiglhofer, Strong-property-fluctuation theory for homogenization of bianisotropic composites: formulation, *Phys Rev E* **62** (2000), 6052–6064; erratum: **63** (2001), 049901.
- [15] T.G. Mackay, A. Lakhtakia and W.S. Weiglhofer, Ellipsoidal topology, orientation diversity and correlation length in bianisotropic composite mediums, *Arch Elektron Übertrag* **55** (2001), 243–251.
- [16] T.G. Mackay, A. Lakhtakia and W.S. Weiglhofer, Homogenisation of similarly oriented, metallic, ellipsoidal inclusions using the bilocal-approximated strong-property-fluctuation theory, *Opt Commun* **197** (2001), 89–95.
- [17] T.G. Mackay, A. Lakhtakia and W.S. Weiglhofer, Third-order implementation and convergence of the strong-property-fluctuation theory in electromagnetic homogenisation, *Phys Rev E* **64** (2001), 066616.
- [18] A. Lakhtakia, Application of strong permittivity fluctuation theory for isotropic, cubically nonlinear, composite mediums, *Opt Commun* **192** (2001), 145–151.
- [19] T.G. Mackay, A. Lakhtakia and W.S. Weiglhofer, Homogenisation of isotropic, cubically nonlinear, composite mediums by the strong-permittivity-fluctuation theory: third-order considerations, *Opt Commun* **204** (2002), 219–228.
- [20] T.G. Mackay, A. Lakhtakia and W.S. Weiglhofer, Electromagnetic homogenization of cubically nonlinear, isotropic chiral composite mediums via the strong-property-fluctuation theory, *Electromagnetics* **23** (2003) (to appear).

- [21] W.S. Weiglhofer, Constitutive characterization of simple and complex mediums, *Introduction to complex mediums for optics and electromagnetics* (W.S. Weiglhofer and A. Lakhtakia, eds), SPIE Press, Bellingham, WA, USA, 2003.
- [22] W.S. Weiglhofer, Electromagnetic depolarization dyadics and elliptic integrals, *J Phys A: Math Gen* **31** (1998), 7191–7196.
- [23] W.S. Weiglhofer and T.G. Mackay, Needles and pillboxes in anisotropic mediums, *IEEE Trans Antennas Propagat* **50** (2001), 85–86.
- [24] P.G. Cottis, C.N. Vazouras and C. Spyrou, Green's function for an unbounded biaxial medium in cylindrical coordinates, *IEEE Trans Antennas Propagat* **47** (1999), 195–199.
- [25] J.A. Sherwin and A. Lakhtakia, Bragg–Pippard formalism for bianisotropic particulate composites, *Microw Opt Technol Lett* **33** (2002), 40–44.
- [26] L. Tsang and J.A. Kong, Scattering of electromagnetic waves from random media with strong permittivity fluctuations, *Radio Sci* **16** (1981), 303–320.
- [27] U. Frisch, Wave propagation in random media, *Probabilistic methods in applied mathematics, Vol. 1* (A.T. Bharucha-Reid, ed), Academic Press, London, UK, 1970.
- [28] B. Michel and A. Lakhtakia, Strong-property-fluctuation theory for homogenizing chiral particulate composites, *Phys Rev E* **51** (1995), 5701–5707.
- [29] L. Tsang, J.A. Kong and R.W. Newton, Application of strong fluctuation random medium theory to scattering of electromagnetic waves from a half-space of dielectric mixture, *IEEE Trans Antennas Propagat* **30** (1982), 292–302.
- [30] M.N. Miller, Bounds for effective electrical, thermal, and magnetic properties of heterogeneous materials, *J Math Phys* **10** (1969), 1988–2004.
- [31] G.W. Milton and N. Phan-Thien, New bounds on effective elastic moduli of two-component materials, *Proc Roy Soc Lond A* **380** (1982), 305–331.
- [32] C.F. Bohren, Isotropic chiral materials, *Introduction to complex mediums for optics and electromagnetics* (W.S. Weiglhofer and A. Lakhtakia, eds), SPIE Press, Bellingham, WA, USA, 2003.
- [33] A. Lakhtakia, *Beltrami fields in chiral media*, World Scientific, Singapore, 1994.
- [34] G.Ya. Slepian, S.A. Maksimenko, F.G. Bass and A. Lakhtakia, Nonlinear electromagnetics in chiral media: self-action of waves, *Phys Rev E* **52** (1995), 1049–1058.
- [35] R.W. Boyd, *Nonlinear optics*, Academic Press, San Diego, CA, USA, 1992.
- [36] H.B. Liao, R.F. Xiao, H. Wang, K.S. Wong and G.K.L. Wong, Large third-order optical nonlinearity in Au:TiO₂ composite films measured on a femtosecond time scale, *Appl Phys Lett* **72** (1998), 1817–1819.
- [37] R.W. Boyd, R.J. Gehr, G.L. Fischer and J.E. Sipe, Nonlinear optical properties of nanocomposite materials, *Pure Appl Opt* **5** (1996), 505–512.

- [38] M.N. Lakhtakia and A. Lakhtakia, Anisotropic composite materials with intensity-dependent permittivity tensor: The Bruggeman approach, *Electromagnetics* **21** (2001), 129–138.
- [39] I.S. Gradshteyn and I.M. Ryzhik, *Table of integrals, series, and products*, Academic Press, London, UK, 1980.



Tom G. Mackay graduated MSci in Mathematics from the University of Glasgow in 1998, after spending the previous 10 years working as a bioengineer at Glasgow Royal Infirmary. He spent the next three years engaged in postgraduate studies in the Department of Mathematics at the University of Glasgow. Upon completing his Ph.D. thesis in 2001, he moved to the University of Edinburgh, where he is currently employed as a lecturer in the School of Mathematics. His research interests are primarily related to the homogenization of complex electromagnetic systems. He also maintains a research interest in cardiovascular bioengineering. His principal teaching commitment is lecturing on dynamical systems.

Negative Phase-Velocity Mediums

Akhlesh Lakhtakia, Martin W. McCall, and
Werner S. Weiglhofer

Abstract

Isotropic dielectric-magnetic mediums wherein the phase velocity vector and the time-averaged Poynting vector of propagating plane waves are oppositely directed have recently become the focus of research on microwave composite materials. The purpose of this overview is to illuminate the background of rapidly occurring developments, to present a clear theoretical foundation of the involved effects, and to clarify some of the confusion—such as regarding terminology—that has arisen in the research literature. Furthermore, a general condition for the particular effect of negative phase-velocity to occur is given; and ramifications for perfect lenses, homogenization approaches, and complex mediums are presented.

1 Introduction

Although electromagnetics is a microscopic science, most research on electromagnetic fields is carried out at macroscopic-length scales. Therefore, the microscopic fields $\tilde{\mathbf{e}}(\mathbf{r}, t)$ and $\tilde{\mathbf{b}}(\mathbf{r}, t)$ are spatially averaged into the macroscopic fields $\tilde{\mathbf{E}}(\mathbf{r}, t)$ and $\tilde{\mathbf{B}}(\mathbf{r}, t)$, respectively [1]. Discrete charges are also averaged similarly into a macroscopic charge density $\tilde{\rho}(\mathbf{r}, t)$, whose time-derivative is equal to the negative of the divergence of a macroscopic current density $\tilde{\mathbf{J}}(\mathbf{r}, t)$; i.e., the continuity condition $\nabla \cdot \tilde{\mathbf{J}} + \partial \tilde{\rho} / \partial t = 0$ is fulfilled.

The current density $\tilde{\mathbf{J}}(\mathbf{r}, t)$ is broken up into three components:

- an externally impressed source current density $\tilde{\mathbf{J}}_e(\mathbf{r}, t)$,
- a polarization term $\frac{\partial}{\partial t} \tilde{\mathbf{P}}(\mathbf{r}, t)$, and
- a magnetization term $\nabla \times \tilde{\mathbf{M}}(\mathbf{r}, t)$.

The charge density, however, comprises just two components:

- an externally impressed source charge density $\tilde{\rho}_e(\mathbf{r}, t)$ that is in accordance with $\nabla \cdot \tilde{\mathbf{J}}_e(\mathbf{r}, t) + \frac{\partial}{\partial t} \tilde{\rho}_e(\mathbf{r}, t) = 0$, and
- a polarization term $-\nabla \cdot \tilde{\mathbf{P}}(\mathbf{r}, t)$.

The magnetization component of $\tilde{\mathbf{J}}(\mathbf{r}, t)$ does not contribute to $\tilde{\rho}(\mathbf{r}, t)$. The terms not attributed to externally impressed sources account for the presence of matter, which is really free space (or vacuum) sprinkled with charges, but is viewed macroscopically as a continuum. Induction fields¹

$$\tilde{\mathbf{D}}(\mathbf{r}, t) = \epsilon_0 \tilde{\mathbf{E}}(\mathbf{r}, t) + \tilde{\mathbf{P}}(\mathbf{r}, t) \quad (1)$$

and

$$\tilde{\mathbf{H}}(\mathbf{r}, t) = \mu_0^{-1} \tilde{\mathbf{B}}(\mathbf{r}, t) - \tilde{\mathbf{M}}(\mathbf{r}, t) \quad (2)$$

thus arise to incorporate continuum material properties, and the Maxwell postulates assume their familiar shape [2-3]:

$$\nabla \cdot \tilde{\mathbf{B}}(\mathbf{r}, t) = 0, \quad (3)$$

$$\nabla \times \tilde{\mathbf{E}}(\mathbf{r}, t) + \frac{\partial}{\partial t} \tilde{\mathbf{B}}(\mathbf{r}, t) = \mathbf{0}, \quad (4)$$

$$\nabla \cdot \tilde{\mathbf{D}}(\mathbf{r}, t) = \tilde{\rho}_e(\mathbf{r}, t), \quad (5)$$

$$\nabla \times \tilde{\mathbf{H}}(\mathbf{r}, t) - \frac{\partial}{\partial t} \tilde{\mathbf{D}}(\mathbf{r}, t) = \tilde{\mathbf{J}}_e(\mathbf{r}, t). \quad (6)$$

¹Magnetic terminology is confused: Whereas $\tilde{\mathbf{B}}$ is the *magnetic field* as appears, for example, in the Lorentz force law, and $\tilde{\mathbf{H}}$ is the *magnetic induction*, it is very common to call $\tilde{\mathbf{H}}$ the magnetic field and $\tilde{\mathbf{B}}$ the magnetic induction.

The macroscopic Maxwell postulates are unsolvable, unless constitutive equations are provided to relate the induction fields $\mathbf{\tilde{D}}(\mathbf{r}, t)$ and $\mathbf{\tilde{H}}(\mathbf{r}, t)$ to the primitive fields $\mathbf{\tilde{E}}(\mathbf{r}, t)$ and $\mathbf{\tilde{B}}(\mathbf{r}, t)$. One can prescribe, in principle, a single set of constitutive relations for the entire universe and for all time—based on the *premise* that the Maxwell postulates are universally and always applicable. Commonly, however, constitutive equations are set up for spatially local and spatiotemporally homogeneous matter, as discussed in an earlier chapter [4]. More commonly, linear constitutive properties are assumed, while time-dependent fields are Fourier-transformed into frequency-dependent phasors. Complexity remains even then, and exact analytical solutions of the Maxwell postulates for linear matter are scarce [2], [5]. One would have imagined that knowledge of electromagnetic phasors in isotropic, homogeneous, linear dielectric-magnetic materials is complete; but a flurry of publications that have appeared from the year 2000 is sufficient to dispel that notion. We refer to *negative phase-velocity mediums*, popularly designated as *left-handed materials* [6-7].

More than three decades ago, Veselago [8] predicted many unusual properties of isotropic mediums with simultaneously negative real relative permittivity and negative real relative permeability at a certain frequency—including inverse refraction, negative radiation pressure, and inverse Doppler effect. The phase-velocity of a plane wave in these mediums is directed opposite to that in mediums with positive real relative permittivity and positive real relative permeability (which are the stuff of textbooks). It appears that the term *left-handed materials*, which amounts to confusing the direction of the phase-velocity with the direction of wave propagation, has its origin in a mistranslation of the Russian original of Veselago's paper into the English language. According to Veselago himself [9], his own description simply related to the handedness of the triplet of vectors spanned by electric and magnetic field phasors and the wave propagation vector. Unaware of the implications, he accepted the term *left-handed materials* for the mediums suggested by him when the translators offered that expression to him as an appropriate equivalent English term.

Veselago's initial suggestion remained completely speculative until a significant breakthrough was announced in 2000. Developing some earlier ideas by Pendry [10] and Pendry *et al.* [11], Smith *et al.* [12] presented evidence for a weakly dissipative composite medium displaying negative values for the real parts of its *effective* permittivity and *effective* permeability. Their materials are lattices of conducting rings and wires arranged with the assistance of printed circuit boards. The microstructures have been variously analyzed [13-14] and even considered to lead to macroscopic bianisotropy [15]. Magnetic microstructural combinations can also be devised to exhibit similar response properties [16]-[18].

Experimental results published in 2001 by Shelby *et al.* [19] on the transmission of a ~ 10 -GHz beam through a wedge provided impetus to the electromagnetics community for a discussion on the concept of a negative index of refraction. Briefly, the beam was made to normally irradiate one planar face of a wedge, while intensity measurements were made on the beam transmitted through the other face

of the wedge. Plots of the transmitted intensity against the observation angle with respect to the normal to the exit face were analyzed. When a Teflon wedge was replaced by a wedge made of the ring-wire composite material, the transmission intensity pattern flipped about the normal to the exit face. The 2001 paper by Shelby *et al.* triggered much research activity, which prompted the inclusion of this chapter in this volume.

2 Phenomenology

2.1 Basic equations

Let us consider a plane wave propagating along the z axis in a linear, homogeneous, isotropic, dielectric-magnetic medium whose relative permittivity and relative permeability are denoted by the complex numbers

$$\epsilon_r = \epsilon'_r + i\epsilon''_r = |\epsilon_r| \exp(i\phi_\epsilon), \quad (7)$$

$$\mu_r = \mu'_r + i\mu''_r = |\mu_r| \exp(i\phi_\mu), \quad (8)$$

respectively. An $\exp(-i\omega t)$ time-dependence is assumed here, with ω as the angular frequency, so that ϵ_r and μ_r are functions of the angular frequency.

The electric and magnetic field phasors associated with the plane wave are as follows:

$$\mathbf{E}(z, \omega) = A \exp(ik_0 n z) \mathbf{u}_x, \quad (9)$$

$$\mathbf{H}(z, \omega) = \frac{n}{\mu_r \eta_0} A \exp(ik_0 n z) \mathbf{u}_y. \quad (10)$$

Here, A is a constant, $k_0 = +\omega\sqrt{\epsilon_0\mu_0}$ is the free-space wavenumber, $\eta_0 = +\sqrt{\mu_0/\epsilon_0}$ is the intrinsic impedance of free space, and the complex refractive index n is to be determined from the dispersion relation $n^2 = \epsilon_r \mu_r$. The Poynting vector is parallel to the z axis and its time-average is given as

$$S_z(n) = \frac{1}{2} \mathbf{u}_z \cdot \text{Re}[\mathbf{E}(z, \omega) \times \mathbf{H}^*(z, \omega)] \quad (11)$$

$$= \text{Re} \left[\frac{n}{\mu_r} \right] \frac{|A|^2}{2\eta_0} \exp(-2k_0 \text{Im}[n]z), \quad (12)$$

where $\text{Re}[\cdot]$ and $\text{Im}[\cdot]$ respectively denote the operations of taking the real and the imaginary part, while $*$ indicates complex conjugation. Dissipation results from the imaginary parts of ϵ_r and μ_r while causality dictates that $\mu''_r > 0$ and $\epsilon''_r > 0$, so that ϵ_r and μ_r lie in the upper half of the complex plane.

2.2 Negative phase-velocity

However, there are two resultant complex refractive indexes,

$$n_{\pm} = \pm \sqrt{\epsilon_r \mu_r}, \quad (13)$$

of which n_+ lies in the upper half of the complex plane and n_- in the lower half. Now $n_{\pm}$ may be written as

$$n_{\pm} = \pm \nu \exp(i\phi_{\nu}), \quad (14)$$

where

$$\nu = +\sqrt{|\epsilon_r||\mu_r|}, \quad \phi_{\nu} = \frac{\phi_{\epsilon} + \phi_{\mu}}{2}. \quad (15)$$

The condition $0 \leq \phi_{\nu} \leq \pi$ follows from $0 \leq \phi_{\epsilon, \mu} \leq \pi$, so that we always have

$$\operatorname{Re}\left[\frac{n_+}{\mu_r}\right] > 0 \quad \Rightarrow \quad S_z(n_+) > 0 \quad (16)$$

and also

$$\operatorname{Re}\left[\frac{n_-}{\mu_r}\right] < 0 \quad \Rightarrow \quad S_z(n_-) < 0. \quad (17)$$

Thus, the choice n_+ *always* relates to power flow in the $+z$ direction, whilst n_- *always* relates to power flow in the $-z$ direction. Since necessarily $\operatorname{Im}[n_+] > 0$ and $\operatorname{Im}[n_-] < 0$, power flow is always in the direction of exponential decrease of the phasors' amplitudes.

We can now identify when the phase-velocity $v_{ph} = (\omega/k_0)/\operatorname{Re}[n]$ is opposite to the direction of power flow. This occurs whenever $\operatorname{Re}[n_+] < 0$ (and consequently $\operatorname{Re}[n_-] > 0$, also). The following condition is straightforwardly derived for such propagation [20]:

$$(|\epsilon_r| - \epsilon'_r)(|\mu_r| - \mu'_r) > \epsilon''_r \mu''_r. \quad (18)$$

Ruppin has derived the equivalent condition [21]

$$\epsilon''_r \mu'_r + \mu''_r \epsilon'_r < 0. \quad (19)$$

Recalling that $\epsilon''_r > 0$ and $\mu''_r > 0$, we see that the satisfaction of the double-negativity condition

$$\{\epsilon'_r < 0, \mu'_r < 0\} \quad (20)$$

leads to the satisfaction of (19). But the parameter space of the permittivity and permeability that fulfils (20) is only a subset of the one fulfilling (18). Clearly, upon inspection of (19), the simultaneous satisfaction of both $\epsilon'_r < 0$ and $\mu'_r < 0$ is a sufficient, but not necessary, requirement for the phase-velocity to be negative. In other words, *both* ϵ'_r and μ'_r do not have to be negative for v_{ph} to be negative, contrary to received wisdom [12], [22-23].

2.3 Dispersion

Let us exemplify the foregoing in detail by an explicit invocation of the Lorentz model for both ϵ_r and μ_r ; thus,

$$\epsilon_r(\lambda_0) = 1 + \frac{p_e}{1 + (N_e^{-1} - i\lambda_e\lambda_0^{-1})^2}, \quad (21)$$

$$\mu_r(\lambda_0) = 1 + \frac{p_m}{1 + (N_m^{-1} - i\lambda_m\lambda_0^{-1})^2}. \quad (22)$$

Here, $\lambda_0 = 2\pi/k_0$ is the free-space wavelength, $p_{e,m}$ are the respective oscillator strengths, $\lambda_{e,m}(1 + N_{e,m}^{-2})^{-1/2}$ are the resonance wavelengths, while $\lambda_{e,m}/N_{e,m}$ are the resonance linewidths.

Figure 1 comprises plots of the real and imaginary parts of ϵ_r and μ_r as functions of λ_0 , when $p_e = 1$, $p_m = 0.8$, $N_e = N_m = 100$, $\lambda_e = 0.3$ mm and

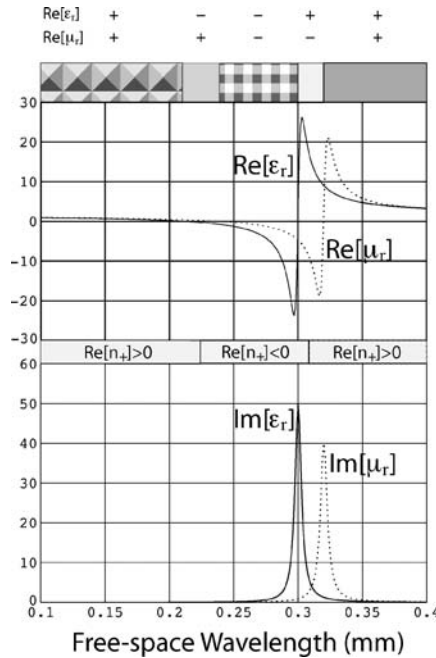


Figure 1 Real and imaginary parts of the relative permittivity and relative permeability according to (21) and (22), respectively, when $p_e = 1$, $p_m = 0.8$, $N_e = N_m = 100$, $\lambda_e = 0.3$ mm and $\lambda_m = 0.32$ mm.

$\lambda_m = 0.32$ mm. Clearly, five separate spectral regions can be identified in this figure. At the either extremity of the horizontal axis are the two regions wherein $\epsilon'_r > 0$ and $\mu'_r > 0$. In the neighborhood of $\lambda_0 = 0.22$ mm, $\epsilon'_r < 0$ but $\mu'_r > 0$. Both $\epsilon'_r < 0$ and $\mu'_r < 0$ in the neighborhood of $\lambda_0 = 0.25$ mm. Finally, $\epsilon'_r > 0$ but $\mu'_r < 0$ around $\lambda_0 = 0.31$ mm. Of course, both $\epsilon''_r > 0$ and $\mu''_r > 0$ for all λ_0 .

Calculations confirm that the spectral region wherein the inequality (18)—or, (19), equivalently—is satisfied is larger than the middle region wherein the condition (20) holds. The former region covers parts of the adjoining regions. Furthermore, at the resonance of the permittivity, $\epsilon'_r = 0$ and $\epsilon''_r > 0$ so that negative phase-velocity is possible at the electric resonance frequency only if $\mu'_r < 0$. Likewise, negative phase-velocity is possible at the magnetic resonance frequency only if $\epsilon'_r < 0$. Whereas the former possibility is exemplified in Figure 1, the latter is not. In other words, the occurrence of a negative phase-velocity depends on the ordering and proximity of the resonances of ϵ_r and μ_r .

2.4 Reflection and refraction

In the five spectral regions identified, the isotropic dielectric-magnetic medium would respond differently to monochromatic electromagnetic excitation. Suppose that a plane wave is normally incident on a half-space occupied by this medium. The reflectance $R(\lambda_0)$ is then given by the standard expression

$$R(\lambda_0) = \left| \frac{+\sqrt{\mu_r(\lambda_0)/\epsilon_r(\lambda_0)} - 1}{+\sqrt{\mu_r(\lambda_0)/\epsilon_r(\lambda_0)} + 1} \right|^2, \quad (23)$$

where $0 \leq R \leq 1$ for all λ_0 by virtue of the principle of conservation of energy. The reflectance spectrum calculated with the constitutive parameters used for Figure 1 is shown in Figure 2. The reflectance is markedly high in the two regions wherein ϵ'_r and μ'_r have opposite signs, but not in the other three regions. The reflectance is particularly low in the leftmost and the rightmost regions ($\epsilon'_r > 0$ and $\mu'_r > 0$) because the ratio μ_r/ϵ_r is close to unity therein. However, the reflectance is somewhat higher in the central region ($\epsilon'_r < 0$ and $\mu'_r < 0$) because $|\mu_r/\epsilon_r| < 0.25$.

The numerical results presented in Figure 2 have been validated by time-domain calculations [24]. This is a significant fact, because time-domain calculations do not involve the refractive index. Furthermore, time-domain calculations also confirmed the direction of power flow and attenuation that must be decided *a priori* to derive (23).

Expressions can be easily derived for reflection and refraction of obliquely incident plane waves as well [25-26]. The direction of the refracted power flow for positive and negative phase-velocity conditions is depicted in Figure 3. The direction of wave propagation is that of the power flow (and attenuation); furthermore, the electromagnetic momentum is in the same direction as the Poynting vector. Confusion of that direction with that of the phase-velocity leads to wrong conclusions such as that “negative phase velocity means that the microwave appears to move backwards” [27] or that radiation pressure can be negative [8].

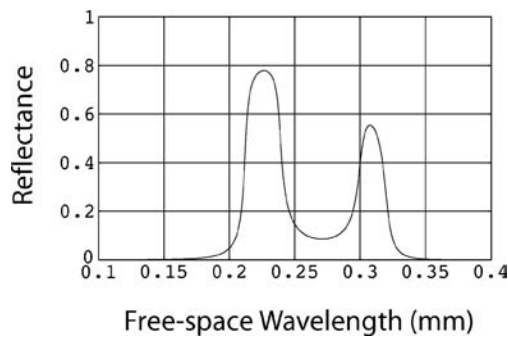


Figure 2 Plane wave reflectance $R(\lambda_0)$ calculated with the constitutive parameters depicted in Figure 1.

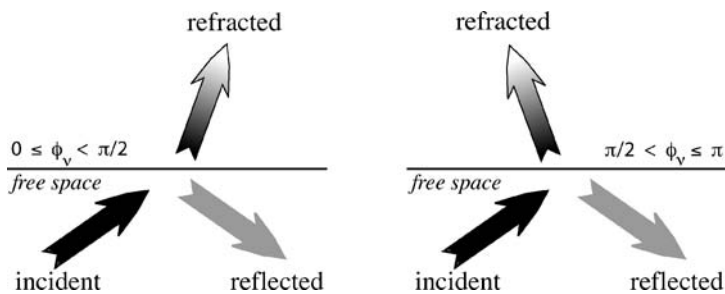


Figure 3 Reflection and refraction of an obliquely incident plane wave at the planar interface of a homogeneous, isotropic dielectric-magnetic medium with free space. The arrows show the directions of the time-averaged Poynting vector, and are graded to indicate attenuation. In the medium of refraction, the phase-velocity is directed co-parallel to the time-averaged Poynting vector when $0 \leq \phi_v < \pi/2$, and anti-parallel when $\pi/2 < \phi_v \leq \pi$.

3 Experimental evidence

3.1 Ring-wire composite material

Although Veselago [8] had predicted several unusual properties of materials with negative real permittivity and permeability, as mentioned earlier, the electromagnetics community bestowed attention on the topic only after Smith *et al.* [12] reported the fabrication of such materials.

The materials of Smith *et al.* consist of various inclusions of conducting rings and wires embedded within printed circuit boards. Predictions of their effective constitutive properties are based on observations from two separate composite mediums.

Medium 1 is a lattice of ringlike inclusions, which for a certain field configuration was presumed to have a resonant relative permeability of the form [11-12]

$$\mu_r(\omega) = 1 - \frac{F\omega^2}{\omega^2 - \omega_0^2 + i\omega\Gamma}, \quad (24)$$

where the resonance angular frequency ω_0 depends principally on the geometry of the rings. In this model, dissipation is described by Γ , and F ($0 < F < 1$) is the ratio of the area occupied by a ring and that of a unit cell. For weak dissipation, the real part of μ_r is negative for $\omega_0 < \omega < \omega_0/\sqrt{1-F}$. The form

$$\mu_r(\omega) = 1 - \frac{G}{\omega^2 - \omega_0^2 + i\omega\Gamma} \quad (25)$$

has also been used [19], [28], but it does not differ significantly from (24) in the resonance regime.

Medium 2 is an included matrix of wires. The effective relative permittivity of this composite medium supposedly displays plasmalike behavior according to [29]

$$\epsilon_r(\omega) = 1 - \frac{\omega_p^2}{\omega(\omega + i\gamma)}, \quad (26)$$

the effective plasma frequency ω_p being dependent on the geometry, and the damping factor γ being small enough to ignore [28]. In such a medium, ϵ_r is negative for $\omega < \omega_p$. A modification of (26) is [19]

$$\epsilon_r(\omega) = 1 - \frac{G'}{\omega^2 - \omega_0'^2 + i\omega\Gamma'}, \quad (27)$$

which is analogous to (25). Equation (26) has been strongly disputed [30-31], however.

Smith *et al.* [12], [32] combined the two foregoing composite materials into a composite material containing both rings and wires. Presenting some numerical and experimental evidence, they argued that the effective properties of their ring-wire composite material would be described by (24)–(27)—i.e., it would exhibit negative real-valued permeability and negative real-valued permittivity within a certain frequency range. We must note that the application of (27) for the ring-wire composite material appears doubtful, according to established notions of homogenization theory [13].

3.2 The crucial observation

A very striking report [19] published in 2001 gave preliminary indications of the observation of the so-called *negative index of refraction*. A ~ 10 GHz beam was made to normally illuminate one face of a wedge, and the angular pattern of the intensity of the beam transmitted through the other face of the wedge was measured. When a Teflon wedge was replaced by a wedge made of the ring-wire composite material, the transmission intensity pattern flipped about the normal to the exit face.

This report did not go unchallenged. First and foremost, the ring-wire composite material is not likely to be isotropic in a strict sense [33–34]. Second, there is concern regarding homogeneity of the ring-wire composite material [33]. Third, losses due to scattering and absorption do not appear to have been seriously accounted for [13], [33–34]. A final issue is that the transmission intensity may not have been measured in the far zone, and so the measurements may have been corrupted by intermediate- or Fresnel-zone effects [13], [34].

Despite these doubts, the crucial flipping of the transmission pattern about the normal to the exit face, when a Teflon wedge was replaced by a ring-wire composite material wedge, appears to be unexplainable in any way other than by resorting to the essence of Veselago’s suggestion. The rules for refraction depicted in Figure 3 can be easily used to explain the flipping, as schematically illustrated in Figure 4.

Examining planewave scattering at planar interfaces, Valanju *et al.* [34] came to the conclusion that the ring-wire composite material of Shelby *et al.* [19] violates causality, because it requires a signal to negotiate a certain distance inside the meta-material instantaneously. However, their conclusion is based on an erroneous definition of the group velocity vector emanating from a “group Snell’s law” that has no foundation in electromagnetic theory; see Chen [2] for the correct definition of the group velocity vector. Lorentz oscillator models, which are by nature causal, permit the real parts of the permittivity and permeability to be negative [35], as we saw in Section 2.3.

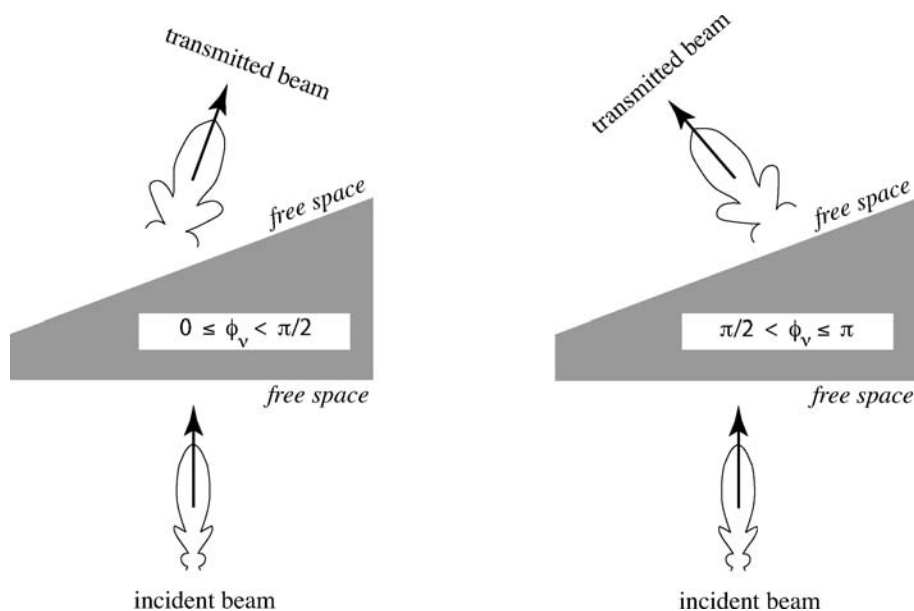


Figure 4 Transmission of a beam through a wedge. The main lobe directions of the incident and the transmitted beams can be derived from the rules of refraction depicted in Figure 3. Note the flipping of the transmission pattern about the normal to the exit face when $\pi/2 < \phi_v \leq \pi$ in relation to that for the usual case $0 \leq \phi_v < \pi/2$.

4 Terminology

The emergence of a clear terminology is often a difficult process with regards to scientific findings relating to novel effects, something that is also apparent in the present instance. The first label for the candidate materials is *left-handed materials* [8]. But chiral materials are important subjects of electromagnetics research and the terms *left-handedness* and *right-handedness* have been applied to the molecular structure of such materials for well over a century [36–37]. The continued use of the term *left-handed materials* (LHMs) for achiral materials will thus confuse crucial microstructural issues [12], [14–15], [17], [19], [28].

The term *backward* (BW) medium has been advanced by Lindell *et al.* [38]. But every wave propagates forward. Regardless of the merits of any *a priori* definitions of forward and backward directions for planewave propagation, this term would be problematic in the context of nonplanar interfaces. Ziolkowski and Heyman [22] introduced the technical term *double negative* (DNG) medium to indicate the condition of the real parts of both permittivity and permeability being negative. As we have seen in Section 2.2, that is not a necessary condition for the phase-velocity to be negative.

The term *negative-index medium* (NIM) of Valanju *et al.* [34] should also be avoided, because it is really the phase-velocity that is negatively directed. Indeed, after a careful study of the relevant constitutive parameters, we have come to the conclusion that the term *negative phase-velocity* (NPV) medium is unambiguous and is not in conflict with any physical scenario to which it is applicable. It also provides a contrast to the emerging *negative group-velocity* (NGV) mediums [39]–[41].

5 Research trends

5.1 Perfect lenses

Excitement about NPV mediums grew after Pendry [42] presented the possibility of fabricating a lens that could circumvent the diffraction limit—i.e., a *perfect lens*—using a medium with $\epsilon_r = \mu_r = -1$. This material has identical intrinsic impedance but negative wavenumber with respect to vacuum [33], which is responsible for the absence of aberrations [43–44]. Attention, often critical, came rapidly from researchers thereafter [22–23], [34], [44]–[48].

Ziolkowski and Heyman [22] concluded from extensive two-dimensional simulations that the condition $\epsilon_r = \mu_r = -1$ cannot be met by realistic materials, even in some narrow frequency range. Distortions due to dissipation inside the desired NPV medium (with $\epsilon_r = \mu_r = -1$) would prove to be a stumbling block in fabricating the desired perfect lenses [44], [48], dissipation being necessarily dispersive [49]. However, the possible use of active (i.e., nonpassive) elements in the ring-wire composite materials may obviate this difficulty to an adequate extent in some frequency range. Chromatic aberrations due to nonfulfilment of the required conditions outside some narrow frequency range will also be important [34], [44].

5.2 Unusual composite materials

At sufficiently low frequencies, it is well known that a homogeneous medium can be prescribed as *effectively* equivalent to a particulate composite material for certain purposes [50]–[52]. Artificial materials with NPV inclusions or host materials are therefore an attractive topic for theoretical research.

The incorporation of NPV mediums in such well-known homogenization approaches, as the ones named after Maxwell Garnett and Bruggeman [50], is mathematically trivial. Whether the theoretical predictions of those approaches will be physically realized is another matter. We do, however, note that surprising results can emerge from the consideration of NPV mediums in homogenization approaches [33]. For instance, the Bruggeman approach forecasts that a certain mixture of a NPV medium with its prosaic positive phase-velocity (PPV) counterpart, both impedance-matched, can function as the medium that is apparently desirable for fabricating perfect lenses [42], [44]. Likewise, the Maxwell Garnett approach predicts that composite mediums with zero permittivity and zero permeability can be made as electrically small NPV inclusions dispersed randomly but homogeneously in a PPV host medium [33].

5.3 Planar technology

As the NPV mediums have been realized in the form of multilaminar slabs, their integration with planar technology is readily possible. New effects are likely to emerge, and extant understanding of electromagnetism shall be refined by theoretical research on the incorporation of NPV mediums in basic multilayer problems.

As an example, the Bragg regime of a multilayer distributed Bragg reflector (DBR) would undergo a blue-shift, if both ϵ'_r and μ'_r were changed from positive to negative [53]. The spectral responses of fractal multilayers are also affected [54]. An underlying cause is the reversal of phase of the reflected and the transmitted plane waves [55]. Multilayer DBRs with NPV constituents could be useful in wavelength regimes that are inaccessible—due to inconvenient mechanical and/or thermal constraints—with DBRs employing PPV constituents exclusively.

5.4 Complex materials

Reversal of the signs of the real parts of both permittivity and permeability can amount to a change of handedness in certain instances. Thus, the circular Bragg phenomenon will be reversed in ferrocholesteric materials with negative real permittivities and permeabilities [56]. Even in isotropic chiral mediums, negative ϵ'_r and μ'_r will lead to a reversed circular dichroism [57].

6 Concluding remarks

This chapter provides an overview of various emerging issues relating to negative phase-velocity (NPV) mediums, and is expected to stimulate different lines of

thought in the electromagnetics research community. In view of claims and counterclaims launched respectively by Shelby *et al.* [19] and the detractors of their experimental results, the scientific situation is presently volatile. Although experimental confirmation from other laboratories appears very desirable, in our opinion it is inadvisable for electromagnetics researchers to ignore the topic.

Postscript by Lakhtakia and McCall: About three months after this chapter was finalized, Parazzoli *et al.* [58] reported their experimental verification of the findings of Shelby *et al.* [19]. The two sets of experiments were conducted on similar samples, and in essentially similar configurations, though the former was in a free-space geometry as opposed to the waveguide setup of the latter. The new results, which include measurements taken at several detector distances, should lay to rest doubts on the corruption of the measured transmission intensities by intermediate-zone effects.

Acknowledgments

A.L. thanks the Weiss Graduate Fellowship Program at the Pennsylvania State University for partial financial support. M.W.M. acknowledges partial support from the *Engineering and Physical Sciences Research Council* of Great Britain (EPSRC grant no. GR/R55078/01).

References

- [1] J.D. Jackson, *Classical electrodynamics*, 3rd ed, Wiley, New York, NY, USA, 1999.
- [2] H.C. Chen, *Theory of electromagnetic waves*, TechBooks, Fairfax, VA, USA, 1992.
- [3] E.J. Post, *Formal structure of electromagnetics*, Dover Press, New York, NY, USA, 1997. [Reprinted from North-Holland, Amsterdam, The Netherlands, 1962.]
- [4] W.S. Weiglhofer, Constitutive characterization of simple and complex mediums, *Introduction to complex mediums for optics and electromagnetics* (W.S. Weiglhofer and A. Lakhtakia, eds), SPIE Press, Bellingham, WA, USA, 2003.
- [5] O.N. Singh and A. Lakhtakia (eds), *Electromagnetic fields in unconventional materials and structures*, Wiley, New York, NY, USA, 2000.
- [6] A. Lakhtakia, M.W. McCall, W.S. Weiglhofer, J. Gerardin and J. Wang, On mediums with negative phase velocity: a brief overview, *Complex mediums III: Beyond linear isotropic dielectrics* (A. Lakhtakia and M.W. McCall, eds), Proc SPIE **4806**, Bellingham, WA, USA, 2002, 176–180.
- [7] A. Lakhtakia, M.W. McCall and W.S. Weiglhofer, Brief overview of recent developments on negative phase-velocity mediums (alias left-handed materials), *Arch Elektron Übertrag* **56** (2002), 407–410.
- [8] V.G. Veselago, The electrodynamics of substances with simultaneously negative values of ϵ and μ , *Sov Phys Usp* **10** (1968), 509–514.

- [9] V.G. Veselago, communication to W.S. Weiglhofer on May 8, 2002. [Veselago answered a question at *Bianisotropics 2002, 9th int conf electromagnetics complex media*, Marrakech, Morocco, 2002. The answer was confirmed by Veselago during a conversation with A. Lakhtakia in Arlington, VA, USA, on May 12, 2003.]
- [10] J. Pendry, Transmission resonances on metallic gratings with very narrow slits, *Phys Rev Lett* **85** (1999), 3966–3969.
- [11] J. Pendry, A.J. Holden, D.J. Robbins and W.J. Stewart, Magnetism from conductors and enhanced nonlinear phenomena, *IEEE Trans Microwave Theory Tech* **47** (1999), 2075–2084.
- [12] D.R. Smith, W.J. Padilla, D.C. Vier, S.C. Nemat-Nasser and S. Schultz, Composite medium with simultaneously negative permeability and permittivity, *Phys Rev Lett* **84** (2000), 4184–4187.
- [13] N. Garcia and M. Nieto-Vesperinas, Is there an experimental verification of a negative index of refraction yet? *Opt Lett* **27** (2002), 885–887.
- [14] P. Markoš and C.M. Soukoulis, Numerical studies of left-handed materials and arrays of split ring resonators, *Phys Rev E* **65** (2002), 036622.
- [15] R. Marqués, F. Medina and R. Rafii-El-Idrissi, Role of bianisotropy in negative permeability and left-handed metamaterials, *Phys Rev B* **65** (2002), 144440.
- [16] G. Dewar, Candidates for $\mu < 0$, $\epsilon < 0$ nanostructures, *Int J Mod Phys B* **15** (2001), 3258–3265.
- [17] S.T. Chui and L. Hu, Theoretical investigation on the possibility of preparing left-handed materials in metallic magnetic granular composites, *Phys Rev B* **65** (2002), 144407.
- [18] P. Gay-Balmaz and O.J.F. Martin, Efficient isotropic magnetic resonators, *Appl Phys Lett* **81** (2002), 939–941.
- [19] R.A. Shelby, D.R. Smith and S. Schultz, Experimental verification of a negative index of refraction, *Science* **292** (2001), 77–79.
- [20] M.W. McCall, A. Lakhtakia and W.S. Weiglhofer, The negative index of refraction demystified, *Eur J Phys* **23** (2002), 353–359.
- [21] R. Rupp, personal communication to A. Lakhtakia on August 13, 2002.
- [22] R.W. Ziolkowski and E. Heyman, Wave propagation in media having negative permittivity and permeability, *Phys Rev E* **64** (2001), 056625.
- [23] J. Paul, C. Christopoulos and D.W.P. Thomas, Time-domain modelling of negative refractive index material, *Electron Lett* **37** (2001), 912–913.
- [24] J. Wang and A. Lakhtakia, On reflection from a half-space with negative real permittivity and permeability, *Microw Opt Technol Lett* **33** (2002), 465–467.
- [25] J.A. Kong, B.-I. Wu and Y. Zhang, Lateral displacement of a Gaussian beam reflected from a grounded slab with negative permittivity and permeability, *Appl Phys Lett* **80** (2001), 2084–2087.
- [26] J.A. Kong, B.-I. Wu and Y. Zhang, A unique lateral displacement of a Gaussian beam transmitted through a slab with negative permittivity and permeability, *Microw Opt Technol Lett* **33** (2002), 136–139. [The shifts reported

- by Kong *et al.* [25–26] should not be confused with Goos-Hänchen shifts discussed by Lakhtakia [55].]
- [27] P.F. Schewe, B. Stein and J. Riordan, *Bulletin of Physics News, Number 534* (April 13, 2001), American Institute of Physics, College Park, MD, USA, 2001. [See <http://newton.ex.ac.uk/aip/physnews.534.html>]
 - [28] D.R. Smith and N. Kroll, Negative refractive index in left-handed materials, *Phys Rev Lett* **85** (2000), 2933–2936.
 - [29] J.B. Pendry, A.J. Holden, W.J. Stewart and I. Youngs, Extremely low frequency plasmons in metallic mesostructures, *Phys Rev Lett* **76** (1996), 4773–4776.
 - [30] R.M. Walser, A.P. Valanju and P.M. Valanju, Comment on “Extremely low frequency plasmons in metallic mesostructures,” *Phys Rev Lett* **87** (1996), 119701. [The authors’ reply, if any, was not published.]
 - [31] A.L. Pokrovsky and A.L. Efros, Electrodynamics of metallic photonic crystals and the problem of left-handed materials, *Phys Rev Lett* **89** (2002), 093901.
 - [32] R.A. Shelby, D.R. Smith, S.C. Nemat-Nasser and S. Schultz, Microwave transmission through a two-dimensional, isotropic, left-handed material, *Appl Phys Lett* **78** (2001), 489–491.
 - [33] A. Lakhtakia, An electromagnetic trinity from “negative permittivity” and “negative permeability,” *Int J Infrared Millim Waves* **22** (2001), 1731–1734; correctly reprinted: **23** (2002), 813–818.
 - [34] P.M. Valanju, R.M. Walser and A.P. Valanju, Wave refraction in negative-index media: always positive and very inhomogeneous, *Phys Rev Lett* **88** (2002), 187401.
 - [35] C.F. Bohren and D.R. Huffman, *Absorption and scattering of light by small particles*, Wiley, New York, NY, USA, 1983.
 - [36] A. Lakhtakia (ed), *Selected papers on natural optical activity*, SPIE Press, Bellingham, WA, USA, 1990.
 - [37] A. Lakhtakia, *Beltrami fields in chiral media*, World Scientific, Singapore, 1994.
 - [38] I.V. Lindell, S.A. Tretyakov, K.I. Nikoskinen and S. Ilvonen, BW media—media with negative parameters, capable of supporting backward waves, *Microw Opt Technol Lett* **31** (2001), 129–133.
 - [39] L.J. Wang, A. Kuzmich and A. Dogariu, Gain-assisted superluminal light propagation, *Nature* **406** (2000), 277–279.
 - [40] K.T. McDonald, Negative group velocity, *Preprint physics/0008013* (<http://arXiv.org/>).
 - [41] C.-G. Huang and Y.-Z. Zhang, Negative group velocity and distortion of a pulse in an anomalous dispersion medium, *J Opt A: Pure Appl Opt* **4** (2002), 263–270.
 - [42] J.B. Pendry, Negative refraction makes a perfect lens, *Phys Rev Lett* **85** (2001), 3966–3969.

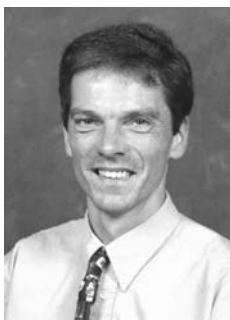
- [43] Z.M. Zhang and C.J. Fu, Unusual photon tunneling in the presence of a layer with a negative refractive index, *Appl Phys Lett* **80** (2002), 1097–1099.
- [44] A. Lakhtakia, On perfect lenses and nihility, *Int J Infrared Millim Waves* **23** (2002), 339–343.
- [45] G.W. 't Hooft, Comment on “Negative refraction makes a perfect lens,” *Phys Rev Lett* **87** (2001), 249701; J. Pendry, Reply, *Phys Rev Lett* **87** (2001), 249702.
- [46] J.M. Williams, Some problems with negative refraction, *Phys Rev Lett* **87** (2001), 249703; J. Pendry, Reply, *Phys Rev Lett* **87** (2001), 249704.
- [47] R. Ruppin, Surface polaritons of a left-handed material slab, *J Phys: Condens Matter* **13** (2001), 1811–1819.
- [48] N. Garcia and M. Nieto-Vesperinas, Left-handed materials do not make a perfect lens, *Phys Rev Lett* **88** (2002), 207403.
- [49] W.S. Weiglhofer and A. Lakhtakia, On causality requirements for material media, *Arch Elektron Übertrag* **50** (1996), 389–391.
- [50] A. Lakhtakia (ed), *Selected papers on linear optical composite materials*, SPIE Press, Bellingham, WA, USA, 1996.
- [51] B. Michel, Recent developments in the homogenization of linear bianisotropic composite materials, *Electromagnetic fields in unconventional materials and structures* (O.N. Singh and A. Lakhtakia, eds), Wiley, New York, NY, USA, 2000.
- [52] T.G. Mackay, Homogenization of linear and nonlinear complex composite materials, *Introduction to complex mediums for optics and electromagnetics* (W.S. Weiglhofer and A. Lakhtakia, eds), SPIE Press, Bellingham, WA, USA, 2003.
- [53] J. Gerardin and A. Lakhtakia, Negative index of refraction and distributed Bragg reflectors, *Microw Opt Technol Lett* **34** (2002), 409–411.
- [54] J. Gerardin and A. Lakhtakia, Spectral response of Cantor multilayers made of materials with negative refractive index, *Phys Lett A* **301** (2002), 377–381.
- [55] A. Lakhtakia, On planewave remittances and Goos-Hänchen shifts of planar slabs with negative real permittivity and permeability, *Electromagnetics* **23** (2003), 71–75.
- [56] A. Lakhtakia, Reversal of circular Bragg phenomenon in ferrocholesteric materials with negative real permittivities and permeabilities, *Adv Mater* **14** (2002), 447–449.
- [57] A. Lakhtakia, Reversed circular dichroism of isotropic chiral mediums with negative real permeability and permittivity, *Microw Opt Technol Lett* **33** (2002), 96–97.
- [58] C.G. Parazzoli, R.B. Greegor, K. Li, B.E.C. Koltenbah and M. Tanielian, Experimental verification and simulation of negative index of refraction using Snell’s law, *Phys Rev Lett* **90** (2003), 107401.



Akhlesh Lakhtakia is a Professor of Engineering Science and Mechanics at the Pennsylvania State University. He has published more than 440 journal articles; has contributed chapters to eight research books; has edited, co-edited, authored or co-authored eight books and four conference proceedings; has reviewed for 65 journals; and was the Editor-in-Chief of the international journal *Speculations in Science and Technology* from 1993 to 1995. He headed the IEEE EMC Technical Committee on Non-sinusoidal Fields from 1992 to 1994; and is a Fellow of Optical Society of America, SPIE—The International Society for Optical Engineering, and the Institute of Physics (UK). He also served as the 1995 Scottish Amicable Visiting Lecturer at the University of Glasgow. He was awarded the PSES Outstanding Research Award in 1996. Since 1999, he has organized four annual SPIE conferences on nanostructured materials and complex mediums. His current research interests lie in the electromagnetics of complex mediums, sculptured thin films, and chiral nanotubes.



Martin W. McCall was born in London, UK, in 1962. He is currently a Reader in Physics at Imperial College, London. Throughout his career he has worked at various times on nonlinear optics, nonlinear dynamics and electromagnetics of complex mediums. He is author or co-author of about 90 papers, and an undergraduate text on classical mechanics.



Werner S. Weiglhofer was born in Bruck an der Mur, Austria where he obtained the degrees of Dipl.-Ing. and Dr. Techn. from the Technische Universität Graz, Austria. He joined the University of Glasgow in 1988 as a Research Assistant in the Department of Mathematics where he became a Professor of Applied Mathematics. He authored/coauthored 135 refereed journal publications and was the author of an undergraduate mathematics textbook. He died in a tragic mountaineering accident on January 12, 2003, in Norway.

Scattering Theory of Photonic Crystals

Didier Felbacq and Frédéric Zolla

Abstract

This chapter is devoted to the theoretical and numerical study of scattering of electromagnetic waves. We begin with a detailed study of the so-called one-dimensional problem, paying special attention to the link between the problem of propagation in an infinite periodic device and the problem of diffraction by a multilayered structure (Bragg mirror). In addition, we deal with the introduction of a defect in a periodic structure. Scattering by parallel rods is described, through the concept of scattering matrix and the Gamow vectors. We then take up the notion of defect for a two-dimensional crystal and tackle boundary effects in finite-sized crystals. The chapter concludes with some current problems encountered for photonic crystals.

1 Introduction

The motion of electrons in natural crystals can be accounted for by band theory, in which the interaction between electrons is neglected. In this theory, one considers the Schrödinger equation with a periodic potential, representing the forces due to the atomic sites. The energies for which electrons can propagate form bands separated by gaps that are bands of forbidden energies. A similar phenomenon exists for photons, or electromagnetic waves, when they propagate in dielectric structures that are periodically nonhomogeneous. In optics, such devices have been used for years under the name of Bragg mirrors. These are artificial devices fabricated as stacks of dielectric slabs of two types: ones with a high refractive index and others with a low refractive index. These structures are one-dimensional. The existence of a photonic bandgap makes them suitable for the realization of filters or mirrors in optics. However, the photonic bandgap is very sensitive to the angle of incidence. In order to remedy this point, the idea of extending the periodicity to the other dimensions of space was suggested in 1987 by John [1] and Yablonovitch [2]. This amounts to designing a periodically modulated device, with forbidden bands of frequencies in all directions of space. Such a structure is called a photonic crystal. In this way, one can imagine that it could be possible to inhibit the spontaneous emission of an atom situated inside such a crystal, and thus to design cavities with very high quality factors. Since 1987, research on photonic crystals has proliferated [3]-[6]. As a matter of fact, it should be understood that the term *photonic crystal* can be applied generically to any periodically structured device. Thus, diffraction gratings [7] can be considered as photonic crystals; and there is indeed a growing interest in the propagation of electromagnetic surface modes on periodically structured surfaces [8].

Unlike natural crystals, which comprise a very large number of atomic sites, actual photonic crystals possess only a limited number of periods, which renders the boundary effects crucial. Another important point is that of the excitation of the internal modes of a device by an external source. The density of states, as obtained through Bloch theory, cannot account for the coupling coefficient between a mode and a source. In order to do so, one has to take into account the fact that the device is an open cavity and that the modes can escape the structure. Smaïli *et al.* [9] have demonstrated the importance of evanescent fields, i.e., non-Bloch modes, in the electromagnetic behavior of finite two-dimensional structures, whereas other authors made headway toward a clear definition of the density of modes for a finite structure and also for generalizing the notion of normal modes to open cavities [10-11]. The development of a coherent theory describing the coupling of a continuum of modes with the discrete leaky modes of an open cavity is of fundamental importance for the expected application of photonic devices to, e.g., quantum information, where the quantization of the electromagnetic field cannot be avoided, and hence a complete set of modes is needed.

In this chapter, our aim is to develop the basic theoretical tools necessary for a specific study of the interaction of classical electromagnetic waves with photonic

crystals. More precisely, we give a rather extended account of the scattering theory of one-dimensional and two-dimensional structures as well as some applications to the study of cavities and guided modes.

2 Scattering theory of photonic crystals

2.1 One-dimensional photonic crystals

2.1.1 The problem of propagation

In order to introduce basic concepts, we start by considering the case of one-dimensional photonic crystals; see Figure 1. For the sake of simplicity, we suppose throughout this chapter that the permeability is equal to μ_0 , which is the permeability of free space. The devices are optically characterized by the relative permittivity $\varepsilon(x)$, which is a periodic function of x with period d : $\varepsilon(x + d) = \varepsilon(x)$. Our aim is to describe the electromagnetic field that can exist in such an infinite structure. More precisely, we look for solutions of the Maxwell equations without an incident field—the so-called eigenmodes of the structure. These solutions are harmonic in time, with a dependence of $\exp(-i\omega t)$. The time-harmonic Maxwell curl equations are

$$\begin{aligned}\nabla \times \mathbf{E} &= i\omega\mu_0\mathbf{H}, \\ \nabla \times \mathbf{H} &= -i\omega\varepsilon_0\varepsilon(x)\mathbf{E},\end{aligned}\tag{1}$$

where ε_0 is the permittivity of free space.

In order to simplify the analysis, we look for solutions that are z -independent. Under this assumption, the vectorial system (1) can be decomposed into two scalar problems. These problems corresponds to linearly polarized waves. The first case

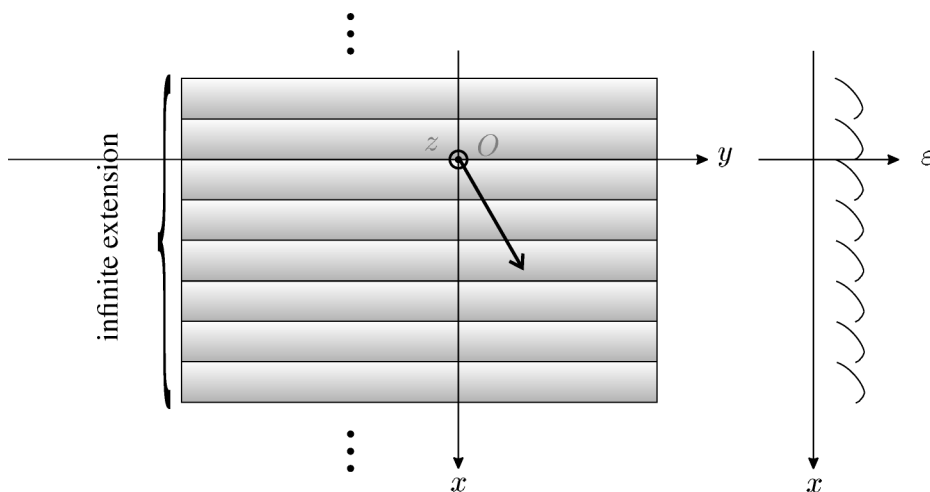


Figure 1 Schematic for propagation in an infinitely extended, periodic medium.

is the so-called $E//$ case in which the electric field is aligned parallel to the z axis: $\mathbf{E} = [0, 0, u(x, y)]e^{-i\omega t}e_z$. The second case is the so-called $H//$ case, wherein the magnetic field is entirely along e_z : $\mathbf{H} = [0, 0, u(x, y)]e^{-i\omega t}e_z$.

The next step consists in using the y invariance of the system by considering the spatial Fourier transform with respect to y . We look for solutions of the form $u(x, y) = v(x)\exp(i\alpha y)$, where α is the Fourier variable associated with y . With these assumptions, we obtain

$$\begin{cases} \frac{d^2v}{dx^2} + [k_0^2\varepsilon(x) - \alpha^2]v = 0, & (E//), \\ \frac{d}{dx}\left[\frac{1}{\varepsilon(x)}\frac{dv}{dx}\right] + \left[k_0^2 - \frac{\alpha^2}{\varepsilon(x)}\right]v = 0, & (H//), \end{cases} \quad (2a) \quad (2b)$$

where $k_0 = \omega\sqrt{\varepsilon_0\mu_0}$ is the free-space wavenumber.

The periodicity of the permittivity allows us to apply the Floquet–Bloch theorem. According to this theorem, all the propagating solutions can be obtained by imposing pseudoperiodic boundary conditions. This means that we can reduce our study to the unit cell $[0, d]$ of $\varepsilon(x)$, and look for solutions v such that $v(x + d) = e^{ikd}v(x)$, where k belongs to the interval $Y' = [-\pi/d, \pi/d]$. In other words, all the propagating solutions are parametrized by Y' .

In order to justify this result, let us remark that a solution of (2a) or (2b) is completely characterized by its initial values $[v(0), v'(0)]$, where $v'(0) = dv/dx|_{x=0}$. It is convenient to introduce the column vector

$$V(x) = \begin{bmatrix} v(x) \\ p(x)v'(x) \end{bmatrix}, \quad (3)$$

where

$$\begin{cases} p \equiv 1, & (E//), \\ p \equiv \varepsilon^{-1}, & (H//). \end{cases} \quad (4a) \quad (4b)$$

Our aim is to construct a transmission matrix T , that is, the 2×2 matrix such that

$$V(d) = TV(0). \quad (5)$$

Remark 2.1 The column vector V is a continuous function of x because equations (2a) and (2b) imply that $v(x)$ and $p(x)v'(x)$ are both continuous functions, even for the points of discontinuity of the function ε . Abusing notation ever so slightly, we write $p(0)v'(0)$ (resp. $p(d)v'(d)$) instead of $\lim_{x \rightarrow 0^+} p(x)v'(x)$ (resp. $\lim_{x \rightarrow d^-} p(x)v'(x)$) in the remainder of this chapter.

In order to obtain the matrix T in closed form, we consider two particular solutions χ_1 and χ_2 defined by their initial conditions as follows:

$$\begin{cases} \chi_1(0) = 1, & p(0)\chi_1'(0) = 0, \end{cases} \quad (6a)$$

$$\begin{cases} \chi_2(0) = 0, & p(0)\chi_2'(0) = 1. \end{cases} \quad (6b)$$

Any solution $v(x)$ of (2a) or (2b) can be written in terms of $\chi_1(x)$ and $\chi_2(x)$ as

$$v(x) = v(0)\chi_1(x) + p(0)v'(0)\chi_2'(x), \quad (7)$$

and we obtain

$$\begin{cases} v(d) = v(0)\chi_1(d) + p(0)v'(0)\chi_2(d) \\ v'(d) = v(0)\chi_1'(d) + p(0)v'(0)\chi_2'(d) \end{cases} \quad (8)$$

These relations are equivalent to the matrix relation (5); hence, we get the transmission matrix

$$T = \begin{bmatrix} \chi_1(d) & \chi_2(d) \\ p(d)\chi_1'(d) & p(d)\chi_2'(d) \end{bmatrix}. \quad (9)$$

The matrix T has the important property that its determinant $\det(T)$ is equal to unity. This result comes by considering the so-called Wronskian $W(x)$ of the system:

$$W(x) = p(x)[\chi_1(x)\chi_2'(x) - \chi_1'(x)\chi_2(x)]. \quad (10)$$

Direct manipulation of $W(x)$ shows that $dW/dx = 0$. From this property, we obtain that $W(x) = 1$, by considering $W(0)$. As a consequence, we have $W(d) = 1$ and thus $\det(T) = 1$. This important property implies that the eigenvalues of T are roots of the so-called characteristic polynomial $X^2 - \text{tr}(T)X + 1$, where $\text{tr}(T)$ denotes the trace of T .

The following three situations may occur, in consequence:

- If $|\text{tr}(T)| < 2$, both eigenvalues of T are complex-valued. They have unit magnitudes, and are the complex conjugates of each other. Therefore they can be written as e^{ikd} and e^{-ikd} , where k is a real number belonging to Y' . All the solutions are linear combinations of v_1 and v_2 , where v_1 and v_2 are such that $v_1(x+d) = e^{ikd}v_1(x)$ and $v_2(x+d) = e^{-ikd}v_2(x)$. All solutions are propagating ones and therefore bounded.
- If $|\text{tr}(T)| > 2$, both eigenvalues are real-valued and the reciprocal of each other. They can be written as γ and γ^{-1} (by convention $|\gamma| < 1$). The solutions are linear combination of v_1 and v_2 , where $v_1(x+d) = \gamma v_1(x)$ and $v_2(x+d) = \gamma^{-1}v_2(x)$. All solutions grow exponentially; and thus all the non-null solutions are unbounded.

- The case $|tr(T)| = 2$ is very particular for which the eigenvalues are 1 and -1 . The solution v_1 is periodic, whereas v_2 is antiperiodic [i.e., $v_2(x + d) = -v_2(x)$]. Both solutions are bounded.

We have obtained a splitting of the set of parameters $(\omega, \alpha) \in \mathbb{R}^+ \times \mathbb{R}$ as follows:

$$\mathcal{G} = \{(\omega, \alpha), |tr(T)| > 2\},$$

$$\mathcal{B} = \{(\omega, \alpha), |tr(T)| < 2\},$$

$$\Delta = \{(\omega, \alpha), |tr(T)| = 2\}.$$

The conclusion is that there can be propagating waves in a photonic crystal if and only if $|tr(T)| \leq 2$. The corresponding sets of wavelengths are the conduction bands $\bar{\mathcal{B}} = \mathcal{B} \cup \Delta$. Conversely, the set $\mathcal{G}$ corresponds to forbidden bands of propagation, that is, photonic bandgaps.

We have thus succeeded in justifying the Floquet–Bloch decomposition for the characterization of propagating modes. As a matter of fact, we have obtained even a stronger result—because we are able to describe the unbounded modes as well. Besides, we have seen that the value $V(x)$ of a mode at any point x in the structure can be computed, provided the value $V(x_0)$ for some x_0 is known. This remark is of fundamental interest when dealing with defects.

2.2 Numerical examples

Let us now turn to some numerical examples in order to illustrate the concepts developed above. We consider a Bragg mirror, which is the simplest example of a periodic one-dimensional structure. Its unit cell is made of two homogeneous layers with two different refractive indexes. The layer widths are chosen equal with a total thickness d . The relative permittivities are respectively 2.25 and 4.5. Figure 2 represents the band diagram for $\alpha = 0$. We can see three gaps, which are printed in gray. Another representation of the dispersion curves consists in plotting the real parts of the eigenvalues of the T matrix versus the free-space wavelength $\lambda_0 = 2\pi/k_0$. We call the resulting curve a *bifurcation diagram*; see Figure 3. This representation offers some advantages over the more traditional *band diagram* because it allows us to represent the eigenvalues in the conduction band and in the gap as well. Besides, this representation shows the connection between the evanescent and propagating modes.

The first two representations that we gave still have a significant drawback: they show nothing about the influence of α on the dispersion curves. A third representation that we call the *binary diagram* remedies this problem; see Figure 4. This diagram represents the evolution of the gaps (white zones) and conduction bands (gray zones) with respect to λ_0 and $\theta = \arcsin(\alpha/k)$. Clearly, the gaps strongly depend upon θ , which became the chief motivation for studying higher-dimensional structures.

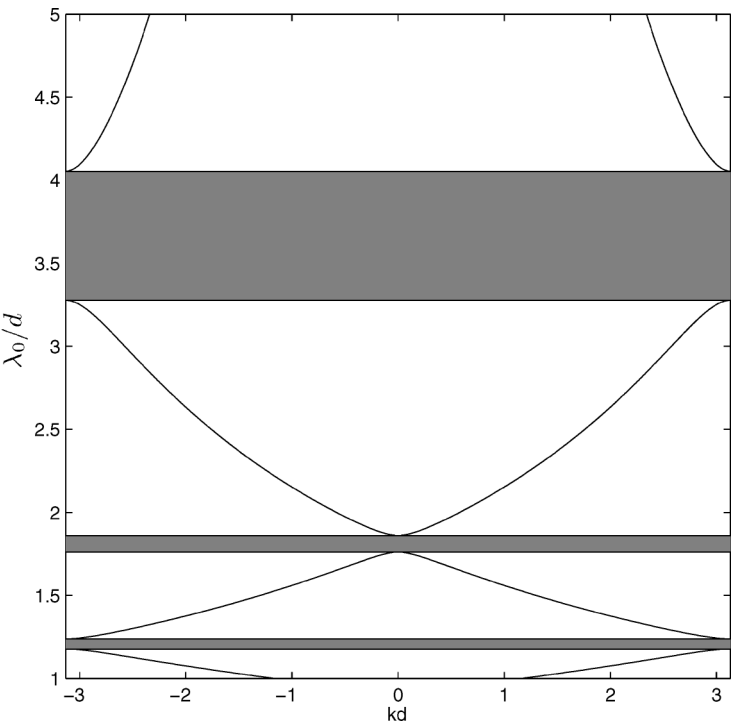


Figure 2 The band diagram for $\alpha = 0$. The normalized free-space wavelength λ_0/d is plotted versus the normalized propagation constant kd . The gaps are represented by gray regions.

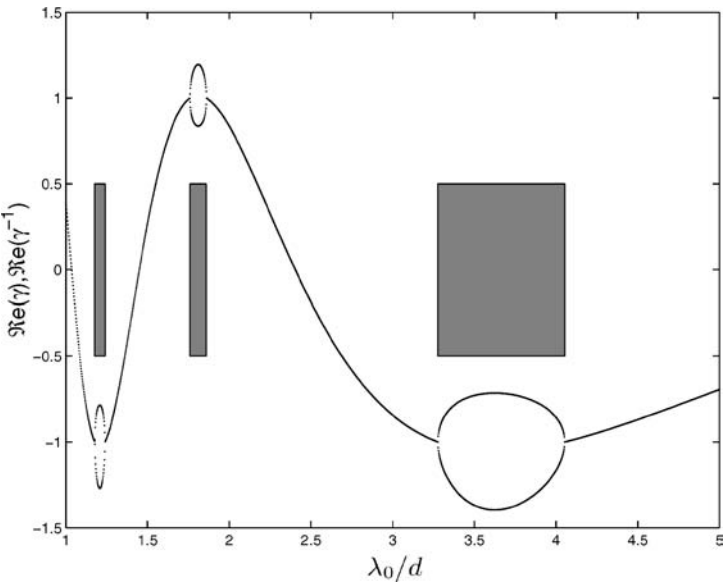


Figure 3 Bifurcation diagram, which is merely another representation of the dispersion curves. The real parts of the eigenvalues γ and γ^{-1} are plotted versus the normalized wavelength λ_0/d . Note that the gaps are represented by gray regions.

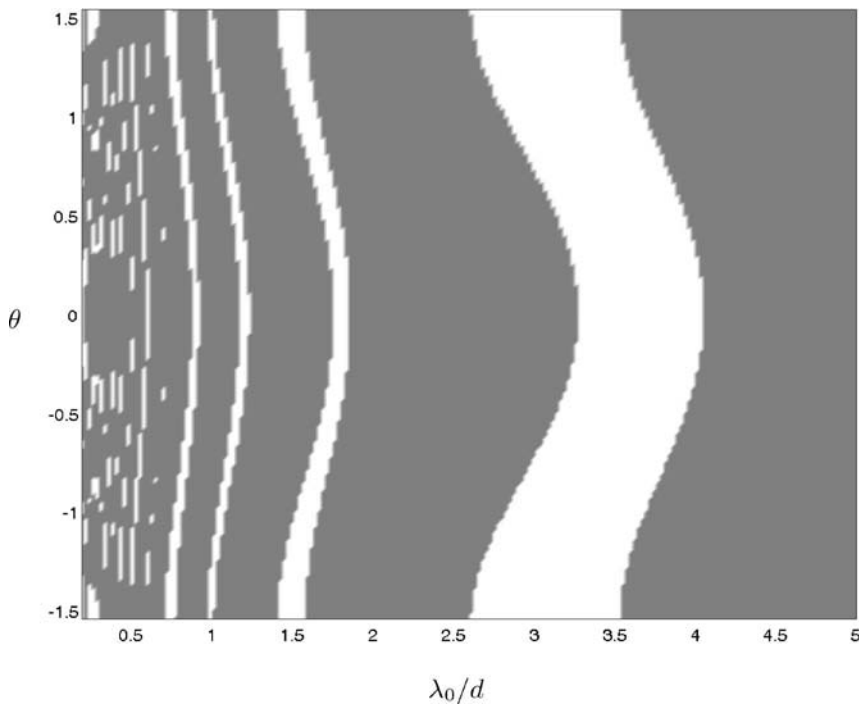


Figure 4 Binary diagram, in which the gaps $\mathcal{G}$ are in white and the conduction band $\mathcal{B}$ in gray, displaying θ and the normalized dependence on free-space wavelength λ_0/d .

2.3 Defect in infinitely extended periodic medium

With the goal of designing new materials that have an influence on light similar to that of semiconductors on electrons, researchers turned to photonic crystals. When the atoms are periodically arranged in a silicon sample, interference between the wavefunctions of electrons create ranges of energies within which electrons cannot propagate. The introduction of “impurities” or defects in a crystal (impure crystal) allows the emergence of a new type of modes: the so-called *localized states*. These modes have a fundamental influence on the flow of electrons and lead to a substantial modification of the conductivity of the crystal.

In order to create an analogous effect with light, some researchers built sets of layers (one-dimensional devices), sets of rods (two-dimensional devices) or sets of judicious galleries in bulk (Yablonovite for the three-dimensional devices) ... If one inserts a defect in a crystal by enlarging a layer or by removing a rod, light can localize in the defect. Loosely speaking, one can say that light is then a prisoner of the defect; and if light “tries” to escape, it is rebuffed by the crystal. That means that one can expect that these special modes lie in the bandgap of a pure crystal.

The simplest example of a defect in an infinitely extended periodic structure is a supernumerary layer, as shown in Figure 5. We can define a transmission matrix T_0 for this defect layer. Our aim is to analyze the influence of the defect on the modes. More precisely, two distinct points should be addressed:

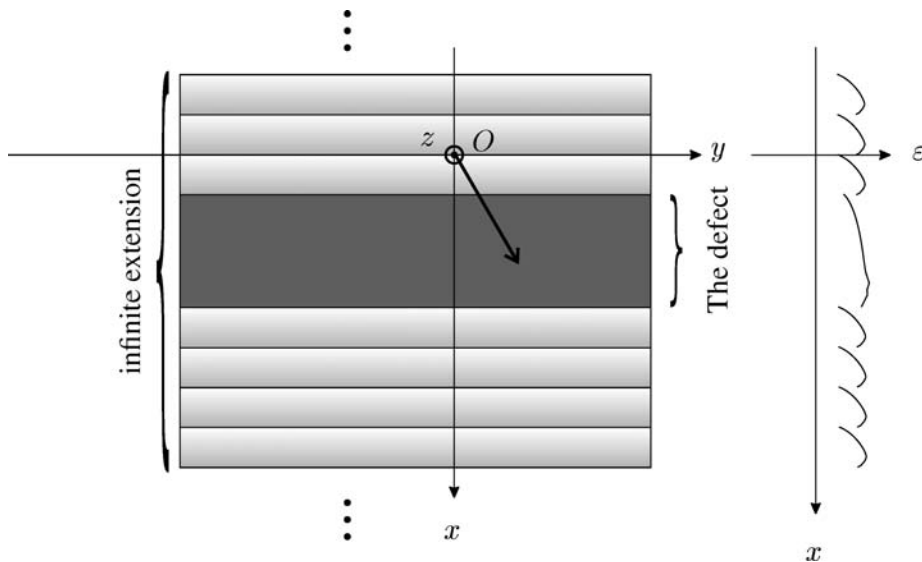


Figure 5 Schematic for propagation in an impure crystal (defect introduced in a pure crystal). The defect lays between x_d and $x_d + h$.

- Does the set $\tilde{\mathcal{B}}$ still correspond to propagating modes only?
- Does the set $\mathcal{G}$ still correspond to unbounded solutions only?

The introduction of the defect breaks the periodicity of the structure; therefore we cannot, strictly speaking, use Bloch waves. However, the transmission matrix can still be used to understand the propagation of waves in an impure crystal.

In order to answer the first question, let us assume that there exists a propagating mode at some point x_0 inside the crystal and for parameters belonging to $\mathcal{B}$. This mode is defined by the initial condition $V(x_0)$. How will this mode evolve spatially in presence of the defect?

Outside the defect, the field can only be that of a propagating mode. It evolves according to the transmission matrix of (5). If the defect is assumed to extend over $[x_d, x_d + h]$, the behavior of the field is obtained by setting $V(x_d + h) = T_0 V(x_d)$. Indeed, for $x > x_d + h$, we have $V(x + nd) = T^n V(x)$. As the matrix T only has propagating eigenvectors, the field remains propagating. As a consequence, the introduction of a defect does not modify the conduction bands.

Let us now turn to the second question. In a pure crystal, no bounded solutions exist. By introducing a defect, we can try to construct a bounded solution for parameters (ω, α) belonging to $\mathcal{G}$. To do so, we have to impose a decreasing behavior on both sides of the defect; this means that the field increases from $-\infty$ to x_d and it also decreases from $x_d + h$ to $+\infty$. In order for such a field to exist, the defect has to switch an exponentially growing eigensolution onto an exponentially decreasing one. To turn this idea into an algebraic equation, let us write the matrix T_0 in the basis of the eigenvectors of T . We choose an ordered pair of eigenvectors $(\mathbf{u}, \mathbf{v})$ such that $T\mathbf{u} = \gamma\mathbf{u}$ and $T\mathbf{v} = \gamma^{-1}\mathbf{v}$, where $|\gamma| < 1$. Denoting by P the matrix

corresponding to this change of coordinates, we have

$$T_0 = P^{-1} \begin{pmatrix} a_0 & b_0 \\ c_0 & d_0 \end{pmatrix} P. \quad (11)$$

The condition of decreasing magnitude away from the defect may then be stated as $d_0(\omega, \alpha) = 0$. The roots (ω_l, α_l) of this dispersion relation define bounded modes of the structures. As these modes decrease exponentially away from the defect, they are localized near it and have a finite energy. A numerical example is given in the next section, when we consider the diffraction of a wave by a finite one-dimensional photonic crystal.

2.4 Scattering photonic crystal of finite thickness

Of course, finite-sized photonic crystals are to be used in scattering experiments, and consequently only finite-size devices are to be considered. By this, we mean that there should be two vacuous half-spaces in any realistic problem. Let the wave vector of the incident plane wave be denoted by $\mathbf{k} = k_0(\cos\theta_{inc}\mathbf{e}_x + \sin\theta_{inc}\mathbf{e}_y)$ as shown in Figure 6.

The total field is described by a single variable function $U_N(x)$. When the electric field is parallel to the z axis (E// case), $U_N(x) \exp(iky \sin\theta)$ represents the z -component of the electric field; and when the magnetic field is parallel to the z axis (H// case), it represents the z -component of the magnetic field. With $\beta(x) = (\omega/c)[\varepsilon(x) - \sin^2(\theta)]^{1/2}$, where c is the speed of light in vacuum, we see

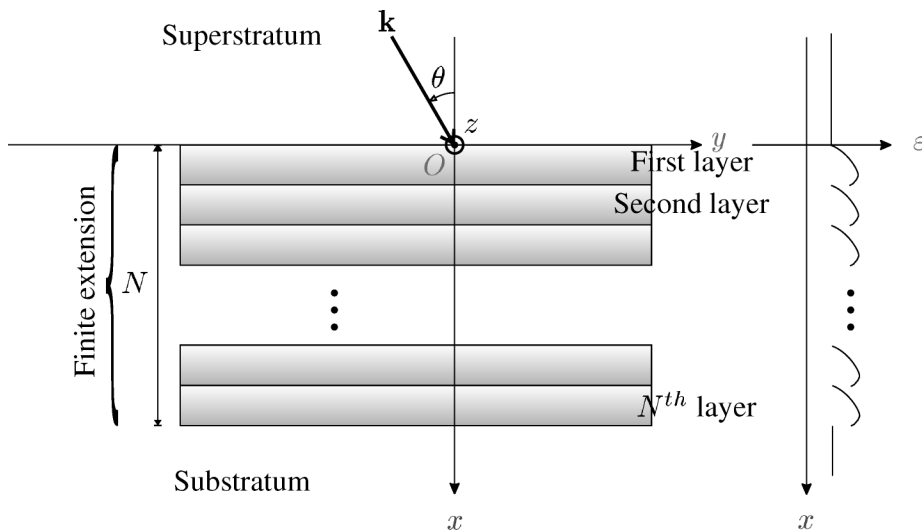


Figure 6 Schematic for the diffraction problem: a given incident field characterized by a vector of propagation $\mathbf{k}$ illuminates a structure made up of N identical layers.

that the total field U_N satisfies the following equation:

$$[p(x)U'_N]' + p(x)\beta^2(x)U_N = 0. \quad (12)$$

This equation and the radiation conditions lead to the following equations:

$$U_N(x) = \begin{cases} \exp(i\beta_0 x) + r_N(k, \theta) \exp(-i\beta_0 x), & x \leq 0, \\ t_N(k, \theta) \exp[i\beta_0(x - N)], & x \geq N. \end{cases} \quad (13)$$

Finally, from (5), we derive the relation

$$T^N \begin{bmatrix} 1 + r_N \\ i\beta_0(1 - r_N) \end{bmatrix} = t_N \begin{pmatrix} 1 \\ i\beta_0 \end{pmatrix}, \quad (14)$$

which permits us to obtain values of the reflection coefficient r_N and the transmission coefficient t_N .

2.4.1 Explicit expressions of the reflection and transmission coefficients

For the sake of clarity, we first introduce the four sets

$$\tilde{\mathcal{B}} = \left\{ (k, \theta) \in \mathbb{R}_+ \times \left(-\frac{\pi}{2}, \frac{\pi}{2} \right) \middle/ (\omega, \alpha) \in \mathcal{B} \right\}, \quad (15)$$

$$\tilde{\mathcal{G}} = \left\{ (k, \theta) \in \mathbb{R}_+ \times \left(-\frac{\pi}{2}, \frac{\pi}{2} \right) \middle/ (\omega, \alpha) \in \mathcal{G} \right\}, \quad (16)$$

$$\tilde{\Delta} = \left\{ (k, \theta) \in \mathbb{R}_+ \times \left(-\frac{\pi}{2}, \frac{\pi}{2} \right) \middle/ (\omega, \alpha) \in \Delta \right\}, \quad (17)$$

and

$$\tilde{\mathcal{P}} = \mathbb{R}_+ \times \left(-\frac{\pi}{2}, \frac{\pi}{2} \right) \setminus \tilde{\Delta} = \tilde{\mathcal{B}} \cup \tilde{\mathcal{G}}. \quad (18)$$

We suppose that (k, θ) belongs to $\tilde{\mathcal{P}}$. Denoting by $(\mathbf{v}, \mathbf{w})$ the basis of eigenvectors of T , we write in the canonical basis of $\mathbb{C}^2$: $\mathbf{v} = (v_1, v_2)$, $\mathbf{w} = (w_1, w_2)$. Eigenvector $\mathbf{v}$ (resp. $\mathbf{w}$) is associated to eigenvalue $\gamma(k, \theta)$ [resp. $\gamma^{-1}(k, \theta)$]. It is of course always possible to choose $(\mathbf{v}, \mathbf{w})$ such that $\det(\mathbf{v}, \mathbf{w}) = v_1 w_2 - v_2 w_1 = 1$.

After tedious but easy algebraic manipulations, we get the coefficients

$$r_N(k, \theta) = \frac{(\gamma^{2N} - 1)f}{\gamma^{2N} - g^{-1}f} \quad (19)$$

and

$$t_N(k, \theta) = \frac{(1 - g^{-1}f)\gamma^N}{\gamma^{2N} - g^{-1}f}, \quad (20)$$

where f and g are functions derived from $\mathbf{v}$ and $\mathbf{w}$ in the following manner. Denoting $q(x_1, x_2) = (i\beta_0 x_1 - x_2)/(i\beta_0 x_1 + x_2)$, we define the functions f and g as follows:

$$f(k, \theta) = \begin{cases} q(\mathbf{v}) & \text{if } |q(\mathbf{v})| \geq |q(\mathbf{w})| \\ q(\mathbf{w}) & \text{if } |q(\mathbf{w})| \geq |q(\mathbf{v})| \end{cases} \quad (21)$$

and

$$g(k, \theta) = \begin{cases} q(\mathbf{v}) & \text{if } |q(\mathbf{v})| \leq |q(\mathbf{w})| \\ q(\mathbf{w}) & \text{if } |q(\mathbf{w})| \leq |q(\mathbf{v})|. \end{cases} \quad (22)$$

Remark 2.2 As $\mathbf{v}$ is associated with the eigenvalue γ , we have $g(k, \theta) = q(\mathbf{v})$ and $f(k, \theta) = q(\mathbf{w})$ for any (k, θ) in $\tilde{\mathcal{G}}$ when $|\gamma| < 1$.

2.4.2 Reflection by a finite one-dimensional photonic crystal

We consider a one-dimensional structure made of N identical layers. A layer is characterized by its relative permittivity $\varepsilon(x)$, which is assumed to be real and positive. For convenience, the thickness of a layer is equal to 1 with an arbitrary unit. The structure is illuminated by a wavepacket of the following form:

$$U^i(x, y, t) = \int_{\tilde{\mathcal{P}}} A(k, \theta) \exp[ik(\cos \theta x + \sin \theta y)] \exp(-i\omega t) dk d\theta. \quad (23)$$

By means of Fourier analysis, the diffraction of such a wavepacket by a finite structure may be reduced to the study of the diffraction of a plane wave, which we developed in Section 2.4.1. The reflected and transmitted fields are expressed as

$$U_N^r(x, y, t) = \int_{\tilde{\mathcal{P}}} A(k, \theta) r_N \exp(i\mathbf{k}^r \cdot \mathbf{r}) \exp(-i\omega t) dk d\theta, \quad x < 0, \quad (24)$$

$$U_N^t(x, y, t) = \int_{\tilde{\mathcal{P}}} A(k, \theta) t_N \exp(i\mathbf{k}^t \cdot \mathbf{r}) \exp(-i\omega t) dk d\theta, \quad x > N, \quad (25)$$

with

$$\mathbf{k}^r = k(-\cos \theta \hat{\mathbf{x}} + \sin \theta \hat{\mathbf{y}}), \quad \mathbf{k}^t = k(\cos \theta \hat{\mathbf{x}} + \sin \theta \hat{\mathbf{y}}). \quad (26)$$

Our aim is now to study the limit behavior of $r_N(k, \theta)$ and $t_N(k, \theta)$ as N tends to infinity, and to obtain explicit asymptotics of the fields.

For this purpose, we split the integrals into two terms: an integral over $\tilde{\mathcal{G}}$ [namely, $U_N^{r, \tilde{\mathcal{G}}}(x, y, t)$ and $U_N^{t, \tilde{\mathcal{G}}}(x, y, t)$], and an integral over $\tilde{\mathcal{B}}$ [namely, $U_N^{r, \tilde{\mathcal{B}}}(x, y, t)$ and $U_N^{t, \tilde{\mathcal{B}}}(x, y, t)$]. For any $(k, \theta) \in \tilde{\mathcal{G}}$, we have $|\gamma| < 1$; and therefore $r_N(k, \theta) \rightarrow g(k, \theta)$, whereas $t_N(k, \theta) \rightarrow 0$ as $N \rightarrow +\infty$. Consequently, we have

$$\lim_{N \rightarrow \infty} U_N^{r, \tilde{\mathcal{G}}}(x, y, t) = \int_{\tilde{\mathcal{G}}} A(k, \theta) g(k, \theta) \exp(i\mathbf{k}^r \cdot \mathbf{r}) \exp(-i\omega t) dk d\theta \quad (27)$$

and

$$\lim_{N \rightarrow \infty} U_N^{t, \tilde{G}}(x, y, t) = 0. \quad (28)$$

The integral over $\tilde{B}$ is more complicated, because r_N and t_N do not converge pointwise.¹ In order to evaluate these integrals, it is convenient to make the following remark:

Remark 2.3 We chose f and g in such a way that $|g| < |f|$; and in a conduction band $f = \bar{g}^{-1}$ (where $\bar{g}$ denotes the conjugate of the complex number g), so that we always have $|g(k, \theta)| \leq 1$.

Writing $\tau = gf^{-1}$ and having in mind that $|\tau| < 1$ in a conduction band, by virtue of (19) and (20) we find

$$r_N(k, \theta) = g + g \sum_{l=1}^{+\infty} \gamma^{2Nl} (\tau^l - \tau^{l-1}), \quad (29)$$

$$t_N(k, \theta) = (1 - \tau) \gamma^N \sum_{l=0}^{+\infty} \gamma^{2Nl} \tau^l. \quad (30)$$

Accordingly, we write

$$\begin{aligned} U_N^{r, \tilde{B}}(x, y, t) &= \int_{\tilde{B}} A(k, \theta) g(k, \theta) \exp i(\mathbf{k} \cdot \mathbf{r} - \omega t) dk d\theta \\ &\quad + \sum_{l=1}^{+\infty} U_{l, N}^{r, \tilde{B}}(x, y, t) \\ U_N^{t, \tilde{B}}(x, y, t) &= \sum_{l=0}^{+\infty} U_{l, N}^{t, \tilde{B}}(x, y, t), \end{aligned} \quad (31)$$

where

$$U_{l, N}^{r, \tilde{B}}(x, y, t) = \int_{\tilde{B}} A(k, \theta) \gamma^{2Nl} g(\tau^l - \tau^{l-1}) \exp [i(\mathbf{k} \cdot \mathbf{r} - \omega t)] dk d\theta \quad (32)$$

for $l \geq 1$, and

$$U_{l, N}^{t, \tilde{B}}(x, y, t) = \int_{\tilde{B}} A(k, \theta) \gamma^{(2l+1)N} (1 - \tau) \tau^l \exp [i(\mathbf{k} \cdot \mathbf{r} - \omega t)] dk d\theta \quad (33)$$

for $l \geq 0$.

¹Actually, they converge weakly!

For sufficiently large N , these last functions correspond to classical optical rays, and it can be proved that $U_{l,N}^{t,\tilde{\mathcal{B}}}(x, y, t)$ [resp. $U_{l,N}^{r,\tilde{\mathcal{B}}}(x, y, t)$] converges uniformly towards zero for every $l \geq 0$ (resp. $l \geq 1$). As a consequence, (27) and (28) still hold when $\tilde{\mathcal{G}}$ is replaced by $\tilde{\mathcal{B}}$; thus,

$$\lim_{N \rightarrow +\infty} U_N^{r,\tilde{\mathcal{B}}}(x, y, t) = \int_{\tilde{\mathcal{B}}} A(k, \theta) g(k, \theta) \exp[i(\mathbf{k}^r \cdot \mathbf{r} - \omega t)] dk d\theta \quad (34)$$

and

$$\lim_{N \rightarrow +\infty} U_N^{t,\tilde{\mathcal{B}}}(x, y, t) = 0. \quad (35)$$

Finally, these relations always hold for the total field. As a conclusion, for sufficiently large periodic structures, the reflected wave is entirely characterized by the function g .

3 Two-dimensional photonic crystals

In the preceding section, we developed in great detail the scattering theory of one-dimensional photonic crystals, not only because they were historically the first ones to be used and still are, but also because the theory takes an explicit form that permits one to describe easily their fundamental properties: the existence of photonic bandgaps and the possibility of localizing light. However, because they are periodic in only one spatial direction, their bandgaps are very sensitive to the angle of incidence. Light cannot therefore be localized in more than one direction.

The essence of photonic crystals, which was described in the seminal articles of Yablonovitch and John, is to extend the periodicity to more spatial dimensions. Examples of such structures are given in Figure 7. We should stress here that what seems to be only an evident generalization has waited a very long time from the original idea of Bragg mirrors to be conceived. With this extension, we enter a new realm of technology, in which total control over the propagation of light could be

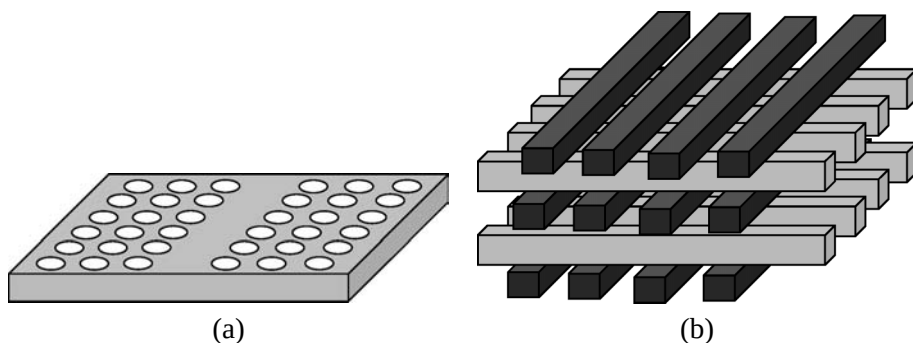


Figure 7 Schematics of higher-dimensional structures: (a) 2D photonic crystal, (b) 3D photonic crystal.

possible. Research activity on these structures has reached quite an incredible level, with more than a thousand papers published each year. Many possible applications or effects have been predicted. Although most of them are still hanging in the air, the rapid development of technology should make them a reality in the near future [12]–[17].

In the rest of this chapter, we describe some of these applications and their physical bases. To do so, we have to present the theoretical aspects of two-dimensional and three-dimensional photonic crystals. The developments of the second section should make that easier.

The first approach to higher-dimensional structures is the spectral approach, by which we mean the characterization of the free modes of an infinitely large crystal. This amounts to extending the method of Bloch waves to two-dimensional and three-dimensional structures, just the way it is done for the Schrödinger equation and the band theory for electrons in natural crystals. Let us therefore recast the time-harmonic Maxwell curl equations as

$$\partial_t \begin{pmatrix} \mathbf{E} \\ \mathbf{H} \end{pmatrix} = \begin{pmatrix} 0 & (\varepsilon_0 \varepsilon)^{-1} \nabla \times \\ \mu_0^{-1} \nabla \times & 0 \end{pmatrix} \begin{pmatrix} \mathbf{E} \\ \mathbf{H} \end{pmatrix}, \quad (36)$$

where ε is a periodic function and the medium is nonmagnetic. This way, the Maxwell curl equations appear in formal Hamiltonian form to which the method of Bloch waves can be efficiently applied. We do not develop this theory here as we did for the one-dimensional case, but rather specialize it to two-dimensional photonic crystals for a specific type of incident polarization.

Let us consider a two-dimensional photonic crystal that is a periodic arrangement of infinitely long, identical, parallel dielectric rods with relative permittivity ε_r . We choose a square periodicity, the unit cell being the square $Y = [0, 1]^2$, and we search for z -independent fields. Let us concentrate on the $H//$ case, where $\mathbf{H} = u(x, y)e^{-i\omega t}\mathbf{e}_z$. Then the scalar $u(x, y)$ satisfies the relation

$$-\nabla \cdot (\varepsilon^{-1} \nabla u) = \left(\frac{\omega}{c} \right)^2 u. \quad (37)$$

Following the method of Bloch waves, the solutions being quasi-periodic ones, we define

$$u_k(\mathbf{r}) = e^{i\mathbf{k} \cdot \mathbf{r}} v(\mathbf{r}), \quad (38)$$

where $\mathbf{r} \equiv (x, y)$ and $\mathbf{k}$ belongs to the first Brillouin zone $Y' = [-\pi, \pi]^2$. The point is now to find the frequency domain where such solutions exist. Just as what happens in one-dimensional structures, there are bands of frequency where no such solutions exist. A complete exposition is given elsewhere [18].

3.1 Optical characterization of photonic crystals

Although the method of Bloch waves can only solve spectral problems (i.e., yield free modes), we are interested in determining the optical characteristics of photonic crystals. By this we mean that, given a finite-sized device, we want to be able to compute the field that is scattered by the device once it is illuminated by some incident field. To do so, various methods can be used. One possible way consists of modeling a photonic crystal as a stack of gratings, for which very efficient numerical methods have been developed [7], [19]. We can also solve for the temporal behavior of the field and then get the harmonic response by a Fourier transform, which is the so-called finite-difference-time-domain (FDTD) method [20]. Finally, and this shall be our approach, one can compute the scattering matrix of the device; that is, the operator S that relates the incident field u^{inc} to the scattered field u^{sca} . This method is particularly well suited to the study of finite-sized devices. It is closely allied to the FDTD method, because the scattering matrix can be obtained from the computations in the time domain [21].

3.2 Construction of the scattering matrix

Let us now describe the scattering theory of a collection of N scatterers that are invariant along the z axis [22-23]. The rods are contained in a domain Ω whose relative permittivity is ε_r . The j th rod has an arbitrary cross-section denoted by C_j in Figure 8, and rod is located at $\mathbf{r}^j = (r^j, \theta^j)$. A local polar coordinate system

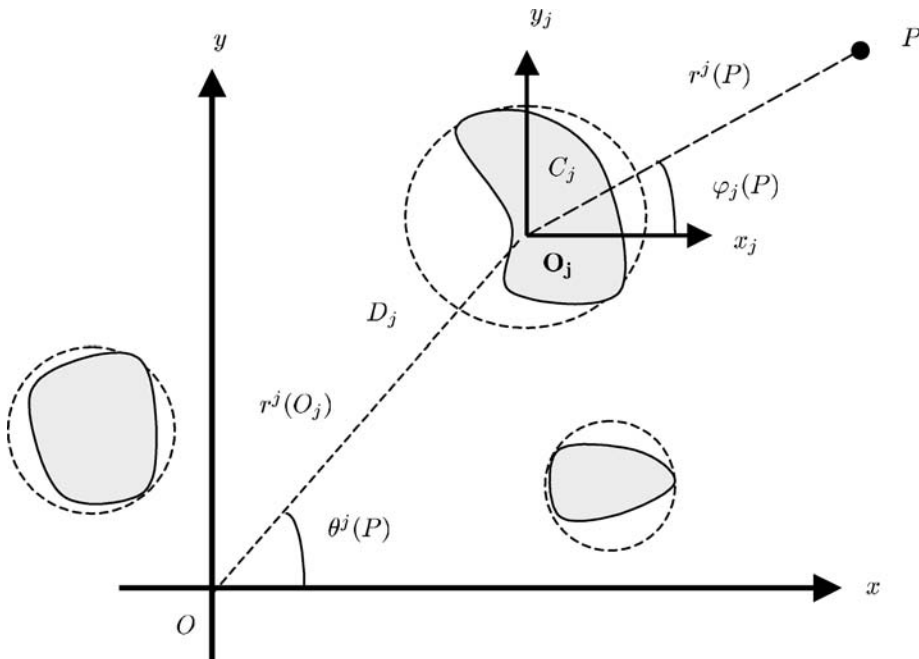


Figure 8 Notation for multiple scattering theory.

(r_j, φ_j) is defined for this rod j . Its relative permittivity is denoted by ε_j . Domain C_j is contained in a disk D_j of radius d_j , and we assume that $D_j \cap D_l = \emptyset$. The permittivity of the ambient medium is supposed to be the permittivity *in vacuo* $\varepsilon_{out} = \varepsilon_0$.

We start by considering only one rod. The rod is illuminated by an incident monochromatic field u^{inc} which is s -polarized. Let u denote the total electric field. We define the diffracted field u^{sca} as $u^{sca} = u - u^{inc}$. The total field satisfies the Helmholtz equation

$$\nabla^2 u + k_0^2 \varepsilon u = 0, \quad (39)$$

while the incident field satisfies $\nabla^2 u^{inc} + k_0^2 u^{inc} = 0$. Keeping in mind that u^{inc} is bounded in all space, we deduce that u^{inc} can be written as the Fourier series

$$u^{inc}(\mathbf{r}_1) = \sum_n a_n J_n(k_0 r_1) e^{in\varphi_1}, \quad (40)$$

where $J_n(\cdot)$ is the cylindrical Bessel function of order n . Furthermore, the diffracted field satisfies the relation

$$\nabla^2 u^{sca} + k_0^2 u^{sca} = k_0^2 (\varepsilon_{out} - \varepsilon) u, \quad (41)$$

everywhere. We can transform this partial differential equation into an integral equation by using the Green function of the Helmholtz equation, that is the function $G_0(\mathbf{r}, \mathbf{r}')$ such that

$$\begin{cases} \nabla^2 G_0 + k_0^2 G_0 = \delta(\mathbf{r}_1 - \mathbf{r}'_1), \\ G_0 \text{ satisfies an outgoing wave condition.} \end{cases} \quad (42)$$

For the two-dimensional case,

$$G_0(\mathbf{r}_1, \mathbf{r}'_1) = -\frac{i}{4} H_0^{(1)}(k_0 |\mathbf{r}_1 - \mathbf{r}'_1|), \quad (43)$$

where $H_0^{(1)}$ is the Hankel function of the first kind [24]. We can then recast (41) as

$$u = u^{inc} + k_0^2 \int_{C_1} (\varepsilon_{out} - \varepsilon_1) G_0(\mathbf{r}_1, \mathbf{r}'_1) u(\mathbf{r}'_1) d\mathbf{r}'_1, \quad (44)$$

everywhere.

At this point, we transform this integral equation into a linear system of equations. To do that, we note from Graf's formula [24]-[26] that

$$H_0^{(1)}(k_0 |\mathbf{r}_1 - \mathbf{r}'_1|) = \sum_n J_n(k_0 r'_1) e^{-in\varphi'_1} H_n^{(1)}(k_0 r_1) e^{in\varphi_1}, \quad \text{for } |\mathbf{r}_1| > |\mathbf{r}'_1|. \quad (45)$$

Equation (44) can then be written as

$$u(\mathbf{r}_1) = \sum_n a_n^{(1)} J_n(k_0 r_1) e^{in\varphi_1} + \sum_n b_n^{(1)} H_n^{(1)}(k_0 r_1) e^{in\varphi_1}, \quad |\mathbf{r}_1| \geq d, \quad (46)$$

where

$$b_n^{(1)} = \frac{ik_0^2}{4} \int_{C_1} (\varepsilon_{out} - \varepsilon_1) J_n(k_0 r_1) e^{-in\varphi_1} d\mathbf{r}_1. \quad (47)$$

Given the conditions at the boundary of the rod, there exists a linear operator $S(k_0)$, the scattering matrix, such that [27]

$$\widehat{b} = S(k_0)\widehat{a}, \quad (48)$$

where $\widehat{b} = (b_n)_n$ and $\widehat{a} = (a_n)_n$. For circular rods, it is possible to find an explicit expression for S .

Let us now describe the multiple-scattering theory for the simplest case of two rods, each rod being characterized by a scattering matrix S_l , $l \in \{1, 2\}$, let us consider the rod numbered l . Around this rod, the total exterior field satisfies (41), and the integral equation

$$\begin{aligned} u(\mathbf{r}_l) &= u^{inc} + k_0^2 \int_{C_1} (\varepsilon_{out} - \varepsilon_1) G_0(\mathbf{r}_l, \mathbf{r}'_l) u(\mathbf{r}'_l) d\mathbf{r}'_l \\ &\quad + k_0^2 \int_{C_2} (\varepsilon_{out} - \varepsilon_2) G_0(\mathbf{r}_l, \mathbf{r}'_l) u(\mathbf{r}'_l) d\mathbf{r}'_l \\ &= u^{inc} + \sum_n b_n^l H_n^{(l)}(k_0 r_l) e^{in\varphi_l} + \sum_n b_n^{l,p} J_n(k_0 r_l) e^{in\varphi_l} \end{aligned} \quad (49)$$

emerges with $p \neq l$, $p = 1$ or 2 . Around a particular rod, the total field is written as the sum of a scattered field (the field that is locally scattered by the rod), and an incoming field (which accounts for the incident field and the field scattered by the other rods). The field scattered by rod l can be expressed in the polar coordinates system attached to rod p , by using Graf's formula, as

$$H_n^{(1)}(kr_l) e^{in\varphi_l} = \sum_m J_{m-n}(k_0 r_p^l) e^{-im\varphi_p^l} H_m^{(1)}(k_0 r_p) e^{im\varphi_p}; \quad l, p \in \{1, 2\}, \quad l \neq p, \quad (50)$$

where (r_p^l, φ_p^l) are the polar coordinates of rod l in the coordinate system of rod p . We obtain the linear relation in matrix form as $\widehat{b}^l = T_{pl} \widehat{b}^p$, where T_{pl} is defined as

$$(T_{pl})_{mq} = \exp[i(q-m)\theta_l^j] H_{m-q}^{(1)}(kr_l^j). \quad (51)$$

The last point consists in expanding the incident field in the polar coordinates systems of each rod. Graf's formula allows us to obtain the Fourier coefficients

$(\hat{a}_1, \hat{a}_2)$ linearly from that in the reference frame: $\hat{a}_l = Q_l \hat{a}$. We finally obtain the following two equations:

$$\begin{aligned}\hat{b}^1 &= S(k_0)(Q_1 \hat{a} + T_{12} \hat{b}^2), \\ \hat{b}^2 &= S(k_0)(Q_2 \hat{a} + T_{21} \hat{b}^1).\end{aligned}\quad (52)$$

This system can be rewritten as follows:

$$\begin{bmatrix} I & -S(k_0)T_{12} \\ -S(k_0)T_{21} & I \end{bmatrix} \begin{pmatrix} \hat{b}^1 \\ \hat{b}^2 \end{pmatrix} = \begin{bmatrix} S(k_0)Q_1 \hat{a} \\ S(k_0)Q_2 \hat{a} \end{bmatrix}. \quad (53)$$

From this relation, we deduce the existence of a linear operator S , the scattering matrix of two nonintersecting rods, such that

$$\begin{pmatrix} \hat{b}^1 \\ \hat{b}^2 \end{pmatrix} = S \begin{pmatrix} Q_1 \hat{a} \\ Q_2 \hat{a} \end{pmatrix}. \quad (54)$$

This method can be generalized to an arbitrary number N of scatterers; it can also be extended to nonlinearly polarized waves [28].

We are now in position to compute the response of a finite-sized photonic crystal to an incident field. Let us consider, for instance, the crystal depicted in Figure 9(a). The rods are dielectric ($\epsilon_r = 11.56$) and the incident field is s -polarized, and the period along the x and y axes is denoted by d . The crystal is illuminated by an incident plane wave, as indicated by the arrow in Figure 9(a). The transmittance is computed as the flux of the real part of the complex Poynting vector through a segment situated below the crystal, normalized by the flux of the real part of the

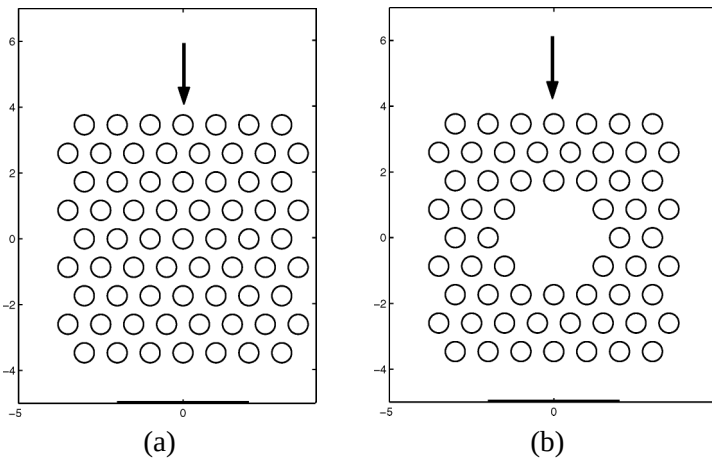


Figure 9 (a) Schematics of a perfect two-dimensional photonic crystal and (b) a photonic crystal with a cavity.

complex Poynting vector of the incident field. This is shown by the bold line in Figure 10(a). We see that the transmission is damped in the region $\lambda_0/d \in [3.25, 4.5]$. This region of wavelengths is a bandgap: the field is exponentially damped inside the crystal. The map of the electric field inside the crystal for $\lambda_0/d = 4$ in Figure 10(b) clearly shows that the field is almost completely reflected.

3.3 Gamow vectors and quasi-normal modes

When dealing with infinitely extended structures, interest is focused on the eigenfrequencies; that is, the frequencies at which propagating modes exist in the structure. For finite-sized structures, there are no longer such real-valued frequencies: the modes are characterized by complex-valued resonant frequencies, which are the poles of the scattering matrix [29]. From theoretical as well as numerical points of view, an important result is that the scattering matrix can be expanded in generalized Laurent series as

$$S(k) = \sum_p \frac{R_p}{k^2 - k_p^2} + S_0(k). \quad (55)$$

The residue operators R_p are obtained through Cauchy integrals

$$S(k) = 1/(2i\pi) \int_{\gamma} S(z) dz, \quad (56)$$

where γ is a loop enclosing the only pole k_p . The absence of poles of order higher than 1 is due to the fact that $\varepsilon - 1$ is null-valued outside a bounded domain. From general scattering theory it is known that the scattering operator is of finite rank. We can now state the following:

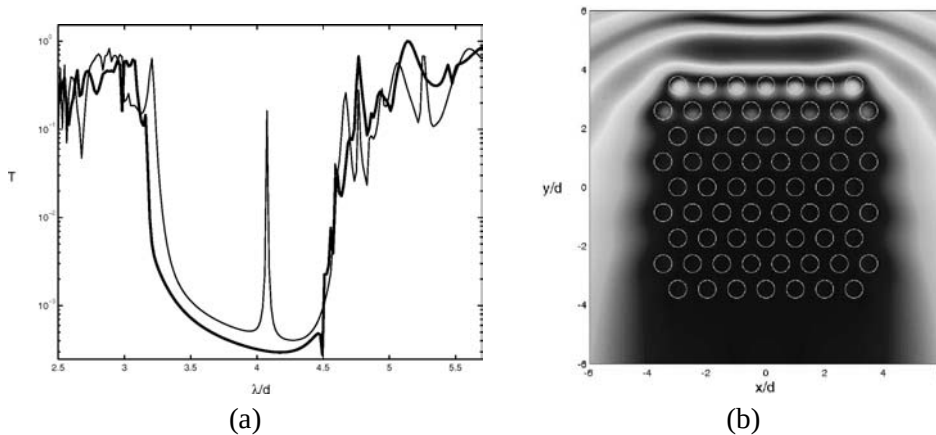


Figure 10 (a) Transmittance of the structures in Figure 3: the bold line corresponds to the perfect crystal and the thin one to the crystal with a cavity. (b) Map of the electric field inside the crystal for $\lambda_0/d = 4$.

Definition 3.1 The Gamow vectors $|\psi_p\rangle$ associated with the resonance k_p are the eigenvectors of R_p .

Let us assume, for simplicity, that the rank of R_p is unity, and let us denote a normalized Gamow vector by $|\phi_p\rangle$. For a given incident field u^{inc} , formally denoted as $|u^{inc}\rangle$, and a wavenumber very close to k_p , the scattered field is $|u^{sca}\rangle = S(k)|u^{inc}\rangle = \frac{R_p|u^{inc}\rangle}{k^2 - k_p^2}$. Now we want to evaluate the vector $R_p|u^{inc}\rangle$, which is proportional to $|\phi_p\rangle$: $R_p|u^{inc}\rangle = \mu|\phi_p\rangle$. We get $\langle R_p|u^{inc}|\phi_p\rangle = \bar{\mu}$, assuming that $|\phi_p\rangle$ is normalized as per $\langle\phi_p|\phi_p\rangle = 1$. This suggests to introduce the adjoint state of $|\phi_p\rangle$ as $|\phi_p^*\rangle = R_p^*|\phi_p\rangle$, where R_p^* is the adjoint operator of R_p . Finally, we get $R_p|u^{inc}\rangle = \langle\phi_p^*|u^{inc}\rangle|\phi_p\rangle$ so that we can write

$$R_p = |\phi_p\rangle\langle\phi_p^*|. \quad (57)$$

We can also note the relation $R_p^2 = \mu R_p$. This gives us an expression of the S matrix in terms of its resonances and Gamow vectors:

$$S(k) = \sum_p \frac{|\phi_p\rangle\langle\phi_p^*|}{k^2 - k_p^2} + S_0(k). \quad (58)$$

4 Resonant modes

Let us now explore the possibility of localizing light inside higher-dimensional photonic crystals, just as we did for one-dimensional structures.

4.1 Isolated defect

We explicate here the basics of the characterization of defect modes in photonic crystals by taking an academic example.² We consider first the case of a single defect, obtained for instance by removing a few rods inside the crystal; see Figure 9(b). In doing so, we create a cavity inside the crystal. How is the optical behavior of the crystal affected by this defect? To answer this question, let us plot the transmission through the crystal with a defect: it is the thin line in Figure 10(a). A sharp peak is located inside the bandgap near $\lambda_0/d \sim 4.07$. This peak corresponds to a pole of the scattering matrix at $\lambda_0/d = 4.077 + 0.0016i$.

This resonance is an intrinsic property of the finite crystal and not a phenomenon linked to the use of a particular incident field or way of computing the transmission ratio. The mode, or Gamow vector, associated with this pole is computed as explained in Section 3.3 using Cauchy integrals. The maps of the modal electric field and the Poynting vector are plotted in Figure 11(a–b). This mode is the fundamental mode of the cavity.

²A more comprehensive treatment may be found elsewhere [29]–[33].

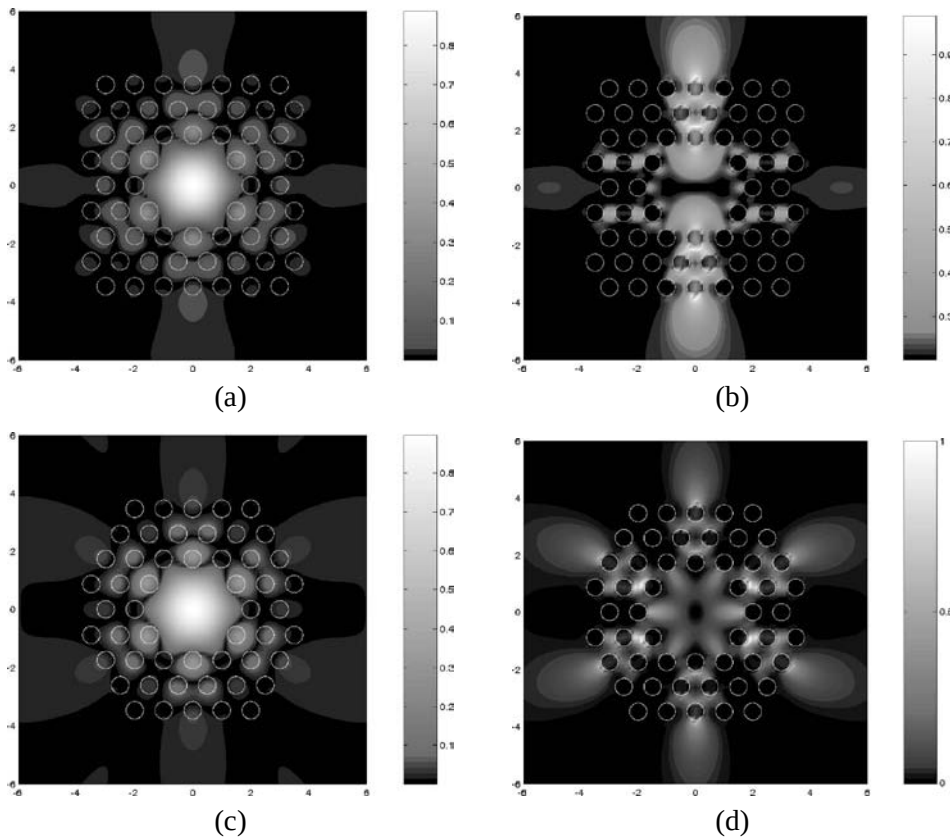


Figure 11 Fundamental mode of the cavity: map of the electric field (left) and norm of Poynting vector (right): (a) and (b) square boundary, $\lambda_0/d = 4.077 + 0.0016i$; (c) and (d) hexagonal boundary, $\lambda_0/d = 4.079 + 0.0036i$.

Let us now consider the same lattice and cavity but with a hexagonal boundary. As the resonant wavelength is $\lambda_0/d = 4.079 + 0.0036i$, the real part is only slightly modified while the imaginary part increases, which is due to the fact that the losses by radiation increase. The maps of the electric field and Poynting vector are given in Figure 11(c–d). A simple comparison with Figure 11(a) shows that the symmetry of the electromagnetic energy of the mode can be greatly increased or decreased by simply modifying the shape of the overall structure, which is of particular interest for spatially selective devices [34–35].

4.2 Photonic waveguides

Extended defect modes can be realized by creating, for instance, a line of defects—as shown in Figure 12(a) [36–37]. That way, a photonic crystal waveguide is created. The underlying physical situation is that of a set of coupled oscillators. If we assume, for instance, that we create two cavities inside the crystal, then the system behaves as two coupled oscillators and beats, the so-called Rabi oscillations, can

be seen between both cavities [33]. When an extended set of defects is created, a band of defect modes appears, and it is possible to guide light inside the crystal by means of the photonic bandgap effect. An example is given in Figure 12(b), where we have plotted the map of the electric field of an extended defect mode ($\lambda_0/d = 3.871 + 0.017i$).

The described physical behavior allows one to design much more complicated devices wherein the propagation of light can be controlled. For instance, it is possible to design add-drop filters, where the propagation of light depends on the wavelength. Figure 13 shows the maps of the electric fields inside an add-drop filter for two different wavelengths. For these different wavelengths, light propagates either forward from the upper branch to the lower branch, or backward. Another example is provided in Figure 14, where light can turn at a right angle and propagate almost without loss [38].

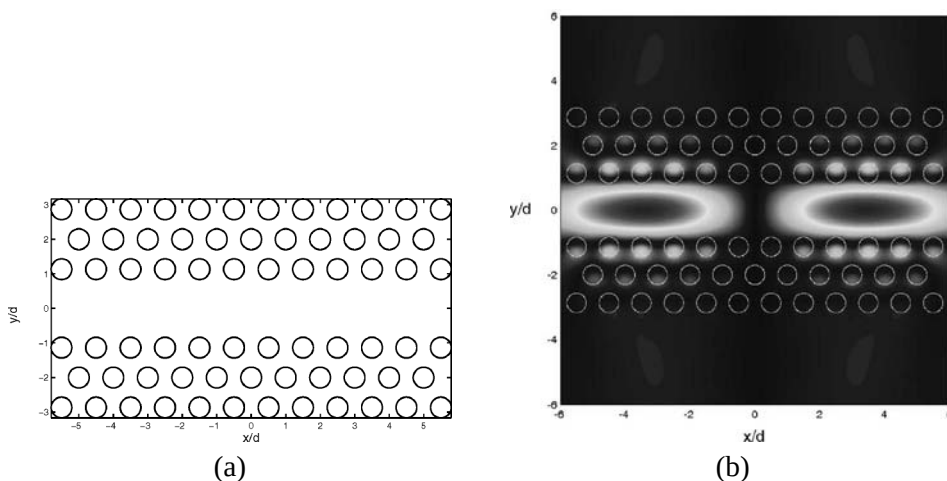


Figure 12 (a) A photonic crystal waveguide and (b) the map of the electric field of a propagating mode.

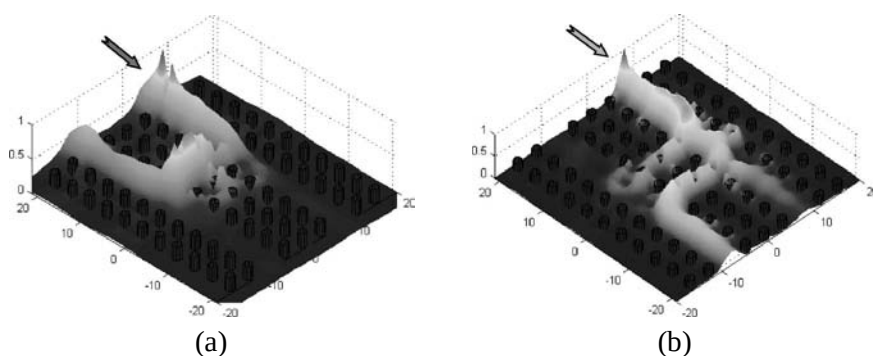


Figure 13 Map of the electric field for two modes inside an add-drop filter. (a) Backward branching, (b) forward branching.

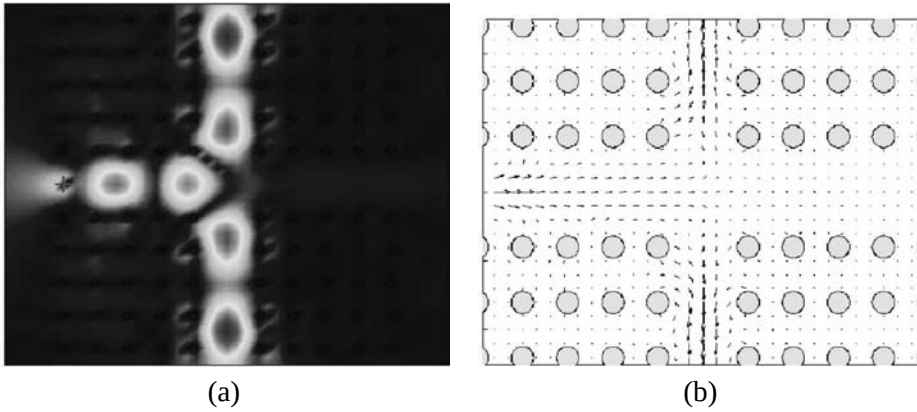


Figure 14 Light propagation in a cross-waveguide. At the intersection, light turns at right angle without reflection. (a) Map of the electric field, (b) map of the Poynting vector.

5 Current problems and future directions

There is currently strong activity in designing photonic crystal waveguides [39]–[41]. The theoretical approach that we have proposed in Section 3 can work only with two-dimensional structures. Of course, actual structures are made of holes in slabs of finite thickness [Figure 7(a)]. Though the physics is basically the same, real structures cannot confine light in the direction of the holes, because there are some losses due to radiation into free space. One then must be able to characterize these losses. A possible method relies on a mixed approach combining scattering theory and grating theory. We give only a brief sketch of this approach here.

The mixed approach amounts to periodically extending the structure in the direction perpendicular to the waveguide. Then we can use a Bloch wave expansion in the horizontal xy plane. Choosing, for instance, the magnetic field, we write

$$H(\mathbf{k}, \mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}} \sum_{nm} h_{nm}(z) e^{ik_x n x + i k_y m y}. \quad (59)$$

Next, we construct the scattering matrix of this *supergrating*, i.e., the matrix relating the incoming plane waves to the outgoing plane waves. Above (superscript u) and below (superscript b) the device, the magnetic field can be written as

$$\begin{aligned} H^u &= e^{i\mathbf{k}\cdot\mathbf{r}} \sum h_{nm}^{u,+} e^{i\alpha_{nm} z} + e^{i\mathbf{k}\cdot\mathbf{r}} \sum h_{nm}^{u,-} e^{-i\alpha_{nm} z} \\ H^b &= e^{i\mathbf{k}\cdot\mathbf{r}} \sum h_{nm}^{b,+} e^{i\alpha_{nm} z} + e^{i\mathbf{k}\cdot\mathbf{r}} \sum h_{nm}^{b,-} e^{-i\alpha_{nm} z} \end{aligned} \quad (60)$$

and the S matrix relates the column vector $(\widehat{h_{nm}^{u,-}}, \widehat{h_{nm}^{b,-}})$ to $(\widehat{h_{nm}^{u,+}}, \widehat{h_{nm}^{b,+}})$. This matrix depends upon both λ_0 and $\mathbf{k}$. By computing the resonances of the S matrix (just

as we did for two-dimensional structures), it is possible to compute the complex-valued propagation constants and then to characterize the losses due to the coupling with the free space modes.

Up to this point, we considered passive structures only. For active devices, it is interesting to consider photonic crystals made of nonlinear materials. The two main effects of interests are (i) the Kerr effect (a third-order effect), that is, the χ_3 tensor; and (ii) the generation of second harmonic, i.e., second-order effects related to χ_2 . Very few results are known for the second-order effects. For the Kerr effect, and with some common approximations, the refractive index of the material depends on the intensity of the field as $n = n_0 + \chi_3 |E|^2$ and makes it possible to control the resonant wavelength. Indeed, if we create a cavity inside a photonic crystal, the resonant wavelength strongly depends on the refractive index of the material. So, by a small change in the refractive index, we can modify this value. An example of this situation is provided in Figure 15. It represents the transmission through the crystal at a given wavelength as the amplitude A^{inc} of the incident field is increased. The chosen wavelengths $\lambda_{1,2}$ are slightly below the resonant wavelength of the cavity for low-intensity fields. We see that at a certain point the field has modified the characteristics of the cavity, so that the resonant wavelength is $\lambda_{1,2}$. At this point, transmission becomes very high. We have thus created a hysteresis cycle, and we have obtained bistability. The interested reader will find a much more detailed study in Ref. 42 and some results on the second harmonic generation in a recent work [43].

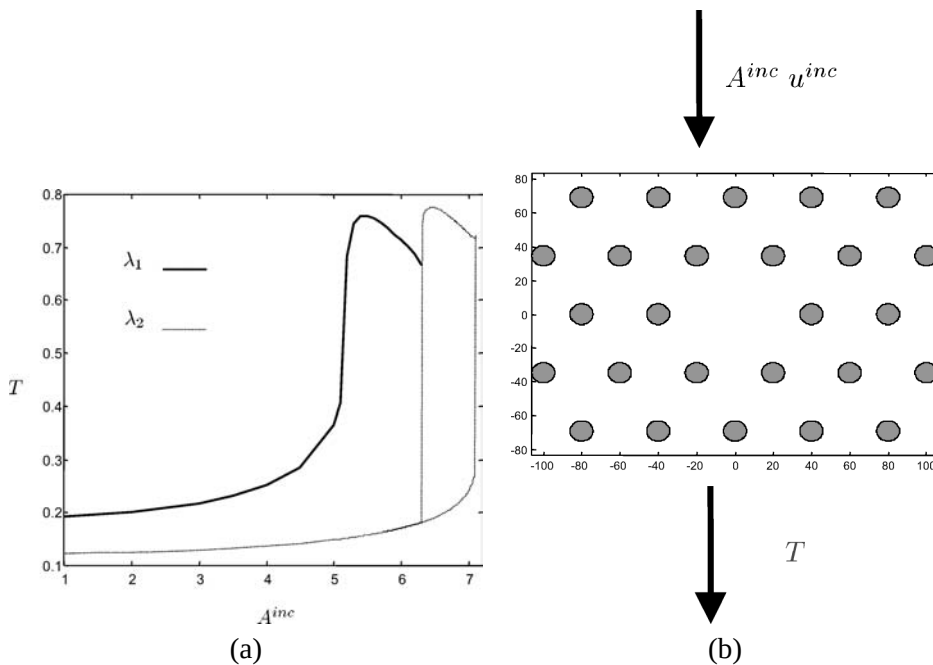


Figure 15 Bistability in a nonlinear two-dimensional photonic crystal. (a) Transmission versus the amplitude of the incident field, (b) schematic of the device.

6 Concluding remarks

Throughout this chapter, we have stressed the properties of photonic crystals considered as clusters of scatterers. Many theoretical tools have been developed in solid state physics, and it is very tempting to use this inheritance as much as possible. However, unlike conduction electrons that barely leave the crystal (scattering is not at stake here), photons do come from free space and go back to it after interacting with matter. The immediate consequence is that photons interact strongly with the boundary of the device, which renders the problem fascinating. Besides, the consideration of an infinitely extended device prevents us from taking into account the near field (i.e., evanescent waves) which is to play a very important role in future technological developments; it suffices to think, for instance, of surfaces plasmons-polaritons [8], [44] and their recently proven ability to preserve quantum entanglement [45]. Many issues, such as the development of microstructured optical fibers [46], have not been addressed. But we must point out the area of quantum optics, where the notion of residue operators could enable a theoretical description of open-cavity quantum electrodynamics that happens in finite-sized photonic crystals [47].

Acknowledgment

Many thanks are due to Dr. E. Centeno for interesting discussions and for providing some of the illustrations.

References

- [1] S. John, Electromagnetic absorption in a disordered medium near a photon mobility edge, *Phys Rev Lett* **53** (1994), 2169–2172.
- [2] E. Yablonovitch, Inhibited spontaneous emission in solid-state physics and electronics, *Phys Rev Lett* **58** (1987), 2059–2062.
- [3] J.D. Joannopoulos, R.D. Meade and J.N. Winn, *Photonic crystals: Molding the flow of light*, Princeton University Press, Princeton, NJ, USA, 1995.
- [4] A. Scherer, T. Doll, E. Yablonovitch, H.O. Everitt and J.A. Higgins (eds), Special section on: Electromagnetic crystal structures, design, synthesis, and applications (optical), *J Lightwave Technol* **17** (1999), 1931–2405.
- [5] A. Scherer, T. Doll, E. Yablonovitch, H.O. Everitt and J.A. Higgins (eds), Special section on: Electromagnetic crystal structures, design, synthesis, and applications (microwave), *IEEE Trans Microw Theory Tech* **47** (11) (1999).
- [6] J. Dowling, An exhaustive bibliography is kept up to date on <http://home.earthlink.net/~jpdowling/pbgbib.html>.
- [7] R. Petit (ed), *Electromagnetic theory of gratings*, Springer, Berlin, Germany, 1980.
- [8] S.I. Bozhevolnyi, J. Erland, K. Leosson, P.M.W. Skovgaard and J.M. Hvam, Waveguiding in surface plasmon polariton bandgap structures, *Phys Rev Lett* **86** (2001), 3008–3011.

- [9] R. Smaâli and D. Felbacq, Bloch waves and non-propagating modes in photonic crystals, *Physica E* **18** (2003), 443–451.
- [10] P.T. Leung, S.Y. Liu and K. Young, Completeness and orthogonality of quasinormal modes in leaky optical cavities, *Phys Rev A* **49** (1994), 3057–3067.
- [11] K.C. Ho, L. Pui-Tang and M. van den Brink, Second quantization of open systems using quasinormal modes, *Phys Rev E* **58** (1998), 2965–2978.
- [12] M. François, J. Danglot, B. Grimbert, P. Mounaix, M. Muller, O. Vanbésien and D. Lippens, Photonic bandgap material for integrated photonic application: technological challenges, *Microelectron Eng* **61-62** (2002), 2169–2172.
- [13] T. Charvolin, E. Hadji, E. Picard, M. Zelsman, M. Assous, B. Dalzotto, M.E. Nier, S. Tedesco, X. Letartre, P. Rojo-Roméo and C. Seassal, Fabrication and optical measurements of silicon on insulator photonic nanostructures, *Microelectron Eng* **61-62** (2002), 529–536.
- [14] K. Busch, Photonic band structure theory: assessment and perspectives, *Compte Rendu Phys Pans* **3** (2002), 53–66.
- [15] A. Femius Koenderink, J.F. Galisteo and W.L. Vos, Three-dimensional photonic crystals as a cage for light, *Compte Rendu Phys Pans* **3** (2002), 67–77.
- [16] J.M. Lourtioz and A. de Lustrac, Metallic photonic crystals, *Compte Rendu Phys Pans* **3** (2002), 79–88.
- [17] H. Benisty, M. Rattier and S. Olivier, Two-dimensional photonic crystals: new feasible confined optical systems, *Compte Rendu Phys Pans* **3** (2002), 89–102.
- [18] M. Plihal and A.A. Maradudin, Photonic band structure of two-dimensional systems: the triangular lattice, *Phys Rev B* **44** (1991), 8565–8571.
- [19] M. Nevière and E. Popov, *Light propagation in periodic media*, Marcel Dekker, New York, NY, USA, 2003.
- [20] M. Qiu and S. He, Numerical method for computing defect modes in two-dimensional photonic crystals with dielectric or metallic inclusions, *Phys Rev B* **61** (2000), 12871–12876.
- [21] N. Moiseyev, Quantum theory of resonances: calculating energies, widths and cross-sections by complex scaling, *Phys Rept* **302** (1998), 211–293.
- [22] D. Felbacq, G. Tayeb and D. Maystre, Scattering by an arbitrary set of parallel cylinders, *J Opt Soc Am A* **11** (1994), 2526–2538.
- [23] E. Centeno and D. Felbacq, Rigorous vector diffraction of electromagnetic waves by bidimensional photonic crystals, *J Opt Soc Am A* **17** (2000), 320–327.
- [24] M. Abramowitz and I. Stegun, *Handbook of mathematical functions*, Dover Press, New York, NY, USA, 1970.
- [25] S. Stein, Addition theorems for spherical wave function, *Quart Appl Math* **19** (1961), 15–24.
- [26] O.R. Cruzan, Translation addition theorems for spherical vector wave function, *Quart Appl Math* **20** (1962), 33–40.

- [27] J. van Bladel, *Electromagnetic fields*, Hemisphere, Washington, DC, USA, 1985.
- [28] D. Felbacq and E. Centeno, Theory of diffraction for 2d photonic crystals with a boundary, *Opt Commun* **199** (2001), 39–45.
- [29] E. Centeno and D. Felbacq, Characterization of defect modes in finite bidimensional photonic crystals, *J Opt Soc Am A* **16** (1999), 2705–2712.
- [30] E. Yablonovitch and T.J. Gmitter, Donor and acceptor modes in photonic band structure, *Phys Rev Lett* **67** (1991), 3380–3383.
- [31] V. Kuzmiak and A. Maradudin, Symmetry analysis of the localized modes associates with substitutional and interstitial defects in two-dimensional triangular photonic crystal, *Phys Rev B* **61** (2000), 10750–10761.
- [32] A. Figotin and A. Klein, Localized classical waves created by defects, *J Stat Phys* **86** (1997), 165–177.
- [33] E. Centeno and D. Felbacq, Rabi oscillations in bidimensional photonic crystals, *Phys Rev B* **62** (2000), 10101–10108.
- [34] E. Centeno and D. Felbacq, Light propagation control by finite-size effect in photonic crystals, *Phys Lett A* **269** (2000), 165–170.
- [35] O. Painter, K. Srinivasan, J.D. O’Brien, A. Sherer and P.D. Dapkus, Tailoring of the resonant mode properties of optical nanocavities in two-dimensional photonic crystal slab waveguides, *J Opt A: Pure Appl Opt* **3** (2001), S161–S170.
- [36] M. Qiu, K. Azizi, A. Karlsson, M. Swillo and B. Jaskorzynska, Numerical studies of mode gaps and coupling efficiency for line-defect waveguides in two-dimensional photonic crystals, *Phys Rev B* **64** (2001), 155113.
- [37] T. Søndergaard and K.H. Dridi, Energy flow in photonic crystal waveguides, *Phys Rev B* **61** (2000), 15688–15696.
- [38] E. Centeno and D. Felbacq, Cross waveguides in biperiodic two-dimensional photonic crystals, *J Opt A: Pure Appl Opt* **3** (2001), S154–S160.
- [39] S. Olivier, C. Smith, M. Rattier, H. Benisty, C. Weisbuch, T. Krauss, R. Houdre and U. Oesterle, Miniband transmission in a photonic crystal coupled-resonator optical waveguide, *Opt Lett* **26** (2001), 1019–1022.
- [40] S. Olivier, H. Benisty, C.J.M. Smith, M. Rattier, C. Weisbuch and T.F. Krauss, Transmission properties of two-dimensional photonic crystal channel waveguides, *Opt Quant Electron* **34** (2002), 171–174.
- [41] X. Letartre, C. Seassal, C. Grillet, P. Rojo Romeo, P. Viktorovitch, M. Le Vasseur d’Yerville, D. Cassagne and C. Jouanin, Group velocity and propagation losses measurement in a single-line photonic crystal waveguide on inorganic membranes, *Appl Phys Lett* **79** (2001), 2312–2314.
- [42] E. Centeno and D. Felbacq, Optical bistability and switching through acceptor mode in a finite-size non-linear 2d photonic crystal, *Phys Rev B* **62** (2000), R7683–R7686.
- [43] C.M. Bowden and A.M. Zheltikov (eds), Nonlinear optics and photonic crystals, *J Opt Soc Am B* (Sept.) **19** (2002), 2046–2296.

- [44] A.K. Sarychev and V.M. Shalaev, Optical properties of metal-dielectric films, *Introduction to complex mediums for optics and electromagnetics* (W.S. Weiglhofer and A. Lakhtakia, eds), SPIE Press, Bellingham, WA, USA, 2003.
- [45] E. Altewischer, M.P. van Exter and J.P. Woerdman, Plasmon-assisted transmission of entangled photons, *Nature* **418** (2002), 304–306.
- [46] J. Broeng, D. Mogilevstev, S.E. Barkou and A. Bjarklev, Photonic crystal fibers: A new class of optical waveguides, *Opt Fiber Technol* **5** (1999), 305–330.
- [47] S.C. Rand, C. Soukoulis and D. Wiersma, Localization, multiple scattering, and lasing in random nanomedia, *J Opt Soc Am B* (in preparation).



Didier Felbacq was born in Cambrai, France, in June 1969. He graduated from the École Supérieure de Physique in 1992 and received the agrégation in mathematics in 1993. His Ph.D. thesis in Physics (1994) was on the theory of photonic crystals. He was an assistant professor in Physics in Clermont-Ferrand University from 1995 to 2002. He joined the University of Montpellier in November 2002 where he is a Professor of Physics. His research interests cover homogenization techniques, scattering theory, nonlinear wave propagation, and scattering by periodic and quasi-periodic structures.



Frédéric Zolla was born in Paris, France, in April 1963. He received a Ph.D. thesis at the University of Marseille in 1993. He is currently an Assistant Professor at the University of Provence and is the author of many papers on theoretical and numerical study of gratings, photonic crystals, and homogenization of the Maxwell equations. An emerging research topic is that of spectral analysis of dielectric waveguides—in particular, numerical schemes for the modeling of exotic fibers such as photonic crystal fibers and twisted fibers.

Optical Properties of Metal-Dielectric Films

Andrey K. Sarychev and Vladimir M. Shalaev

Abstract

Optical response of metal–dielectric inhomogeneous films is considered. A Generalized Ohm’s Law is formulated to relate the electric and magnetic fields *outside* to the currents *inside* the film. Computer simulations show that the local electric and magnetic fields experience giant spatial fluctuations. The fields are localized in small spatially separated peaks: electric and magnetic hot spots. Optical transmittance through a periodically inhomogeneous metallic film is strongly enhanced when the incident wave is in resonance with surface plasmon polaritons excited in the film, and can be exploited for optical switching. An analytical theory for extraordinary light transmittance through an optically thick metallic film with subwavelength holes is developed. The transmittance has sharp peaks due to the internal resonances in the holes. At a resonance, electric and magnetic fields are dramatically enhanced in the holes. These resonances are proposed to circuit light over a metallic film for distances on the nanometer length scale.

1 Introduction

The last two decades have been a time of immense improvement in our understanding of the optical properties of inhomogeneous materials. An important representative class of such materials is constituted by metal-dielectric composite materials near the percolation threshold. Nanostructured materials of this class have attracted attention because of their unique electromagnetic properties. Many fundamental phenomena, such as localization and delocalization of electrons and optical excitations, play important roles in random materials. The light-induced plasmon modes in metal-dielectric composite materials can result in dramatic enhancement of optical responses over broad spectral ranges. In particular, percolating metal-dielectric films can be employed for (i) surface-enhanced spectroscopy with unsurpassed sensitivity and (ii) novel optical elements, such as optical switches and efficient optical filters [1]-[3].

We consider the optical properties of random metal-dielectric films, also known as *semicontinuous metallic films*, in this chapter. These films are usually produced by either thermal evaporation or sputtering of a metal onto an insulating substrate. During the growth process, small clusters of metallic grains are formed first; and, eventually, a continuous conducting path appears between the ends of the sample at a percolation threshold, indicating a metal-insulator transition in the system. At high surface coverage, the film is mostly metallic with voids of irregular shapes; and finally the film becomes uniformly metallic. Although these films have been intensively studied both experimentally and theoretically for some time [4]-[19], the important role of giant local-field fluctuations, resulting from plasmon localization, was recognized only recently [20]-[38].

A two-dimensional inhomogeneous film is a thin layer within which the local physical properties are not uniform. The response of such an inhomogeneous layer to impinging plane electromagnetic waves crucially depends on (i) the spatial scale of the inhomogeneity compared to the wavelength and (ii) the angle of incidence. Usually, when the wavelength is smaller than the inhomogeneity length scale, the incident wave is scattered in various directions. The total field scattered in a certain direction is the sum of the elementary waves scattered in that direction by each elementary scatterer on the surface. As each elementary wave is specified not only by its amplitude, but also by its phase and direction of propagation, the sum of the elementary waves is a vector sum. The scattered energy is then distributed in various directions, though certain privileged directions may receive more energy than others.

By contrast, when the inhomogeneity spatial scale is much smaller than the wavelength, the resolution of the wave is too small in contrast to the nonuniformities. Therefore, the incident plane wave is then reflected specularly and transmitted in a well-defined direction, as if the film were a homogeneous layer with bulk effective physical properties (conductivity, permittivity, and permeability), as has been studied for about two centuries [39] and is considered elsewhere in this book [39]-[41]. The electromagnetic field is coupled to the inhomogeneities in such a

way that irregular currents are excited on the surface of the layer. Strong distortions of the field then appear near the surface, but they decay exponentially so that the scattered field asserts a plane wave character when far away from the surface.

2 Generalized Ohm's law approximation and giant fluctuations of local electromagnetic fields

A new analytical approach to the calculation of optical properties of metal-dielectric films, referred to as the Generalized Ohm's Law (GOL), has recently been proposed [1]. We restrict ourselves, for simplicity, to the case where all the external fields are parallel to the plane of the film (i.e., when a plane wave is normally incident on a flat surface). The z axis is perpendicular to the metallic film (with possible holes, voids, or other inhomogeneities) of thickness h . The external electromagnetic wave is incident on the front interface $z = -h/2$, and the transmitted wave is emitted from the back interface $z = h/2$, as it is shown in Figure 1. The typical spatial scale D of the metallic grains is supposed to be much smaller than the free-space wavelength λ_0 ; i.e., $D \ll \lambda_0$.

We consider first the electric and magnetic fields in the close vicinity to the film in the front reference plane $z = -h/2 - a$ and in the back reference plane $z = h/2 + a$, as it is shown in Figure 1. The electric and magnetic fields considered at distance a in front of the film are designated as

$$\mathbf{E}_1(\mathbf{r}) = \mathbf{E}(\mathbf{r}, -h/2 - a), \quad \mathbf{H}_1(\mathbf{r}) = \mathbf{H}(\mathbf{r}, -h/2 - a), \quad (1)$$

and those at a distance a behind the film are denoted by

$$\mathbf{E}_2(\mathbf{r}) = \mathbf{E}(\mathbf{r}, h/2 + a), \quad \mathbf{H}_2(\mathbf{r}) = \mathbf{H}(\mathbf{r}, h/2 + a). \quad (2)$$

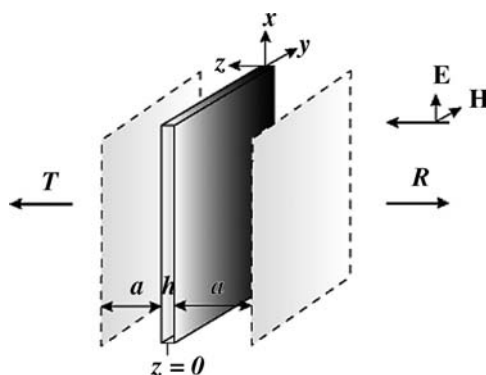


Figure 1 Electromagnetic wave of wavelength λ_0 impinges on a metal-dielectric film. It is partially reflected and absorbed, and the remainder is transmitted through the film. Electric and magnetic fields are considered in the plane $z = -h/2 - a$ in front of the film and in the plane $z = h/2 + a$ behind the film.

All the fields are monochromatic with the usual $\exp(-i\omega t)$ time-dependence. The vector $\mathbf{r} = \{x, y\}$ is a two-dimensional vector in the xy plane. In the case of laterally inhomogeneous films, the average electric displacement

$$\mathbf{D}(\mathbf{r}) = k_0 \int_{-h/2-a}^{h/2+a} \mathbf{D}(\mathbf{r}, z) dz = k_0 \int_{-h/2-a}^{h/2+a} \varepsilon(\mathbf{r}, z) \mathbf{E}(\mathbf{r}, z) dz \quad (3)$$

and the average magnetic induction

$$\mathbf{B}(\mathbf{r}) = k_0 \int_{-h/2-a}^{h/2+a} \mathbf{B}(\mathbf{r}, z) dz = k_0 \int_{-h/2-a}^{h/2+a} \mu(\mathbf{r}, z) \mathbf{H}(\mathbf{r}, z) dz \quad (4)$$

are functions of $\mathbf{r}$, where $k_0 = 2\pi/\lambda_0 = \omega/c$ is the free-space wavenumber. Vectors $\mathbf{D}$ and $\mathbf{B}$, introduced by (3) and (4), have the same dimension as electric displacement and magnetic induction. We assume hereafter, for simplicity, that the permittivity ε is a scalar and the permeability $\mu = 1$. Note that Gaussian units are used in this chapter.

In the GOL approximation, the local electromagnetic field is a superposition of two plane waves propagating in $+z$ and $-z$ directions. This superposition is different in different regions of the film. We neglect scattered and evanescent waves that propagate in the xy plane since their amplitudes are proportional to $(\lambda_0/D)^2$. Thus, we use the two-plane-wave approximation when both $\mathbf{E}(\mathbf{r}, z)$ and $\mathbf{H}(\mathbf{r}, z)$ have components in the xy plane only. The superposition of two plane waves is determined by the fields $\mathbf{E}_1$ and $\mathbf{H}_1$ or $\mathbf{E}_2$ and $\mathbf{H}_2$, defined in the front or back reference planes respectively [see (1), (2)]. Therefore, in the GOL approximation the fields $\mathbf{E}(\mathbf{r}, z)$ and $\mathbf{H}(\mathbf{r}, z)$ are completely determined by either $\mathbf{E}_1(\mathbf{r})$ and $\mathbf{H}_1(\mathbf{r})$ or $\mathbf{E}_2(\mathbf{r})$ and $\mathbf{H}_2(\mathbf{r})$ or any two linear combinations of these fields.

Then, the Maxwell equations $\text{curl} \mathbf{E}(\mathbf{r}, z) = ik_0 \mathbf{B}(\mathbf{r}, z)$ and $\text{curl} \mathbf{H}(\mathbf{r}, z) = -ik_0 \mathbf{D}(\mathbf{r}, z)$, on integration from $z = -h/2 - a$ to $z = h/2 + a$, yield

$$\{\mathbf{n} \times [\mathbf{E}_1(\mathbf{r}) - \mathbf{E}_2(\mathbf{r})]\} = i\mathbf{B}(\mathbf{r}), \quad \{\mathbf{n} \times [\mathbf{H}_1(\mathbf{r}) - \mathbf{H}_2(\mathbf{r})]\} = -i\mathbf{D}(\mathbf{r}), \quad (5)$$

where $\mathbf{n} = \{0, 0, 1\}$ is the unit normal to the xy plane. The fields $\mathbf{E}_{1,2}(\mathbf{r})$ and $\mathbf{H}_{1,2}(\mathbf{r})$ have components only in the xy plane, as their z -directed components vanish in the two-wave approximation; hence, these fields as functions of the vector $\mathbf{r}$ are curl-free (otherwise they should have z components according to the Maxwell equations).

It is convenient to introduce the linear combinations of the fields $\mathbf{E}_{1,2}(\mathbf{r})$ and $\mathbf{H}_{1,2}(\mathbf{r})$; namely the fields

$$\mathbf{E}(\mathbf{r}) = \mathbf{E}_1(\mathbf{r}) + \mathbf{E}_2(\mathbf{r}), \quad \mathbf{H}(\mathbf{r}) = \mathbf{H}_1(\mathbf{r}) + \mathbf{H}_2(\mathbf{r}) \quad (6)$$

that are also curl-free:

$$\text{curl} \mathbf{E}(\mathbf{r}) = \mathbf{0}, \quad \text{curl} \mathbf{H}(\mathbf{r}) = \mathbf{0}, \quad (7)$$

and also determine the fields inside the film. The continuity equations for the average “electric displacement” $\mathbf{D}(\mathbf{r})$ and the average “magnetic induction” $\mathbf{B}(\mathbf{r})$ in (5) are obtained by integration of the local (3D) equations $\text{div } \mathbf{D} = 0$, $\text{div } \mathbf{B} = 0$ over the coordinate “ z ” between the reference planes $z = -h/2 - a$ and $z = h/2 + a$, which results in

$$\text{div } \mathbf{D}(\mathbf{r}) = 0, \quad \text{div } \mathbf{B}(\mathbf{r}) = 0, \quad (8)$$

where we take into account that the components $D_z = 0$ and $B_z = 0$ at the planes $z = -h/2 - a$ and $z = h/2 + a$.

Equations (5), (7) and (8) are the system of GOL equations that connect electric and magnetic fields, determined in (2D) reference planes, to the average electric displacement (electric current) and average magnetic induction (magnetic current) flows in the film. Thus, the entire physics of 3D inhomogeneous film, which is described by the full set of Maxwell equations, has been reduced to a set of quasi-static equations (7) and (8). To solve these equations we just need the constitutive equations connecting the fields $\mathbf{E}$ and $\mathbf{H}$ to the “currents” $\mathbf{D}$ and $\mathbf{B}$.

Since the fields $\mathbf{E}$ and $\mathbf{H}$ can completely determine the fields inside the film in the considered GOL (two-wave) approximation, the average (2D) electric displacement $\mathbf{D}(\mathbf{r})$ and the average magnetic induction $\mathbf{B}(\mathbf{r})$ can be presented as linear combinations $\mathbf{D} = u\mathbf{E} + g_1\mathbf{H}$ and $\mathbf{B} = v\mathbf{H} + g_2\mathbf{E}$, where u , v , g_1 , and g_2 are dimensionless *ohmic* parameters. For simplicity, we consider films having mirror symmetry with respect to reflection in the $z = 0$ plane. For such films, parameters $g_1 = 0$ and $g_2 = 0$ [18-19], [22]. Therefore, we can write

$$\mathbf{D}(\mathbf{r}) = u(\mathbf{r})\mathbf{E}(\mathbf{r}), \quad \mathbf{B}(\mathbf{r}) = v(\mathbf{r})\mathbf{H}(\mathbf{r}), \quad (9)$$

where dimensionless ohmic parameters $u(\mathbf{r})$ and $v(\mathbf{r})$ are expressed in terms of the local refractive index $n = \sqrt{\varepsilon(\mathbf{r})}$ as

$$u = \frac{\tan(Dk_0/4) + n \tan(hk_0n/2)}{1 - n \tan(Dk_0/4) \tan(hk_0n/2)} \quad (10)$$

and

$$v = \frac{n \tan(Dk_0/4) + \tan(hk_0n/2)}{n - \tan(Dk_0/4) \tan(hk_0n/2)}. \quad (11)$$

The refractive index n takes the values $n_m = \sqrt{\varepsilon_m}$ and $n_d = \sqrt{\varepsilon_d}$ for the metallic and the dielectric regions of the film. The distance to the reference planes is set $a = D/4$ in (10) and (11) [16], [19], [35]. Note that this choice of a is close to $a = 2D/(3\pi)$, which is obtained from comparison of GOL and exact results [42] for the diffraction by a small aperture in a perfectly conducting screen. Equations (9) have the form typical for constitutive equations in electromagnetism [43], but include parameters u and v which incorporate local geometry of the film.

In order to find the observable optical properties of the film, such as transmittance and reflectance, we average (5) over the film plane and introduce the effective film parameters u_e and v_e , by spatial averaging; thus,

$$u_e \langle \mathbf{E} \rangle = \langle u \mathbf{E} \rangle, \quad v_e \langle \mathbf{H} \rangle = \langle v \mathbf{H} \rangle. \quad (12)$$

Thereby, we obtain the relations

$$[\mathbf{n} \times (\langle \mathbf{E}_2 \rangle - \langle \mathbf{E}_1 \rangle)] = i v_e \langle \mathbf{H} \rangle, \quad [\mathbf{n} \times (\langle \mathbf{H}_2 \rangle - \langle \mathbf{H}_1 \rangle)] = -i u_e \langle \mathbf{E} \rangle, \quad (13)$$

which relate the spatially averaged fields on both sides of the film.

We suppose that the incident plane wave lies in the half-space $z < 0$, so that its electric field depends on z as $\exp(ik_0 z)$ (see Figure 1). The incident wave is partially reflected and partially transmitted through the film. The electric field amplitude in the half-space $z < 0$, away from the film, can be written as $\tilde{\mathbf{E}}_1(z) = \mathbf{E}_0 \{\exp[ik_0(z + h/2 + a)] + r \exp[-ik_0(z + h/2 + a)]\}$, where r is the reflection coefficient and $\mathbf{E}_0$ is the amplitude of the incident wave. Well behind the film, the electric field acquires the form $\tilde{\mathbf{E}}_2(z) = t \exp[ik_0(z - h/2 - a)]$, where t is the transmission coefficient. In the reference planes $z = -h/2 - a$ and $z = h/2 + a$, the average electric field equals $\langle \mathbf{E}_1 \rangle$ and $\langle \mathbf{E}_2 \rangle$, respectively; therefore, $\langle \mathbf{E}_1 \rangle = \tilde{\mathbf{E}}_1(-h/2 - a) = (1 + r)\mathbf{E}_0$ and $\langle \mathbf{E}_2 \rangle = \tilde{\mathbf{E}}_2(h/2 + a) = t\mathbf{E}_0$. The same matching for the magnetic fields gives $\langle \mathbf{H}_1 \rangle = (1 - r)[\mathbf{n} \times \mathbf{E}_0]$ and $\langle \mathbf{H}_2 \rangle = t[\mathbf{n} \times \mathbf{E}_0]$. Substitution of these expressions for $\langle \mathbf{E}_{1,2} \rangle$ and $\langle \mathbf{H}_{1,2} \rangle$ in (6) and then in (13) gives two linear equations for t and r , whence

$$R \equiv |r|^2 = \left| \frac{(u_e - v_e)}{(i + u_e)(i + v_e)} \right|^2, \quad T \equiv |t|^2 = \left| \frac{1 + u_e v_e}{(i + u_e)(i + v_e)} \right|^2. \quad (14)$$

Thus, the effective ohmic parameters u_e and v_e completely determine the *observable* optical properties of inhomogeneous films.

Since the fields $\mathbf{E}(\mathbf{r})$ and $\mathbf{H}(\mathbf{r})$ are curl-free, they can be represented as gradients of certain scalar potentials as follows:

$$\mathbf{E} = -\nabla \varphi, \quad \mathbf{H} = -\nabla \psi. \quad (15)$$

By substituting these expressions, first in (9) and then in (8), we obtain the equations

$$\nabla \cdot [u(\mathbf{r}) \nabla \varphi(\mathbf{r})] = 0, \quad \nabla \cdot [v(\mathbf{r}) \nabla \psi(\mathbf{r})] = 0, \quad (16)$$

which are solved independently for the potentials φ and ψ . These equations are solved subject to the constraints

$$\langle -\nabla \varphi \rangle = \langle \mathbf{E} \rangle \equiv (1 + r + t)\mathbf{E}_0, \quad \langle -\nabla \psi \rangle = \langle \mathbf{H} \rangle \equiv (1 - r + t)[\mathbf{n} \times \mathbf{E}_0],$$

where the spatially uniform field $\mathbf{E}_0$ is the amplitude of the incident wave plane wave, and r and t are reflection and transmission coefficient, respectively.

The local electromagnetic fields obtained from the numerical solution of (16) are plotted in Figure 2 [35]. The maximums of the local fields are larger than the intensity of the incident plane wave by 4 to 5 orders of magnitude. The giant field fluctuations result in several new physical effects such as percolation-enhanced Rayleigh scattering, nonlinear scattering, percolation-enhanced Raman scattering, and huge enhancement of Kerr as well as other optical nonlinearities [1-2].

3 Surface plasmon polaritons

We consider now electromagnetic waves propagating on the surface of a metallic film. In the optical and infrared regimes, the collective excitation of the electron density (which is coupled to the near field) results in a surface plasmon polariton (SPP) (also known as a surface polariton [44-45]) traveling on the metal surface. These surface waves are excited when the real part of the metallic permittivity $\varepsilon_m = \varepsilon'_m + i\varepsilon''_m$ is negative (i.e., $\varepsilon'_m < 0$) and dissipation is small (i.e., $\kappa = \varepsilon''_m/|\varepsilon'_m| \ll 1$), which is typical for a metal in the optical regime. Let us denote $\varepsilon_m = -\nu^2$, where $\nu = -in$ is almost positive since the losses are small. At the metal-air interface, the SPP is an H wave, with its magnetic field parallel to the interface [45]. In the direction perpendicular to the interface, SPPs exponentially decay on both sides of the interface. The relation between the angular frequency ω and the wavenumber k_p of the SPP can be found from the following consideration.

We assume that the SPP propagates in the x direction, with $\mathbf{H}$ parallel to the y axis: $\mathbf{H} = \{0, H, 0\}$. The half-space $z > 0$ is vacuum while the metal fills the half-space $z < 0$. We seek solutions of the form

$$H_1 = H_0 \exp(ik_p x - \Lambda_1 z), \quad z > 0, \quad (17)$$

$$H_2 = H_0 \exp(ik_p x + \Lambda_2 z), \quad z < 0, \quad (18)$$

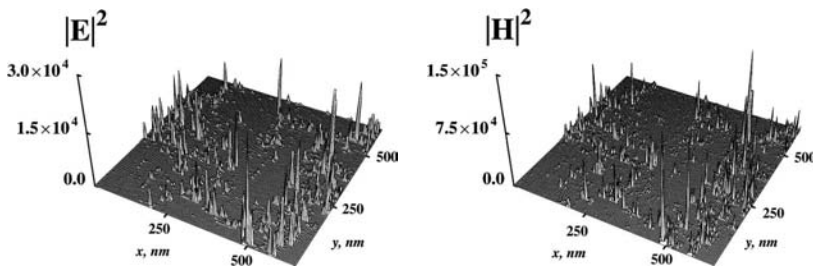


Figure 2 Distribution of local field intensities $|E|^2 = |E_1(\mathbf{r})|^2$ and $|H|^2 = |H_1(\mathbf{r})|^2$ in silver-dielectric percolation film for wavelength $\lambda_0 = 1 \mu\text{m}$ and thickness $h = 50 \text{ nm}$; concentration of silver particles is at the percolation threshold (silver surface concentration $p \approx 0.5$). Average field amplitudes are normalized to unit magnitude (i.e., $|\langle \mathbf{E}_1 \rangle| = |\langle \mathbf{H}_1 \rangle| = 1$).

where $\Lambda_1 = (k_p^2 - k_0^2)^{1/2}$ and $\Lambda_2 = [k_p^2 + (k_0 v)^2]^{1/2}$. Thus, the boundary condition on the continuity of the tangential component of the magnetic field is automatically satisfied. The continuity of the tangential component of the electric field results in the condition

$$\frac{\partial H_1}{\partial z} = -\frac{1}{v^2} \frac{\partial H_2}{\partial z} \quad (19)$$

for $z = 0$. This equation yields the dispersion equation

$$k_p = \frac{k_0 v}{\sqrt{v^2 - 1}} \quad (20)$$

for the SPP wavenumber k_p . SPP propagation on the metal surface then requires that $|v| > 1$.

There are two kinds of SPP modes in a metal film of finite thickness h , which correspond to symmetric and antisymmetric (with respect to reflection in the plane $z = 0$) oscillations of the electron density on both interfaces. Hereafter, it is supposed that $|v| > 1$. With the assumption of a strong skin effect, i.e., $\exp[-hk_0 \operatorname{Re}(v)] \ll 1$, the propagation of SPP is determined by the relation

$$k_{1,2} = k_p \left[1 \pm \frac{2v^2}{v^4 - 1} \exp(-hk_p v) \right], \quad (21)$$

where the wavenumbers k_1 and k_2 correspond to the symmetric and antisymmetric modes, respectively, and k_p is defined by (20). The phase velocities of the symmetric and antisymmetric SPPs are less than the speed of light c , and neither type can be excited by an external electromagnetic wave, because that would violate the principle of conservation of momentum. In a sense, the SPPs represent a *hidden reality* that is invisible, since an SPP does not interact with impinging light.

4 Resonant transmission

The situation changes dramatically when the metallic film is periodically inhomogeneous. The permittivity modulation provides the momentum needed to compensate for the difference between the momentums of the photon and the SPP. Hence, an SPP can be excited by an incident plane wave regardless of film thickness. An example of such spatial modulation is the square array of nanoholes punched into a metallic film [46]–[49]. Another example of a regular modulation of the refractive index, which we propose in Section 5, is light-induced modulation due to optical Kerr nonlinearity.

As before, a plane wave of interest is incident normally on the film. Then, the electromagnetic field inside the film is spatially modulated with the period L because of the film inhomogeneity, whether specially fabricated or light-induced. When the frequency of an incident wave is such that either of the SPP wavelengths

$\lambda_{1,2} = 2\pi/k_{1,2}$ coincides with L , the corresponding SPPs are excited in the film. If the film is optically thick, the SPP is excited initially on the front interface. Finally, it spreads out on *both* sides of the film.

There is a straightforward analogy between the front and back SPP components on the one hand and two identical oscillators coupled together on the other hand. The coupling can be arbitrarily weak; nevertheless, if we push the first oscillator, then after some time (which depends on the coupling) the second oscillator begins to oscillate with the same amplitude as the first oscillator. By the same token, the front and back SPP will eventually have the same amplitudes. When the SPP propagates on the back interface, it interacts with the permittivity modulation and, as a result, converts its energy back to a plane wave re-emitted from the film. Therefore, at resonance, the film becomes almost transparent, regardless of its thickness; however, the width of the transmittance resonance shrinks when the film thickness increases.

The amplitude $g \sim \Delta\varepsilon/\varepsilon$ of the permittivity modulation does not play any role in this scenario. Although g could be arbitrarily small, the front and back SPPs could be excited and the film would become transparent. Moreover, we do not need nanoholes through the film for resonant transmission to occur. All that is required is that both sides of the metal film are modulated with the same spatial period. The minimum g needed for resonant transmission depends on the loss in the metal; but the loss can be relatively small if the skin effect is strong. The transmittance maximum typically has a doublet structure corresponding to the excitation of symmetric and antisymmetric SPPs, as we show next.

A periodic modulation can always be represented by a Fourier series. Resonant transmission takes place when the wavelength of the impinging light is such that one of the SPP wavenumbers $k_{1,2}$ equals the wavenumber q of a spatial harmonic. The resonant interaction of an SPP with the q th spatial harmonic results in enhanced transmission. Since other spatial harmonics are nonresonant, we can consider light interaction with q th harmonic only. We suppose that magnetic field $\mathbf{H}$ in the wave normally impinging on the film has only a y component: $\mathbf{H} = \{0, H, 0\}$. We have to consider the interaction of the incident plane wave with a metal film whose permittivity varies as

$$\varepsilon(\mathbf{r}) = -v^2(1 + g \cos qx), \quad (22)$$

with $g \ll 1$.

The amplitude of a normally incident plane wave depends on z only. In the course of the interaction with the periodic permittivity (22), an electromagnetic harmonic varying as $\cos qx$ is generated. The amplitude of this harmonic is proportional to $g \ll 1$. This harmonic, in turn, interacts with the film modulation and thus generates a harmonic varying as $\cos 2qx$ and other higher harmonics. Thus, the whole spectrum of the electromagnetic wave is excited in the film when the incident plane wave interacts with the permittivity modulation. The amplitudes of the

$\cos pqx$ harmonics are proportional to g^p , ($p = 1, 2, 3, \dots$). Resonant transmission occurs when these harmonics are converted back to the plane wave transmitted through the film.

We are interested in the electromagnetic harmonics that can be converted back to the plane wave in such a way that this optical process is proportional to the lowest power of the modulation amplitude g . Therefore, we restrict our attention to the $\cos qx$ harmonic and write the magnetic field as $\tilde{H}(\mathbf{r}, z) = H(z) + H_q(z) \cos qx$, where $H(z)$ and $H_q(z)$ are two unknown functions. Substituting the field $\mathbf{H} = \{0, H(z) + H_q(z) \cos qx, 0\}$ in the Maxwell equations, equating the terms that have the same dependence on x [50], and neglecting the generation of higher harmonics, we obtain the system of two differential equations that determine the fields inside the film as follows:

$$\begin{aligned} \frac{d^2}{dz^2} H(z) - (k_0 v)^2 H - \frac{g}{2} \frac{d^2}{dz^2} H_q(z) &= 0, \\ \frac{d^2}{dz^2} H_q(z) - [(k_0 v)^2 + q^2] H_q - g \frac{d^2}{dz^2} H(z) &= 0. \end{aligned} \quad (23)$$

The transmitted magnetic field $H(z) = t \exp(ik_0 z)$, ($z > h/2$), where $T = |t|^2$ is the transmittance. For the $\cos qx$ harmonic, we use the radiative boundary conditions; namely,

$$H_q(z) = \begin{cases} Y_3 \exp\left(-i\sqrt{k_0^2 - q^2}z\right), & z < -h/2 \\ Y_4 \exp\left(i\sqrt{k_0^2 - q^2}z\right), & z > h/2 \end{cases}, \quad (24)$$

where Y_3 and Y_4 are constants.

We neglect nonresonant (direct) transmittance, which allows us to obtain a rather simple expression for the resonant transmittance:

$$T(\tilde{\Delta}) = \frac{4\tilde{g}^4}{[(\tilde{\Delta} - 1)^2 + (\tilde{g}^2 + \tilde{\kappa})^2][(\tilde{\Delta} + 1)^2 + (\tilde{g}^2 + \tilde{\kappa})^2]}. \quad (25)$$

This quantity depends on

$$\tilde{\Delta} = g^2 \frac{|v|(|v| - m)^2(|v| + |v|^3 + 2m)}{8(1 + |v|^2)\zeta} - \frac{\Delta}{\zeta} \frac{m(1 + |v|^2)}{2|v|}, \quad (26)$$

which is the normalized detuning from the SPP frequency; and it also depends on the renormalized modulation amplitude

$$\tilde{g} = \frac{g|v|\sqrt{m}(|v| - m)}{2\sqrt{1 + |v|^2}\sqrt{\zeta}} \quad (27)$$

as well as the renormalized loss factor

$$\tilde{\kappa} = \frac{(1 + |\nu|^2)\kappa}{4|\nu|^2\zeta}, \quad (28)$$

where $m = \sqrt{|\nu|^2 - 1}$, $\Delta = k_0/q - m/|\nu|$, and $\zeta = \exp(-h|\nu|q)$. Recall that we wrote $\varepsilon_m = -\nu^2$, and we also assumed small losses (i.e., $\kappa = \varepsilon_m''/|\varepsilon_m'| \ll 1$) and $|\nu| > 1$.

To analyze the resonant transmission, for simplicity, we set $\tilde{\kappa} = 0$ in (25). Then, for $\tilde{g} < 1$, it follows that $T(\tilde{\Delta})$ has two maximums; namely, $T(\tilde{\Delta}_1) = T(\tilde{\Delta}_2) = 1$ for $\tilde{\Delta}_{1,2} = \pm\sqrt{1 - \tilde{g}^4}$. Therefore, a lossless film becomes absolutely transparent at resonances, regardless of its thickness.

It is instructive to consider how the transmittance changes when $\tilde{g}$ increases. The distance between the two maximums, given by $\tilde{\Delta}_1 - \tilde{\Delta}_2 = 2\sqrt{1 - \tilde{g}^4}$, decreases with increasing $\tilde{g}$. The film remains transparent at both resonances. Finally, when $\tilde{g}$ exceeds unity, the two maximums merge together; and the transmission spectrum displays just one maximum, with transmittance $T_m = 4\tilde{g}^4/(1 + \tilde{g}^2)^2$ that decreases with further increase of $\tilde{g}$, as shown in Figure 3. The reason is that interaction with the permittivity modulation results in the radiative decay of the SPP and its conversion to the emitted plane wave. The radiative losses, conveyed by the $\tilde{g}^2$ term in the denominator of (25), lead to a damping of the SPP. As a result, resonant transmittance increases with a decreasing permittivity modulation $\tilde{g}$, and *vice versa*.

Resonant transmittance of a silver film is shown in Figure 4. The plot holds for $\kappa = 1.6 \times 10^{-3}$, which is realistic at cryogenic temperatures, when the electron mean-free path is determined by the film thickness h . Clearly, the transmittance has sharp two peaks corresponding to the SPP excitation, although the transmittance is less than 100% due to losses. Outside the resonance regime, the transmittance is estimated as $T \sim \exp(-2|\nu|k_0d) \sim 10^{-6}$. Thus, at resonance, the transmittance is enhanced by five orders of magnitude.

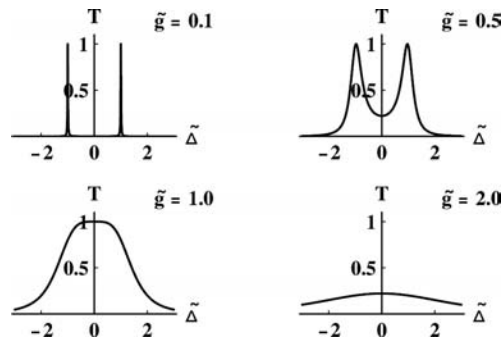


Figure 3 Transmittance as a function of normalized detuning $\tilde{\Delta}$ for different modulations $\tilde{g}$.

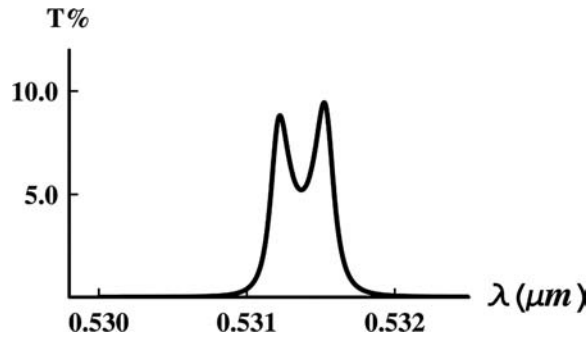


Figure 4 Transmittance of a silver film of thickness $d = 0.18 \mu\text{m}$, modulation $g = 0.1$, and spatial period $a = 2\pi/q = 0.5 \mu\text{m}$.

5 Light-induced resonant transmission

In Section 4, we assumed that the periodic permittivity modulation to be fabricated. Let us now consider the creation and control of this modulation by light itself through optical nonlinearity of the Kerr type [51]. We first suppose that film has a “seed” modulation $g_0 \ll 1$ and then determine the increase in g due to the film nonlinearity. Exactly at resonance, the transmitted intensity $I_t = T I_0$ is of the same order of magnitude as the intensity I_0 of the incident plane wave. The transmitted wave is generated by the SPP which propagates on the front interface ($z = -h/2$) because of interaction with the permittivity modulation. Therefore, the intensity I_p of the SPP is estimated as $I_p \sim I_t/g^2 \sim I_0/g^2 \gg I_0$. At the back interface ($z = h/2$), the SPP intensity is of the same order of magnitude. The electric field E_p of the SPP is spatially modulated with the resonance wavenumber k_p . Thus, the field-induced modulation of the permittivity is estimated via $g \sim 24\pi\chi^{(3)}|E_p|^2$, where $\chi^{(3)}$ is the nonlinear susceptibility quantifying the Kerr effect. The induced modulation of the permittivity increases the transmittance and, therefore, the intensity of the SPP. This positive feedback may result in the bistability depicted in Figure 5.

When dimensionless intensity of the impinging light I^* becomes larger than $I_1^* \simeq 6 \times 10^{-3}$, the transmittance T jumps from nearly zero to almost unity, and the film suddenly becomes transparent. If I^* is reduced thereafter, the film remains transparent even for $I^* < I_1^*$, since the SPP has been already excited in the film. Transmission declines steeply for $I^* \sim I_2^* \simeq 10^{-3} < I_1^*$. Thus, optical bistability can occur in periodically nonhomogeneous metallic films. The susceptibility $\chi^{(3)}$ is rather large for noble metals (typically, $\chi^{(3)} > 10^{-8}$ esu [52-53]), and the intensity I_0 required for the bistability can be easily achieved with conventional lasers. We also note that the seed modulation g_0 can be created by the interference of two additional control laser beams that are incident on the surface from the different sides with respect to the normal.

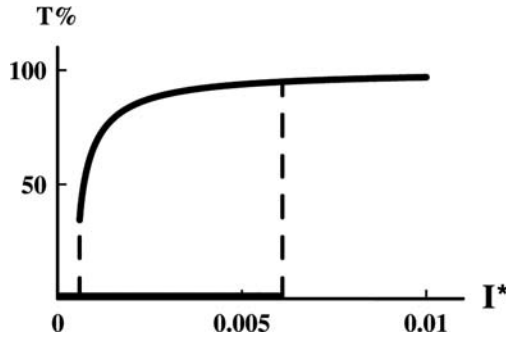


Figure 5 Nonlinear transmittance as a function of the intensity I_0 of the incident plane wave; $I^* = 24\pi^2 I_0 \chi^{(3)}/c$.

6 Extraordinary optical transmittance through nanoholes

Let us now apply the approach developed in Section 2.4 in order to investigate the interaction of light with nanoholes punched in an otherwise optically thick metal film. We show that the transmittance of a metal film with subwavelength-sized holes has sharp resonances corresponding to the excitation of the localized surface waves that are specific to a metal film with holes [54].

This extraordinary optical transmittance (EOT) phenomenon was discovered by Ebbesen *et al.* [46], who also subjected it to intensive examination; see also Sonnichsen *et al.* [55]. Various models—mostly numerical simulations—were suggested to explain the EOT [35], [49], [56]–[60]. Despite the very sophisticated simulation algorithms used, the physical nature of this phenomenon is not fully understood. Here we use a GOL approximation to develop a physical model that provides a simple qualitative picture.

Our objective is to find the transmittance of a metal film with subwavelength-sized holes. The local electric and magnetic fields, $\mathbf{E}(\mathbf{r})$ and $\mathbf{H}(\mathbf{r})$, can be determined from the potentials involved in (16); and then the effective parameters u_e and v_e can be obtained from the definitions (12). Since both $\mathbf{E}(\mathbf{r})$ and $\mathbf{H}(\mathbf{r})$ are curl-free, while both $\mathbf{D}(\mathbf{r})$ and $\mathbf{B}(\mathbf{r})$ are divergence-free, the quasi-static approximation can be used [1], [35]. Therefore, many efficient analytical and numerical methods, which were developed in quasi-static percolation theory, are at our disposal for the calculation of the local fields and the effective parameters [1], [4].

Here we use the simplest approximation—the Maxwell Garnett (MG) approach—that holds when the surface hole concentration p is small (i.e., $p \ll 1$). In two dimensions, the MG approach leads to the following expressions:

$$E_h = \frac{2u_m}{u_m + u_h} E_m, \quad H_h = \frac{2v_m}{v_m + v_h} H_m, \quad (29)$$

where u_m , v_m and v_h , v_h are the ohmic parameters for the metal and holes, E_m and H_m are electric and magnetic fields averaged over the metal, E_h and H_h are the

fields averaged over the holes. From (12) and (29) we obtain the effective *electric* parameter:

$$u_e \equiv \frac{\langle uE \rangle}{\langle E \rangle} = \frac{(1-p)u_m E_m + pu_h E_h}{(1-p)E_m + pE_h}. \quad (30)$$

Repeating the same procedure, we find the effective *magnetic* parameter

$$v_e \equiv \frac{\langle vH \rangle}{\langle H \rangle} = \frac{(1-p)v_m H_m + pv_h H_h}{(1-p)H_m + pH_h}, \quad (31)$$

analogously.

Hereafter we consider films whose thickness h is much larger than the skin depth δ in bulk metal ($h \gg \delta = 1/\text{Im}[n]k_0$). This case corresponds to most of the experiments with subwavelength holes reported so far. With the assumption of strong skin effect, (10) and (11) readily yield

$$u_m = -\cot(ak_0), \quad v_m = \tan(ak_0), \quad (32)$$

so that $u_m = -1/v_m$. We substitute the parameters u_e and v_e in (14) to obtain the transmittance

$$T = \frac{16p^2|u_m|^2(1+u_h v_h)^2}{|\Sigma_1 \Sigma_2|^2}, \quad (33)$$

where

$$\Sigma_1 = u_h - pu_h + (1+p)(1-iu_h)u_m - i(1-p)u_m^2 \quad (34)$$

and

$$\Sigma_2 = (i+u_m)(u_m v_h - 1) + p(i-u_m)(u_m v_h + 1). \quad (35)$$

In deriving (33), we assumed that (i) $p \ll 1$ and (ii) $|u_m| \gg 1$, which follows from (32) when the hole diameter $D \sim a \ll \lambda$.

The electric field E_h in a hole tends towards infinity in the limit $u_m \rightarrow -u_h$ if there are no losses; see (29). By utilizing the relation $u_m = -u_h$ in (33), we obtain transmittance $T = 4|u_m|/|1+u_m^2|$, which does not depend on the hole concentration p and, therefore, remains finite even as $p \rightarrow 0$. Similarly, when magnetic resonance takes place (i.e., $v_m = -1/u_m = -v_h$), the resonant transmittance also remains finite as $p \rightarrow 0$. Consequently, we conclude that the electric and magnetic resonances in the holes can result in EOT.

Note, however, that a strong electric field in a hole could excite an SPP that emerges from the hole. Then, the resonance would be damped due to radiative losses. As radiative losses are ignored in the following analysis, the expressions for the transmittance obtained in the remainder of this section should be considered as merely estimates.

7 Electric and magnetic resonances

To estimate the transmittance we should find ohmic parameters u_h and v_h for a hole. The parameters u_h , and v_h depend on the field distribution inside the hole. The internal field is a superposition of different eigenmodes for a subcritical waveguide that characterizes a hole. At the hole entrance, the internal field is similar to a plane wave, even though its amplitude can be significantly different from the amplitude of the incident wave. Deeper inside the hole, only the mode with the least eigenvalue survives. To simplify further semiquantitative analysis, we assume that the internal field is a plane wave near the entrance of the hole and it matches the fundamental internal mode at a distance a from both ends of the hole. We use for this matching the same distance a that we used earlier to match local fields with the incident plane wave. As a result, we obtain

$$u_h = v_h = \tan[(a + h/2)k_0] \quad (36)$$

for $h < 2a$, and

$$u_h = \frac{k_0 \tan(2ak_0) - \sqrt{\kappa^2 - k_0^2} \tanh[(h/2 - a)\sqrt{\kappa^2 - k_0^2}]}{k_0 + \sqrt{\kappa^2 - k_0^2} \tan(2ak_0) \tanh[(h/2 - a)\sqrt{\kappa^2 - k_0^2}]}, \quad (37)$$

$$v_h = \frac{\sqrt{\kappa^2 - k_0^2} \tan(2ak_0) + k_0 \tanh[(h/2 - a)\sqrt{\kappa^2 - k_0^2}]}{\sqrt{\kappa^2 - k_0^2} - k_0 \tan(2ak_0) \tanh[(h/2 - a)\sqrt{\kappa^2 - k_0^2}]} \quad (38)$$

for $h > 2a$. Here, $\kappa = 3.68/D$ is the eigenvalue for the basic mode in a cylindrical waveguide of diameter D [45, Ch. 92]. We denote the hole diameter and diameter of a metal particle (see Section 2) by the same symbol D , since in both cases D characterizes the typical size of an inhomogeneity.

Consider first shallow holes of depth $h < 2a$. By substituting (32) and (36) into (33) and invoking the limit $p \ll 1$, we obtain

$$T(k_0) = \sum_{j=1,2,3,\dots} \frac{A_j(k_0)}{A_j(k_0) + (4a + h)^2(k_0 - k_j)^2}, \quad (39)$$

where

$$A_j(k_0) = 4p^2 \sin^4\left(\frac{2aj\pi}{4a + h}\right) \quad (40)$$

and

$$k_j = \frac{j\pi}{4a + h} - \frac{p}{4a + h} \sin\left(\frac{4aj\pi}{4a + h}\right). \quad (41)$$

Clearly, the transmittance in (39) is the sum of the resonances located at $k_0 = k_j$, as shown in Figure 6. The transmittance T is almost a periodic function of k_0 for $p \ll 1$, though the peak width depends on k_0 . A peak can disappear when the corresponding numerator $A_j(k_j)$ in (39) vanishes. The odd-numbered resonances in (39) correspond to a high electric field, whereas the even-numbered resonances indicate a high magnetic field in the holes of the film. This classification is shown clearly in Figure 7.

For deep holes ($h > 2a$), we obtain the transmittance by substituting the ohmic parameters from (32) and (38) into (33). The transmittance T is plotted in Figure 8 as a function of Dk_0 . It follows from this plot that the k_0 -dependence of T can be rather peculiar when the thickness of the film increases: the peaks corresponding to high electric and high magnetic fields in the holes can move closer and even merge together.

With the assumption of losslessness, the magnitude of the electric or magnetic field in the holes tends toward infinity at a transmittance resonance. In any real film, however, the resonant fields must remain of finite magnitudes by virtue of losses. Impedance boundary conditions [45, Ch. 87] assist in the incorporation of losses in the foregoing analyses. The transmittance thus obtained for a square array

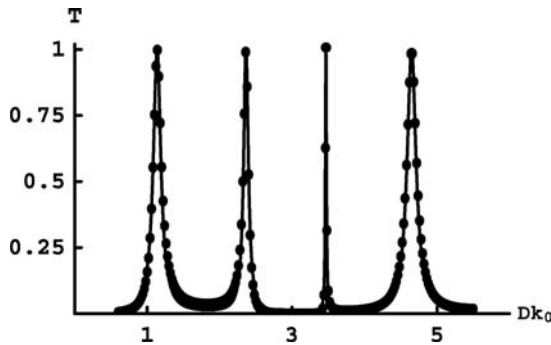


Figure 6 Transmittance through a metallic film with shallow holes ($h < 2a$); $a/D = 0.6$, $h/D = 0.8$, $p = 0.1$. The solid line is the resonance approximation (39), while the dots represent calculations with (33).

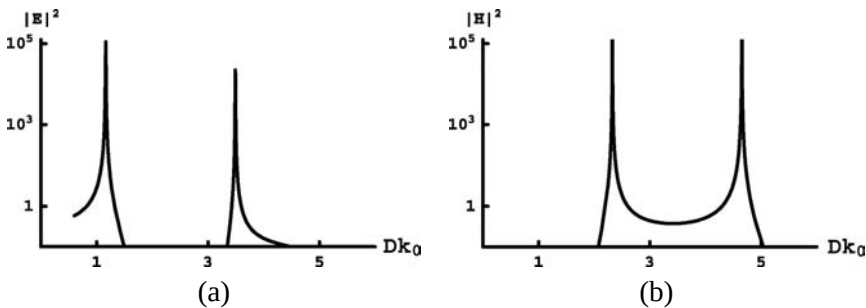


Figure 7 Magnitudes of the (a) electric field and (b) the magnetic field in a hole, when the incident field amplitude equals unity. See Figure 6 for the system parameters.

of nanoholes in a silver film is in qualitative agreement with experiments [46-47], [62], as illustrated in Figure 9. Most peaks in the transmittance are due to excitation of the SSP in the holes.

In calculating the transmittance presented in Figure 9, we also accounted for the SPP, which is not localized in a hole but propagates over the entire surface of the film. The propagating SPP is excited when the distance between adjacent holes coincides with $2\pi/k_{1,2}$ [see (20) and (21)]. This results in a peak at $\lambda_0 \simeq 0.6 \mu\text{m}$ in Figure 9(a), as well as for a small change in amplitude of the peak at $\lambda_0 \simeq 0.8 \mu\text{m}$ in Figure 9(b).

8 Light circuiting in nanoholes

We have considered so far the uniform illumination of the metal films by a plane wave. It is interesting to consider another possibility when only one of the holes

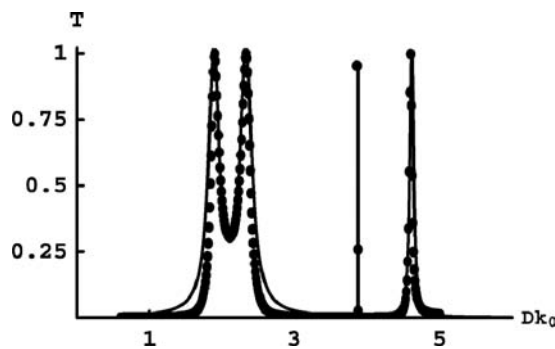


Figure 8 Transmittance through a film with deep holes ($h > 2a$); $a/D = 0.6$, $h/D = 1.45$, $p = 0.1$.

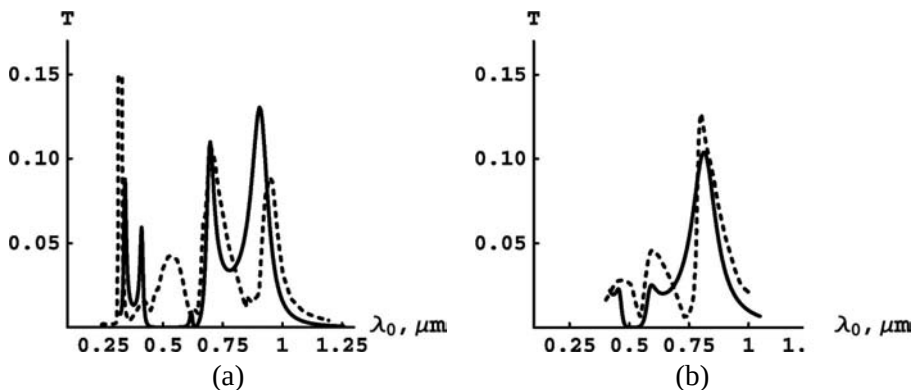


Figure 9 Figure 9: Extraordinary optical transmittance through a square array of nanoholes in (a) silver film on a quartz substrate ($D = 0.15 \mu\text{m}$, $h = 0.2 \mu\text{m}$, $p = 0.049$, $a/D = 0.6$), and (b) free-standing silver film ($D = 0.28 \mu\text{m}$, $h = 0.32 \mu\text{m}$, $p = 0.049$, $a/D = 0.6$). The dashed lines represent experimental data. The solid lines show theoretical results.

is illuminated by a nanosized light source. This can be accomplished, for example, with a nanometer-size probe of a near-field scanning optical microscope [61]. At resonance, the electric field (and/or the magnetic field) spreads out from the illuminated hole toward other holes, because of interactions between the holes via plasmons. Such holes can be arranged into any desired pattern to localize light and circuit the propagation of the electromagnetic energy, as shown in Figure 10. Thus, in Figures 10(c) and 10(d), we show how light circuiting depends on the polarization of the source. When the electric field, which excites the central hole, changes its polarization from parallel to “ T ” stem [Figure 10(c)] to perpendicular [Figure 10(d)], the light changes the direction of its propagation at a scale that is much smaller than the wavelength of the guided light.

9 Concluding remarks

The observable optical properties of thin metal-dielectric films were reviewed in this chapter. An approximation premised on the Generalized Ohm’s Law (GOL) allows calculation of the field distributions in and the transmittances of inhomogeneous metal films. Computer simulations show giant electromagnetic field

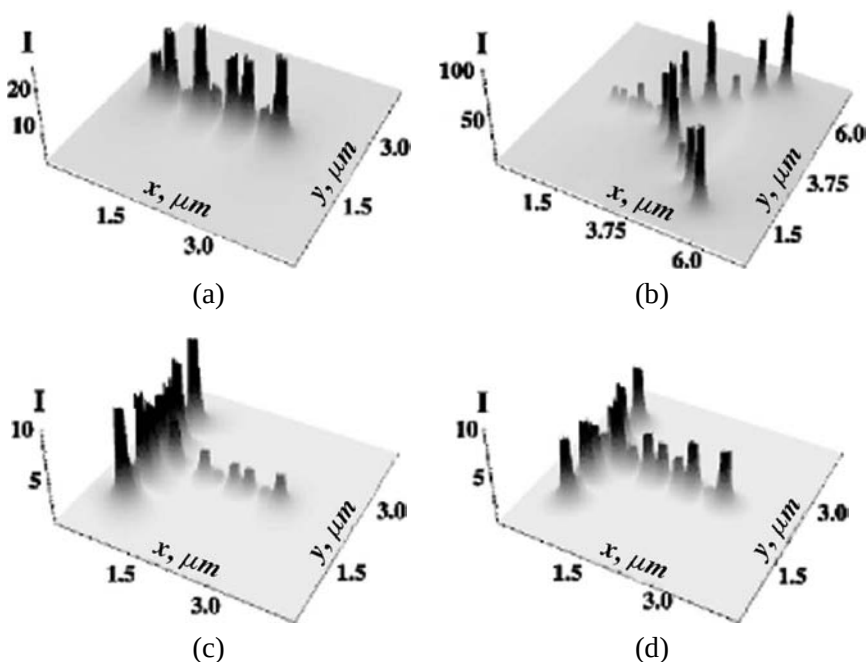


Figure 10 Nanocircuits formed by a pattern of nanoholes in a metallic film. (a) A line waveguide, wherein the first left hole is excited; (b) a fork waveguide, wherein first left hole is excited, (c) and (d) switch; center hole in top “ T ” is excited by electric field E : (c) $E \perp$ top T , (d) $E \parallel$ top T . Input parameters: $D = 0.15 \mu\text{m}$, $h = 0.2 \mu\text{m}$, $\lambda = 0.78 \mu\text{m}$, and distance between the holes centers $= 0.3 \mu\text{m}$.

fluctuations in such films. Indeed, giant electric field fluctuations near the percolation threshold in metal-dielectric films have already been observed in the microwave [22] and optical [32-33], [37-38] regimes. These fluctuations result in a huge enhancement of various optical effects. For example, surface enhancement for Raman scattering is proportional to the fourth moment of the field [24], [26], which is strongly enhanced. The same is valid for the Kerr nonlinearity, which is also proportional to the fourth moment [25], [31].

The excitation of a surface plasmon polariton (SPP) in periodically modulated metal films can result in resonant transmission, so that an optically thick film can become transparent. The transmittance can be increased by factors of $> 10^5$ at the resonance.

Resonant transmittance spectrums have a characteristic double-peak structure due to the splitting of SPPs into symmetric and antisymmetric modes. The resonance transmittance increases with decreasing losses in the system, which can be achieved by cooling the film to cryogenic temperatures. Exactly at the resonance, the amplitude of the excited SPP can be larger than the amplitude of the incident wave by several orders of magnitude. Then, optical nonlinearity can be significantly enhanced. We predicted that at sufficiently large intensities of the impinging light, the film can manifest the optical bistability, a phenomenon that can be exploited for optical switching.

Extraordinary light transmittance through optically thick metallic films with subwavelength-sized holes is possible. The transmittance has sharp resonances corresponding to the excitation of various surface waves. Some of these waves are similar to SPP, while others are localized surface waves that are specific for a perforated metal and have not been discussed elsewhere. Nanoholes can be arranged into any desired pattern to localize and circuit em energy. Such nanoengineered structures could be used as integrated elements in various optoelectronic and photonic devices, including most sophisticated ones such as optical computers.

Acknowledgment

The authors are grateful to Academician A.M. Dykhne and Dr. V. Podolskiy for useful discussions. This work was supported by Battelle (DAAD19-02-D-0001), NASA (NCC-1-01049), ARO (DAAD19-01-1-0682), and NSF (ESC-0210445).

References

- [1] A.K. Sarychev and V.M. Shalaev, Electromagnetic field fluctuations and optical nonlinearities in metal-dielectric composites, *Phys Rept* **335** (2000), 275–371.
- [2] A.K. Sarychev and V.M. Shalaev, Field distribution, Anderson localization, and optical phenomena in random metal-dielectric films, *Optics of nanostructured materials* (V.A. Markel and T. George, eds), Wiley, New York, NY, USA, 2000.

- [3] V.M. Shalaev, *Nonlinear optics of random media: Fractal composites and metal-dielectric films*, Springer, Heidelberg, Germany, 2000.
- [4] D.J. Bergman and D. Stroud, Physical properties of macroscopically inhomogeneous media, *Solid State Phys* **46** (1992), 147–269.
- [5] J.P. Clerc, G. Giraud, J.M. Laugier and J.M. Luck, The AC electrical-conductivity of binary disordered-systems, percolation clusters, fractals and related models, *Adv Phys* **39** (1990), 191–308.
- [6] R.W. Cohen, G.D. Cody, M.D. Coutts and B. Abeles, Optical properties of granular silver and gold films, *Phys Rev B* **8** (1973), 3689–3701.
- [7] G.A. Niklasson and C.G. Granqvist, Optical-properties and solar selectivity of coevaporated Co-Al₂O₃ composite films, *J Appl Phys* **55** (1984), 3382–3410.
- [8] L.C. Botten and R.C. McPhedran, Phase constraints for lossy symmetric structures, *Optica Acta* **32** (1985), 595–605.
- [9] M. Gajdardziska-Josifovska, R.C. McPhedran, D.R. McKenzie and R.E. Collins, Silver magnesium fluoride cermet films. Optical and electrical-properties, *Appl Opt* **28** (1989), 2744–2753.
- [10] C.A. Davis, D.R. McKenzie and R.C. McPhedran, Optical-properties and microstructure of thin silver films, *Opt Commun* **85** (1991), 70–82.
- [11] P. Gadenne, A. Beghadi and J. Lafait, Optical crossover analysis of granular gold-films at percolation, *Opt Commun* **65** (1988), 17–21.
- [12] P. Gadenne, Y. Yagil and G. Deutscher, Transmittance and reflectance insitu measurements of semicontinuous gold-films during deposition, *J Appl Phys* **66** (1989) 3019–3025.
- [13] Y. Yagil, M. Yosefin, D.J. Bergman, G. Deutscher and P. Gadenne, Scaling theory for the optical-properties of semicontinuous metal-films, *Phys Rev B* **43** (1991), 11342–11352.
- [14] Y. Yagil, P. Gadenne, C. Julien and G. Deutscher, Optical-properties of thin semicontinuous gold-films over a wavelength range of 2.5 to 500 μm , *Phys Rev B* **46** (1992), 2503–2511.
- [15] T.W. Noh, P.H. Song, Sung-Il Lee, D.C. Harris, J.R. Gaines and J.C. Garland, Far-infrared studies of 2-dimensional random metal-insulator composites, *Phys Rev B* **46** (1992), 4212–4221.
- [16] A.K. Sarychev, D.J. Bergman and Y. Yagil, Optical and microwave properties of metal-insulator thin films: possibility of light localization, *Physica A* **207** (1994), 372–378.
- [17] A.K. Sarychev, D.J. Bergman and Y. Yagil, Theory of the optical and microwave properties of metal-dielectric films, *Phys Rev B* **51** (1995), 5366–5385.
- [18] R. Levy-Nathansohn and D.J. Bergman, Studies of the generalized Ohm's law, *Physica A* **241** (1997), 166–172.
- [19] R. Levy-Nathansohn and D.J. Bergman, Decoupling and testing of the generalized Ohm's law, *Phys Rev B* **55** (1997), 5425–5439.

- [20] S. Blacher, F. Brouers and A.K. Sarychev, Multifractality of giant electric field fluctuations in semicontinuous films, *Fractals in the natural and applied sciences* (M.M. Novak, ed), Chapman and Hall, London, UK, 1995.
- [21] E.M. Baskin, M.V. Entin, A.K. Sarychev and A.A. Snarskii, Enhancement of high-frequency field in near-ideal metal mixture, *Physica A* **242** (1997), 49–56.
- [22] A.N. Lagarkov, K.N. Rozanov, A.K. Sarychev and A.N. Simonov, Experimental and theoretical study of metal-dielectric percolating films at microwaves, *Physica A* **241** (1997), 199–206.
- [23] F. Brouers, A.K. Sarychev, S. Blacher and O. Lothaire, Giant field fluctuations and characteristic lengths in optical-properties of nano metal-dielectric films, *Physica A* **241** (1997), 146–153.
- [24] F. Brouers, S. Blacher, A.N. Lagarkov, A.K. Sarychev, P. Gadenne and V.M. Shalaev, Theory of giant Raman scattering from semicontinuous metal films, *Phys Rev B* **55** (1997), 13234–13245.
- [25] V.M. Shalaev and A.K. Sarychev, Nonlinear optics of random metal-dielectric films, *Phys Rev B* **57** (1998), 13265–13288.
- [26] P. Gadenne, F. Brouers, V.M. Shalaev and A.K. Sarychev, Giant stokes fields on semicontinuous metal films, *J Opt Soc Am B* **15** (1998), 68–72.
- [27] V.M. Shalaev, E.Y. Poliakov, V.A. Markel, V.P. Safonov and A.K. Sarychev, Surface-enhanced optical nonlinearities of nanostructured fractal materials, *Fractals* **5** (1997), 63–82.
- [28] V.M. Shalaev, V.A. Markel, E.Y. Poliakov, R.L. Armstrong, V.P. Safonov and A.K. Sarychev, Nonlinear Optical Phenomena in Nanostructured Fractal Materials, *J Nonlin Opt Phys Mat* **7** (1998), 131–152.
- [29] F. Brouers, S. Blacher and A.K. Sarychev, Giant field fluctuations and anomalous scattering in semicontinuous metal films, *Phys Rev B* **58** (1998), 15897–15903.
- [30] A.K. Sarychev, V.A. Shubin and V.M. Shalaev, Percolation-enhanced nonlinear scattering from metal-dielectric composites, *Phys Rev E* **59** (1999), 7239–7242.
- [31] A.K. Sarychev, V.A. Shubin and V.M. Shalaev, Anderson localization of surface plasmons and nonlinear optics of metal-dielectric composites, *Phys Rev B* **60** (1999), 16389–16408.
- [32] S. Grésillon, L. Aigouy, A.C. Boccara, J.C. Rivoal, X. Quelin, C. Desmarest, P. Gadenne, V.A. Shubin, A.K. Sarychev and V.M. Shalaev, Experimental observation of localized optical excitations in random metal-dielectric films, *Phys Rev Lett* **82** (1999), 4520–4523.
- [33] S. Grésillon, J.C. Rivoal, P. Gadenne, X. Quelin, V.M. Shalaev and A.K. Sarychev, Nanoscale observation of enhanced electromagnetic field, *Phys Stat Sol (a)* **175** (1999), 337–346.
- [34] A.K. Sarychev, V.A. Shubin and V.M. Shalaev, Giant high-order field moments in metal-dielectric composites, *Physica A* **266** (1999), 115–122.

- [35] V.A. Shubin, A.K. Sarychev, J.P. Clerc and V.M. Shalaev, Local electric and magnetic fields in semicontinuous metal films: Beyond the quasistatic approximation, *Phys Rev B* **62** (2000), 11230–11244.
- [36] M.I. Stockman, S.V. Faleev and D.J. Bergman, Localization versus delocalization of surface plasmons in nanosystems: Can one state have both characteristics? *Phys Rev Lett* **87** (2001), 167401.
- [37] M. Breit, V.A. Podolskiy, S. Gresillon, G. von Plessen, J. Feldmann, J.C. Rivoal, P. Gadenne, A.K. Sarychev and V.M. Shalaev, Experimental observation of percolation-enhanced nonlinear light scattering from semicontinuous metal films, *Phys Rev B* **64** (2001), 125106.
- [38] K. Seal, M.A. Nelson, Z.C. Ying, D.A. Genov, A.K. Sarychev and V.M. Shalaev, Growth, morphology, and optical and electrical properties of semicontinuous metallic films, *Phys Rev B* **67** (2003), 035318.
- [39] A. Lakhtakia (ed), *Selected papers on linear optical composite materials*, SPIE Press, Bellingham, WA, USA, 1996.
- [40] G.B. Smith, Nanostructured thin films, *Introduction to complex mediums for optics and electromagnetics* (W.S. Weiglhofer and A. Lakhtakia, eds), SPIE Press, Bellingham, WA, USA, 2003.
- [41] T.G. Mackay, Homogenization of linear and nonlinear complex composite materials, *Introduction to complex mediums for optics and electromagnetics* (W.S. Weiglhofer and A. Lakhtakia, eds), SPIE Press, Bellingham, WA, USA, 2003.
- [42] H.A. Bethe, Theory of diffraction by small holes, *Phys Rev* **66** (1944), 163–182.
- [43] W.S. Weiglhofer, Constitutive characterization of simple and complex mediums, *Introduction to complex mediums for optics and electromagnetics* (W.S. Weiglhofer and A. Lakhtakia, eds), SPIE Press, Bellingham, WA, USA, 2003.
- [44] V.D. Freilikher, E. Kanzieper and A.A. Maradudin, Coherent scattering enhancement in systems bounded by rough surfaces, *Phys Rept* **288** (1997), 127–204.
- [45] L.D. Landau, E.M. Lifshitz and L.P. Pitaevskii, *Electrodynamics of continuous media, 2nd ed*, Pergamon Press, Oxford, UK, 1984.
- [46] T.W. Ebbesen, H.J. Lezec, H.F. Ghaemi, T. Thio and P.A. Wolff, Extraordinary optical transmission through sub-wavelength hole arrays, *Nature* **391** (1998), 667–669.
- [47] H.F. Ghaemi, T. Thio, D.E. Grupp, T.W. Ebbesen and H.J. Lezec, Surface plasmons enhance optical transmission through subwavelength holes, *Phys Rev B* **58** (1998), 6779–6782.
- [48] T.J. Kim, T. Thio, T.W. Ebbesen, D.E. Grupp and H.J. Lezec, Control of optical transmission through metals perforated with subwavelength hole arrays, *Opt Lett* **24** (1999), 256–258.
- [49] L. Martin-Moreno, F.J. Garcia-Vidal, H.J. Lezec, K.M. Pellerin, T. Thio, J.B. Pendry and T.W. Ebbesen, Theory of extraordinary optical transmission through subwavelength hole arrays, *Phys Rev Lett* **86** (2001), 1114–1117.

- [50] Lord Rayleigh, *The theory of sound, 2nd ed*, Macmillan, London, UK, 1896.
- [51] H.E. Brandt (ed), *Selected papers on nonlinear optics*, SPIE Press, Bellingham, WA, USA, 1991.
- [52] S. Debrus, J. Lafait, M. May, N. Pincon, D. Prot, C. Sella and J. Venturini, Z-scan determination of the third-order optical nonlinearity of gold: silica nanocomposites, *J Appl Phys* **88** (2000), 4469–4475.
- [53] P.J. Bennett, V. Albanis, Y.P. Svirko and N.I. Zheludev, Femtosecond cubic optical nonlinearity of thin nickel films, *Opt Lett* **24** (1999), 1373–1375.
- [54] A.K. Sarychev, V.A. Podolskiy, A.M. Dykhne and V.M. Shalaev, Resonance transmittance through a metal film with subwavelength holes, *IEEE J Quant Electron* **38** (2002), 956–963.
- [55] C. Sonnichsen, A.C. Duch, G. Steininger, M. Koch, G. von Plessen and J. Feldmann, Launching surface plasmons into nanoholes in metal films, *Appl Phys Lett* **76** (2000), 140–142.
- [56] W.C. Tan, T.W. Preist, R.J. Sambles and N.P. Wanstall, Flat surface-plasmon-polariton bands and resonant optical absorption on short-pitch metal gratings, *Phys Rev B* **59** (1999), 12661–12666.
- [57] W.C. Tan, T.W. Preist and R.J. Sambles, Resonant tunneling of light through thin metal films via strongly localized surface plasmons, *Phys Rev B* **62** (2000), 11134–11138.
- [58] L. Salomon, F.D. Grillot, A.V. Zayats and F. de Fornel, Near-field distribution of optical transmission of periodic subwavelength holes in a metal film, *Phys Rev Lett* **86** (2001), 1110–1113.
- [59] R. Wannemacher, Plasmon-supported transmission of light through nanometric holes in metallic thin films, *Opt Comm* **195** (2001), 107–118.
- [60] Q. Cao and P. Lalanne, Negative role of surface plasmons in the transmission of metallic gratings with very narrow slits, *Phys Rev Lett* **88** (2002), 057403.
- [61] F. de Fornel, *Evanescent waves: From Newtonian optics to atomic optics*, Springer, Heidelberg, Germany, 2001.
- [62] D.E. Grupp, H.J. Lezec, T.W. Ebbesen, K.M. Pellerin and T. Thio, Crucial role of metal surface in enhanced transmission through subwavelength apertures, *Appl Phys Lett* **77** (2000), 1569–1571.



Andrey K. Sarychev is a Senior Research Scientist in the School of Electrical and Computer Engineering at Purdue University. He obtained his Ph.D. in Physics and Mathematics in 1979 from Moscow Institute of Physics and Technology. In 1990, he was decorated with the Medal of Honor for his investigations of microwave composite materials. In 1993, he was awarded the degree of Doctor of Science from Russian Academy of Science; in 2000, he was nominated for member of the Academy. A.K. Sarychev has worked as a Visiting Professor

in Tel Aviv University, Liege University, and the University of Provence in 1991–1995. During 1995–1998, he was head of the Theoretical Department at Institute of Theoretical and Applied Electrodynamics Russian Academy of Science. During 1998–2001, he was College Professor in the Physics Department at New Mexico State University. He has served as a Member of Program Committees for international conferences and presented numerous invited talks. Dr. Sarychev has authored 2 books and 4 book chapters, more than 100 original papers, and holds 5 patents. His current research interests include new optical and microwave materials and nanophotonics.



Vladimir M. Shalaev is a Professor of the School of Electrical and Computer Engineering at Purdue University. He is a Fellow of the American Physical Society and a Fellow of the Optical Society of America. Dr. Shalaev received his Ph.D. degree in Physics and Mathematics from University of Krasnoyarsk, Russia in 1983, where he then held an Assistant Professor position. In 1990 he was awarded the International Humboldt Fellowship and worked at University of Heidelberg, Germany, and University of Paris-Sud, France. In 1991–1993, he worked as

a Research Professor at University of Toronto and Ontario Laser and Lightwave Center, Canada. From 1993 to 1997 he was Associate Professor and, from 1997 to 2001, Professor of Physics at New Mexico State University (NMSU). Since 1997 he has been George W. Gardiner Professor of Physics at NMSU. He has served as a Chair and Member of Program Committees for numerous international conferences and held invited professorships in several leading universities in France, Hong Kong, and Germany. Since 2001 he has been a Professor of Electrical and Computer Engineering at Purdue University. Dr. Shalaev has authored 2 books and 10 book chapters, nearly 200 technical papers and conference presentations (50 invited), and edited 2 books. His current research interests include nanophotonics and optical nanomaterials.

Nanostructured Thin Films

Geoff B. Smith

Abstract

Nanostructured thin films containing metal or compound conductors are scientifically intriguing and technologically important. Their optical properties are linked to applications such as spectrally selective solar absorbers, solar control glazing, angular selective filters, optical biosensors and decorative paints. Film and coating morphologies discussed include nanograined thin metal layers; cermet and polymers containing metal, oxide and boride conductor nanoparticles; oblique nanometal columns in oxide; clusters and arrays of conducting nanoparticles; nanoholes in metal, and thin metal layers on nanostructures. Situations where quasi-static effective medium theories of optical response can be used, and those where they may be inadequate due to induced currents, including surface plasmon polaritons, are outlined. The wide variety of responses provided by coupling between surface plasmons in nanoparticle arrays to form new polarization modes is outlined. Very thin nanostructured conductor layers on dielectric nanoparticles are shown to allow broadband tuning as film thickness changes. Other advantages compared to metal particles are found. Metal nanoshells in practice exemplify a common attribute of nanostructured film systems, which often embody a hierarchy of nanofeatures that combine to determine final overall optical behavior.

1 Introduction

1.1 History and scope

Thin film- and coating-based techniques are crucial to many emerging nanotechnologies. Some nanostructured coatings are displaying properties that have not yet been explained or modeled, while some of the latest and exciting theoretical ideas cannot be put to the test until the required coating structures can be engineered. It is thus a vibrant, challenging and growing field. This review focuses on nanostructured layers containing conductors, which can be produced by vacuum techniques, from solution, by lithography and etching, by coating onto nanostructures formed by imprinting, and simply by depositing layers of nanoparticles. The nonvacuum techniques may include electrolytic and chemically deposited layers, polymer coatings, and dip or spin coating of dense layers of preformed nanoparticles, possibly aided by chemical modifications that induce self-assembly into patterns. Liquid crystal layers are not addressed in this review. Nanostructured films containing engineered arrays of well-defined nanoholes can also be produced simply in otherwise thin continuous layers after deposition, not only by standard lithographic etching but also by depositing onto a controlled planar array of nonmetallic nanoparticles that can then be subsequently removed by selective chemical etching or mechanically [1]. The etching technique is used in my laboratory.

It should be noted that even routine vacuum deposition techniques such as sputtering can produce nanostructures, depending on deposition conditions and materials employed. Resultant film material properties, such as apparent values of optical constants in some spectral ranges, thermal emittance, and resistivity, can differ significantly from those values expected for bulk materials. This may occur even if nanovoids occupy a small volume fraction and nanograins or columns are tightly packed. In practice, a hierarchy or multiple types of nanostructures often coexist in thin films. Thus, the properties achieved by forming deliberately engineered nanofeatures can be influenced by other incidental nanostructures in the film. For example, we may form after deposition, or during deposition, nanoholes 20 nm to 100 nm across in a thin electrically continuous metal layer. Overall response is quite sensitive to the electromagnetic properties of the material surrounding the holes. These properties are shifted from bulk values because of its 20- to 40-nm grains. Sometimes a small number of very fine voids around 0.5 nm wide might also be present. Another example is provided by isolated nanoparticles and nanocolumns that have influential substructure. For good quantitative fits to data, any models produced to describe the influence of *bigger* nanofeatures must take account of coexisting *smaller* nanostructures or other effects that make the pertinent material properties shift from bulk values. Specific examples come up later in this chapter, including structures in films on nano- and microparticles, and changes in energy-loss mechanisms. Nanostructural hierarchies can also develop under those deposition conditions where nanocolumns grow competitively [2].

Up to the mid-1980s, columnar nanostructure in vacuum-deposited thin films, often with associated voids, while of scientific interest [3]-[5] and common, was

generally regarded as a nuisance. The rapid uptake of ion-assisted deposition techniques in the period prior to and around 1990 was primarily motivated by the need to eliminate these columnar voids for stability and durability in precision optics. Also at that time, energetic deposition techniques, such as the cathodic arc and unbalanced magnetron sputter deposition, were found to give dense (approaching bulk density) arrays of nanograins and nanocolumns [6]. This led to the realization that there was value in some nanostructured films from a mechanical properties viewpoint (e.g., enhanced hardness and scratch resistance). A more recent development has been the use of nanocomposite films such as TiN nanosize inclusions in Si_3N_4 tissue (*nc*-TiSiN) to achieve super- and ultra-hardness, approaching the hardness of diamond [7]. This field of nanostructured films, like diamond films, may have much to offer optics in the future where robust surfaces are obviously of major value. A few of the nanostructured films we deal with are not robust and need protective layers which, in turn, play a role in the resultant optical response. Most, however, are quite strong.

1.2 Effective-medium models

In contrast to the 1980s, nanostructured thin films are now recognized as a source of many new and practical opportunities for applications, especially in the area of optics, but also mechanics, chemical and biochemical sensing and catalysis. Given their diversity, modeling is important; but unlike the world of classical thin film systems, current design tools have limited scope. Our interest in columnar thin films thus came not only from their potential for novel applications [8], but also because the new structures they provided gave an excellent testbed for development and critical analysis of effective-medium models. These are used for a description of effective optical response (or relative permittivities) of the total structure. In some metal-containing nanostructures, given recent developments, it is timely that the concept and zone of applicability of an *effective medium* be re-examined. The current controversies concerning the possibilities of effective mediums that may have an *apparent* negative refractive index exemplify the need [9]–[11]. Controversy has arisen because physicists in both supporting and opposing camps have treated these nanostructured materials as effective mediums in situations where a homogeneous representation of the composite material is questionable. In basic operational terms, if an incident plane wave onto an inhomogeneous medium produces to a good approximation a plane reflected wave in the far-field zone, then the medium can be represented by effective refractive indexes. If the medium transmits light, a plane transmitted wave should also arise. In conceptual terms, *the effective-medium representation describes how electromagnetic energy appears to traverse a composite medium rather than how it actually traverses the medium*. Thus, these uniform reflected and transmitted plane waves, at distances many wavelengths removed from the layer, appear to come from a homogeneous material. Intensities and phases of these waves can then be used with standard thin-film models to find

the effective optical constants. In the near-field zone and inside the composite material, energy flows become inhomogeneous but on a scale much smaller than the wavelength λ_0 in vacuum.

To highlight this, it seems that one needs to define the appropriate effective relative permittivity to use according to *where measurements are made* or, from an imaging perspective, the distance sources are located from a material. On treating the measurement or source point as (x, y, z) with $z - z_0 = h$, the height above the surface allows two or three types of dielectric functions to be defined for an isotropic layer between $z = z_0$ and $z = -t$. The traditional effective medium is with two components labeled by index $n = 1, 2$ given in functional form $\varepsilon^* = F(\varepsilon_1, \varepsilon_2, f_1, f_2, S_{n,n'})$ for $h \gg \lambda$, with ε_n and f_n the component relative permittivities and volume fractions, respectively; and $S_{n,n'}$, with all combinations of indexes n and n' of importance, representing all internal structural information that needs to be included to describe the particles and arrangements of elementary material components with respect to each other. This includes such issues as particle shape and layout of elements or particles, relative to other elements or particles of the same and different type. There is no explicit dependence on wavelength if the composite's surface is smooth and has approximate specularity. Energy in this zone is uniform in space and determined by ε^* . Very near and inside the material, however, energy distribution in space is influenced by the actual nanostructure according to either $\varepsilon^*(x, y, z) = F'[\varepsilon_1, \varepsilon_2, f_1, f_2, S_{n,n'}, (x, y, z)]$ if a quasi-static approach works or $\varepsilon^*(x, y, z) = F''[\lambda_0, \varepsilon_1, \varepsilon_2, f_1, f_2, S_{n,n'}, (x, y, z)]$ if dynamic effects are present due to induced currents or polaritons. F'' is therefore distinct in its explicit wavelength dependence, which can still be associated with specularity to a good approximation in the far-field zone when the composite has a surface that is sufficiently smooth. In the use of F to describe reflectance and transmittance data, effective or average boundary conditions can be used which are independent of location on the surface. In contrast, F' and F'' at $z = z_0$ and $-t$ give the actual fields and need point-by-point boundary conditions at the interface. A complete electromagnetic model or simulation will enable a prediction of how F' and F'' at $z - z_0 = h \leq \lambda_0$ vary with x and y . Its measurement or mapping in the various (x, y, h) planes is needed to test predictions and is directly relevant to some emerging nanotechnologies. A near-field scanning optical microscope in collection mode with the tip at height h above the surface can, within the spatial resolution dictated by its aperture size, sample these field profiles.

Thus, for any engineering ideas based on near-field properties, such as evanescent waves (e.g., the superlens), use of the far-field effective-medium model seems inappropriate, but neither can *effective left-handed* materials be dismissed on the basis of what would happen internally or at the interfaces in a homogeneous left-handed medium. Such data as specular transmittance, specular reflectance or the ellipsometric parameters Δ and Ψ obtained from field amplitude ratios and phase differences can always be inverted to obtain apparent refractive indexes. It is the task of an effective-medium theory to link these far-field results to the complex breakup of electromagnetic energy within and on a nanostructure. Traditional quasi-static

models did this by calculating an average internal electric field in terms of actual fields induced within different components. In this chapter, we cover examples where this works well and other examples where it fails so that new approaches are required.

While it has been known how to deposit oblique columnar grains for some time, it has also been shown lately that oblique voids in otherwise continuous films are interesting and can be engineered directly during or after deposition. Anisotropic effective-medium theories were essential for modeling columnar films and voids; and soon followed [12], enabling important new ideas [13]. Models depend on the detailed topology in the nanostructure, and the few simple models available are often applied uncritically. We show here that apparent fitting of data can be misleading, since the most widely used models have generic mathematical resonance features that commonly allow fitting. That is, unless the models used have their fitting parameters shown to be consistent with observed topological and composition data, fitting must be interpreted warily. Another check on the validity of the simpler models is to compare their predictions with experimental spectrums across a bandwidth that is much broader than the immediate wavelength range used in fitting, which is quite commonly limited to wavelengths around a single resonance. For instance, additional multipole resonances might not be observed because they lie at ultraviolet wavelengths, while experimental data is confined to just the visible or the dominant solar energy wavelengths (visible and near-infrared). In Section 3 it is shown that the fitting of data over a limited spectral range in a single resonance model can be understood when other bands are present, provided the fitting parameters are reinterpreted.

The most common approaches to effective-medium models assume quasi-static fields. Recent developments in various nanostructured metal [14] films, metal particle cluster arrays [15], and lattice structures with photonic band gaps show that smallness of features relative to wavelength alone cannot always be used to justify the quasi-static assumption, and even basic Mie theory [16] may be inadequate. That is, dynamic models with full wave character and exact local structure may be needed at times, even when features are much smaller than the wavelength of impinging radiation. The underlying reason is that charge excitations such as extended or localized current excitations may be induced. These include polaritons, which are waves involving coupling of photons and extended charge oscillations, here called surface plasmon polaritons (SPPs). If the structure allows them to form, these new hybrid waves can have many wavelengths between the wavelength of the incident light wave and the spacing between electrons. SPP waves may be damped by emitting radiation and by electron collisions. These collisions can occur with surfaces, and possibly by nonlocality of the mobile electron response. Surface interactions and nonlocal response are more likely to occur in nanostructures. We also cover situations where the traditional quasi-static models are well justified and provide excellent models for technology development. Localized *static* surface plasmon (SP) excitations occur in the quasi-static limit, commonly on isolated metal nanoparticles. Quasi-static models also imply no significant scattering.

Scattering onsets and the breakdown of quasi-static effects with size for isolated particles are well understood using Mie type models and should be checked if in doubt. Extinction by particle absorption will be much larger than that due to scattering in the quasi-static limit.

Nanostructured films have become such a wide-ranging, fruitful area of growth in the last decade that this review needs to be limited in its scope. It does not cover nanostructured systems that contain only dielectric material, such as sculptured thin films [17] (though such structures should also be possible with refractory metals). It focuses on those nanostructured layers and coatings where plasmon and surface plasmon effects enabled or enhanced by the nanostructure provide the basis for the properties of interest. *Thus, materials with negative relative permittivity ε must be present. Bulk materials without structural defects and with $\varepsilon < 0$ have high reflectance and propagate light internally mainly by decaying evanescent waves.* These are mainly either conducting materials or ionic materials in their Reststrahlen band [18]. Nanostructure in the latter may enable new infrared technologies and is ripe for study, but we focus here on conductors. The nanostructured topologies of interest include isolated and dense arrays of conducting spheres and spheroids embedded in dielectric material, metal columnar arrays, granular metal films, thin metal films containing cylindrical voids, layers of nanoparticles, and thin metal films on nanopatterned dielectric substrate. We touch on a new class of nanostructured film that has recently come into the spotlight. It has the potential to impact general optics for any region of the spectrum, from ultraviolet to deep infrared [19]. It comprises very thin layers of a conductor on a dielectric nanoparticle. If such coated particles form a film then there is a three-level hierarchy of nanostructure to deal with: the internal structure and thickness of the particle coating, the coated particle itself, and the array of coated particles. The latter is seen to behave differently than all-metal particle arrays. Continuous and discontinuous dielectric layers 20 nm to 150 nm thick on very thin metal flakes and metal films are also of interest. These can provide multicolored energy-efficient paints [20-21] and new surface plasmon effects.

2 Nanostructured films containing conductors: an overview

A summary of the main issues for different generic classes of nanostructured films defined by component conducting particle topology appears in Table 1. It contains, for various generic divisions, relevant theoretical physical models for optical properties along with the author's view of their maturity. Production approaches and applications are also given.

The great geometric diversity and, hence, variety of optical responses that the nanoscale brings is illustrated in Figure 1. It shows the evolution of individual nanostructural elements that can appear in films and that need unique optical response models. The various particle arrangements include isolated random arrays and clusters; a close pair which may also touch or intersect; a homo or hetero pair (in size and material); bigger clusters of these subtypes including chains and

Table 1 Classes of nanostructured films and layers with relevant optical models, physics and applications.

Generic nanostructure	Theoretical models/physics	Model Status	Production approaches	Example Applications
Very dilute (<0.5% to 1%) nanoparticles in dielectric matrix	Noninteracting spheroid, isolated SPs	M	Codeposition polymer foil extrusion	Glazing, filters (UV, laser, color)
Dilute/dispersed (2% to 25%) nanoparticles in dielectric matrix	Generalized Maxwell Garnett MG, weak coupling	M	Vacuum co-deposition, coating solutions	Solar selective absorbers, nonlinear materials
Columnar, oblique columnar metals plus voids or dielectric	Isotropic, Anisotropic MG or Bruggemann	M, D	Oblique co-deposition	Angle selectivity, sensors, catalysts
Granular metal films, metal films with nanoholes, oblique voids in metal	Coupled SPs across film, cavity resonant SPs, polaritons, generalized Ohm's law	D, N	Vacuum deposition of very thin films, lithography and etching	Strong IR transmission, nanoscale imaging, lighting, sensors
Thin metal layer on dielectric nanostructure	Thickness-dependent SP resonance	M, D	Deposition on polymer imprint or nanoparticle array	Bio and chemical sensors, color tuning
Dense nanoparticle arrays, clusters, coatings (nontouching, touching, intersecting)	Multipoles strong coupled SPs, lattice models, photonic band gaps	M, D, N	Particle vacuum coating, solution dip, dense self assembly	UV and solar control, nanoscale photonics, (waveguides, antennas)
Thin metal on dielectric nanoparticles (arrays and clusters)	Coated spheroid, damped SP coupling between particles	M, D, N	Chemical coating onto seeded particle	Wide-range narrow-band tuning with two materials, chemical and biosensors
Thin dielectric on metal flake, metal nanoparticle, nanohole	Classic thin film if smooth	M	Chemical vapor deposition	Energy-efficient paint, color vs angle
Nanostructured dielectric on very thin metal	Enhanced SP generation and cross film coupling	D	Oblique deposition, or etching	Induced transmittance, novel metal nanooptics

Abbreviations used in Model Status column: M = mature, D = under development/some progress, N = recent new phenomena, insights. Two or more letters refer to different subclasses of these nanostructures (SP = surface plasmons).

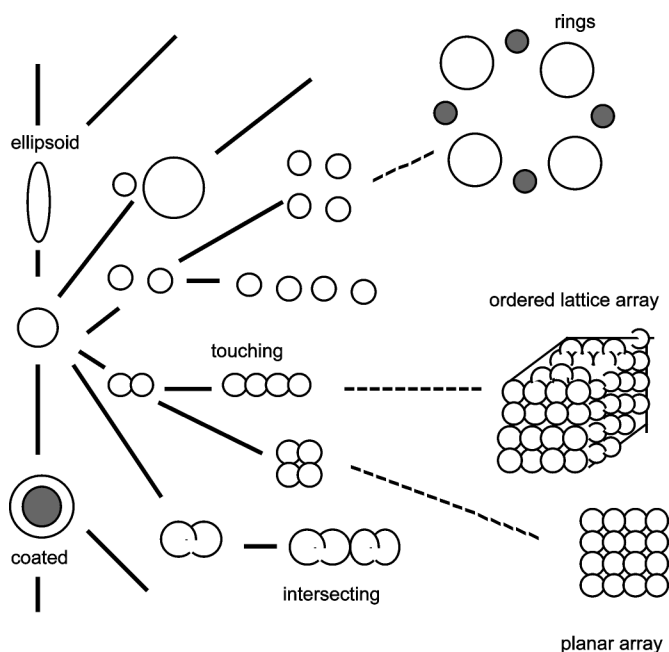


Figure 1 Nanoparticle, clusters and arrays: families with unique plasmon character and optical responses.

rings; percolating arrays in two and three dimensions; and finally, complete planar and solid lattices of spheres that can touch or involve gaps. The branches shown for spheres can be duplicated for each individual type of particle, such as the ellipsoid and metal-coated dielectric particle shown, and for columnar voids. Ellipsoidal particles and sphere chains have optical anisotropy and ellipsoids may align obliquely to give angular selectivity. Select models for elements in Figure 1 are discussed later. *All arrays will support surface plasmons, but only some will support polaritons.* Dielectric layers above or below nanostructured films impact strongly the coupling of radiation to the system modes. The matrix material surrounding nanoparticles has a similarly strong influence. Thus, nanostructured metal films and metal nanoparticle arrays tend to have their special properties amplified by dielectric overlayers or surrounding matrixes. Production and modeling research problems are thus almost limitless.

Theory is needed with such diversity to help sort out the benefits of specific substructures and make the development process more efficient. With a few important exceptions, theory has not led the way in most recent advances. Simulation is commonly used when theory cannot handle all structural issues. It is useful in exploring possibilities, but without basic understanding it too can be just as inefficient as empirical experimentation. Figure 2 shows just a few examples of actual nanostructured thin films produced in my laboratories over the past decade.

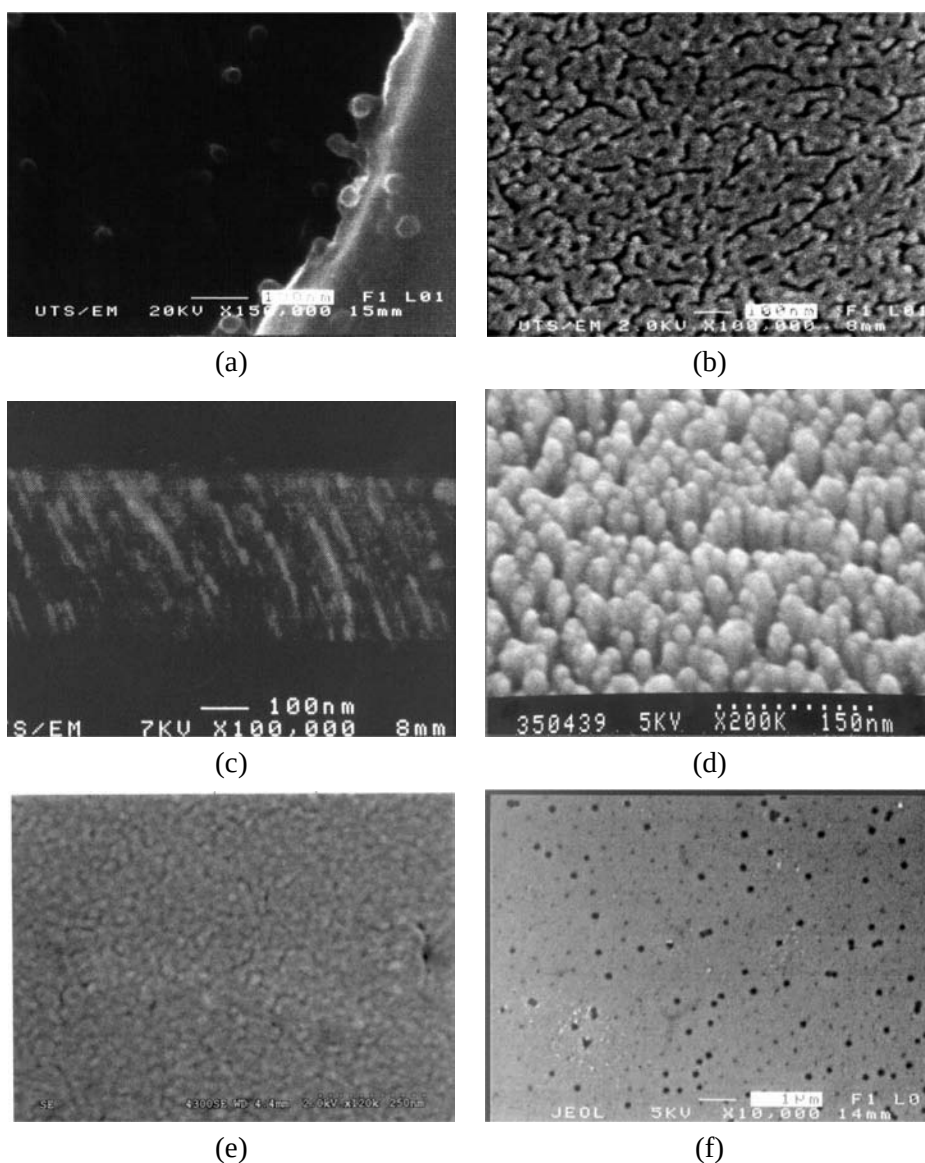


Figure 2 Examples of nanostructured thin films containing conductors: (a) dilute antimony tin oxide (ATO) nanospheres in polymer film; (b) oblique granular Ag metal; (c) Ag metal columns embedded in SiO_2 ; (d) Ta metal nanocolumns; (e) 9-nm-thick Ag film with nanograins; and (f) random round nanoholes in the metal film in image (e).

3 Thin films containing nanoparticles

3.1 General issues

To understand complex nanostructured thin films, it is best to begin with an understanding of the least complex cases, which quickly become nontrivial. We focus

for simplicity on the effective-medium limit where materials appear to be optically homogeneous in terms of their far-field behavior. In that case, far-field data can be described in terms of uniform refractive indexes. Scattering or haze must be negligible or very weak for optical homogenization to be used. Thus, theoretical scattering studies can be used to establish the bounds on structure for validity of an effective-medium approach. Three generic classes of effective models need to be identified, one being relatively new. The first two are quasi-static limit models. This limit was believed to apply whenever characteristic sizes D of particles, and other features such as voids or clusters, are such that $D/\lambda_0 \ll 1$. As noted previously, the latter is now a necessary but insufficient condition. First, there are dilute separated-particle systems, where single-particle Mie-type scattering models apply *and wavelike polaritons cannot form*. The quasi-static limit emerges simply. Next, interparticle interactions occur but gaps are still too large to support polaritons. Quasi-static models can still apply but their validity criterions need to be modified. This is the second class of effective mediums. The *random cell* concept [22], which incorporates average local structure and spacing (discussed further in Section 3.4), can be used in this case. Consideration of scattering off such a cell has allowed validity criterions [23-24] to include average spacing. The third and final class involves induced currents, including surface plasmon waves or polaritons.

Experimental lack of significant scattering is now necessary but not sufficient to justify quasi-static modeling. Until recently, one might have assumed it was sufficient, but in those classes of nanostructure that support surface extended currents, such as plasmon polaritons, dynamic models appear to be unavoidable. This is the case even if data on the far-field profile indicates weak or insignificant scattering. *That means a non-quasi-static effective medium can occur, which involves averages over fields associated with induced polarization waves or currents, that have characteristic spatial periods larger than the finest structural features but much smaller than the wavelength.* Such systems include touching spheres [15], narrow tapered grooves [25], simple nanoholes [1], [14] through metal, and very thin dense metal films with nanofeatures. These all support SPP waves which may have wavelengths much smaller than the exciting optical wavelength. A new SPP mode can arise from coupling of excitations on either side of nanogaps in metal, or on opposite surfaces of thin metal films. Such features will not appear in classical quasi-static models that can however embody localized surface plasmons. *Thus, mathematically exact quasi-static solutions may for some nanoscale geometries never agree with dynamic models.* The touching-particle case detailed in Section 4 may exemplify this. New types of effective-medium models, and hence effective relative permittivities, can still apply if the far-field averages out all local near-field and current fluctuations to reproduce a specular plane wave. Experimental data often implies this averaging has happened. This is a relatively new field, but progress and simple expressions would be of great value to experimenters. One approach has been a Generalized Ohm's Law [26-27]. A feature of this new class of effective optical constants will be their dependencies on the angles of incidence and

polarization sensitivity. Light coupling to SPPs in percolating particle arrays or on metals with isolated hole arrays should depend on the incidence angles.

3.2 Polarization eigenmodes in arrays

In quasi-static models, the natural polarization eigenmodes of an arbitrary multi-particle nanostructure can be very different to those of the multipole modes of a constituent isolated particle. *The fundamental system eigenmode is not an array of simple dipoles, once two or more modes are present.* Exact quasi-static solutions for two simple clusters appear in Figure 3 [28]. Potentials can, if desired, be expanded as an admixture of isolated particle modes (dipole, quadrupole, etc.) on each particle. For example, in any sphere array the potential outside sphere l can

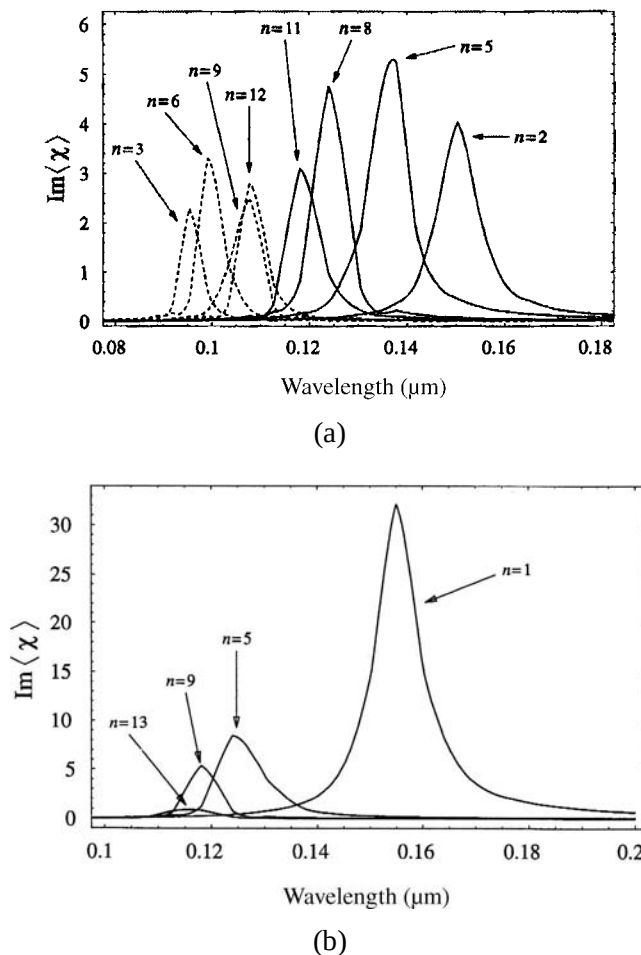


Figure 3 Exact polarization modes for clusters of three aluminium cylinders with a smallest gap of $0.2R$ (a) equilateral triangle (b) chain. Mode order $n = 1$ has zero spectral weight for the triangle, and $n = 9, 12$ are symmetric modes close to the doubly degenerate isolated cylinder mode.

be expressed in the form

$$\begin{aligned}\Phi_{out}^{(l)} = & V_{ext}(\mathbf{R}) - \sum_{nm} B_{nm}^{(l)} |\mathbf{R} - \mathbf{R}_l|^{-n-1} Y_{nm}(\mathbf{R} - \mathbf{R}_l) \\ & - \sum_{j \neq l} \sum_{nm} B_{nm}^{(j)} |\mathbf{R} - \mathbf{R}_j|^{-n-1} Y_{nm}(\mathbf{R} - \mathbf{R}_j),\end{aligned}\quad (1)$$

where the first term is the potential of the external field, the second term is the potential created by particle l at the point $\mathbf{R}$, and the third term is the potential created by the rest of the particles at this point. $Y_{nm}(\mathbf{R} - \mathbf{R}_l)$ are the scalar spherical harmonics with n as the polar order. The expansion coefficients B_{nm} depend on the array geometry and must be solved to determine the effective dielectric response. Such a mathematical decomposition is common in computational work but clumsy when interactions are strong. In that case, the final modes have little in common with these starting functions, and numerical accuracy requires high-order sums for an accurate solution. If the fundamental eigenmode spans a wavelength range in isolation, as in the example in Figure 3(b), it can easily be confused for a classical Maxwell Garnett (MG) or Bruggemann (BR)-type dipole if the higher-order modes are not picked up experimentally. However, if it is isolated, the fundamental mode alone gives the additional polarizability and hence dipole moment per unit volume across a limited frequency range. The existence of modes at other frequencies with nonvanishing spectral weight (for equations, see Section 3.3) means that the weight of this isolated fundamental is shifted from that of a dipole (the volume fraction). Thus, fitting parameters may need to be reinterpreted.

3.3 Isolated nanoparticles

The origin of strong surface plasmon resonance in isolated metal nanoparticles is well known but worth reiterating, as it also embodies key issues pertinent to more complex systems and to practical optical engineering in metal nanostructures. The combination of a negative real part of the relative permittivity and a particle geometry in which the induced internal field is constant (this only applies to spheres and spheroids) creates conditions for a single-dipole surface plasmon resonance. For a sphere with complex relative permittivity $\varepsilon_m = \varepsilon'_m + i\varepsilon''_m$ embedded in a medium of (real) relative permittivity ε_h this can occur only if the equation

$$\varepsilon'_m = -2\varepsilon_h \quad (2)$$

can be satisfied. For a conductor in a spectral region sufficiently below interband transition peak frequencies such that their residual impact is finite but almost independent of frequency, one can write

$$\varepsilon_m = \varepsilon_\infty - \left[\frac{\omega_p^2}{\omega(\omega + i\omega_\tau)} \right]. \quad (3)$$

Therein, ω_τ is the relaxation frequency, ε_∞ accounts for the residual effects of the high-frequency transitions, and

$$\omega_p = \frac{n_e e^2}{\varepsilon_0 m^*} \quad (4)$$

is the plasma frequency. The symbol m^* represents the effective mass of electron carriers, and n_e their density. The central material parameter is n_e since it allows control of ω_p which we need to do for controlled spectral tuning such as is required in glazing applications (Table 1). Thus, two key frequencies depend on ω_p . First is that frequency at which ε'_m actually becomes negative, $\omega_{p,scr} = \omega_p / \sqrt{\varepsilon_\infty}$, called the screened plasma frequency, which is usually significantly lower than ω_p . It is the frequency below which bulk conductors start to reflect strongly. The second is the surface plasmon resonance frequency given by (2) and (3) as $\omega_{sp} = \omega_p / \sqrt{\varepsilon_\infty + 2\varepsilon_h}$. A higher-index matrix can thus shift the SP resonance to longer wavelengths. It is easily shown that the resonance's absorption width is proportional to ε''_m and its peak intensity is proportional to $1/\varepsilon''_m$. Both of these quantities are governed by the relaxation frequency ω_τ . Therefore, a higher relaxation rate will broaden the resonance and lessen its peak intensity. Metal films with nanocrystalline grains can be described simply by (3) with an ω_τ enhanced over bulk values. This enhancement arises from electron collisions with grain boundaries since the grain size is smaller than the bulk mean-free path. A nanoparticle material or the metal component in a film containing nanoholes will require use of such modified optical constants to get good overall fits to data if nanograins are present. This is a simple example of hierarchical nanoeffects. It is usual in nanoparticles to increase the relaxation rate due to collisions with particle surfaces if it is faster or comparable to either typical bulk collision frequencies or the modified bulk values just introduced. These simple equations derived for spherical particles have been extended to describe light coupling to various nanostructures, including dense arrays of particles on a surface [29]. That treatment was based on observation of apparently isolated absorption modes, and data fitting was then based on resonance equations for an isolated sphere. However, fitting required some juggling of model parameters, which was not easily explained. This appears to be an example of a spectrally isolated plasmon resonance mode, with a spectral weight shifted from that of a simple dipole. Other modes must be present in such a dense system, as noted above. The erroneous treatment of isolated SP resonances in multimode nanostructures as simple dipole resonances may be quite common.

Intensive study of nanoparticles embedded in dielectric films on metal substrates occurred in the 1970s and 1980s (see a review from 1983 [30]). It was then found to be the best way to engineer solar selective absorbers that require high absorptance from 400 nm to 2500 nm and also a sharp transition to low film absorptance in the black-body range, so that the low substrate emittance then comes into play. Absorptance (A) is the quantity that is extracted directly from spectrophotometric measurements of reflectance (R) and transmittance (T), and it is the fraction

of incident energy that is absorbed; so $A = 1 - T - R$ with $T = 0$ for opaque substrates. Two leading commercial electrodeposited films for solar collectors, back chrome and nickel pigmented anodized alumina, have complex nanostructures that produce high absorption efficiency in a well-defined spectral range provided by nanoparticles of chromium and nickel, respectively, which are denser towards the bottom of the film. Vacuum-deposited nanometal cermet films were developed for the same purpose but were mainly used in evacuated tubular collectors, for higher-temperature concentrator use. In recent years such surfaces have been used for solar power stations utilizing parabolic-trough concentrating collectors, and have been refined by using double layers of different metal nanoparticle density to maximize absorption, while providing a sharper spectral transition near $\lambda_0 = 2.5 \mu\text{m}$ [31]. The substrate metal, which provides low emittance, is usually also deposited, and should be *engineered to avoid nanostructures* (which often occur), as they can raise emittance by 100% or more. That is, they reduce infrared reflectance from the 92% to 96% range to 84% to 90%. Two factors cause this shift in emittance in metal nanostructures: first, infrared (IR) radiation couples more readily into nanostructured metal layers that are not flat (via SPPs); and second, relaxation rates and hence absorption losses are higher due to scattering of electrons at nanograin boundaries or surface protuberances.

It is important to note that *deep blacks* can occur with quite low metal particle content, because light at ω_{SP} resonates strongly with the particle's dipole surface plasmon. *The absorption cross-section is much larger at resonance than the geometric cross-section of the particle.* Much of the resultant local buildup of fields and energy is dissipated as heat (i.e., not as the alternative decay process for SPs, which is by radiation) with charge carrier scattering, thereby damping the excited surface plasmons. The high efficiency and high cross-section near resonance of surface plasmon particle absorption in isolated conducting particles, combined with its well-defined spectral range defined by n_e , is the key to their usefulness since very low densities of particles can provide required absorption strengths. Consequently, high performance can come at low cost, even if exotic expensive particles are needed. The weaknesses of this approach are the limitations in being able to tune or shift the resonance in a controlled manner and in the ability to easily control the breadth of the resonance and the peak cross-section. Gold nanoparticles in a low-index matrix material do not produce black but transmit red. Gold blacks as thin layers, however, have a long history [32-33] but rely on the more complex mechanisms that occur in clusters and strings of nanoparticles, which are not so easy to manage for applications, and are only now being fully understood. Clustering raises many new problems and opportunities, some of which are very recent [15]. To tune at low density with a single material nanoparticle in a defined medium requires changing the particle material to shift the carrier density, or changing the particle shape. These are the only parameters that define the resonance at densities where interactions between particles are weak.

Vacuum-based co-deposition of metal or compound conductors, with a dielectric material, or the creation of layers from a polymer or monomer solution that

contains nanoparticles of these materials, can provide the desired nanocomposite layers. Use of nanoparticles made from compound materials with lower conductivity than noble metals, due to reduced carrier density n_e , such as indium tin oxide (ITO) and antimony tin oxide (ATO), shifts ω_{SP} to wavelengths near $1\ \mu\text{m}$ or longer. These nanoparticles have some use in solar control glazing, when combined with certain standard glasses for blocking solar heat but not light [34]. Better results are obtained using particles of other materials with ω_{SP} in the crucial solar near-infrared (NIR) range of 0.75 to $1\ \mu\text{m}$. Examples are currently under investigation in my laboratories, and a particular one suggested by our work [35], LaB_6 , is now available in polymeric form for windows, since this technology is very easy to implement. Quasi-static effective-medium models for isolated dipoles on spheres with no interactions between particles give excellent NIR data fits for spherical nanoparticles, since the densities needed are very low ($<0.5\%$ by volume) in the thicker films and the polymer foils that are used for windows. Then, both clustering and interactions can be neglected. An example of a laminated window optical transmittance spectrum, which simple theory can model almost exactly, is shown in Figure 4 [34]. It uses low-density ITO and ATO.

This is the strength of plasmon resonance effects: with layers or foils in the 100 to $700\ \mu\text{m}$ thickness range, densities of nanoparticles can be low enough that interparticle effects may be neglected, and the attenuation coefficient is then linear in the particle volume fraction.

3.4 Increasing density and clustering effects

When density increases to the point where polarized particles interact, the optical response can be modeled simply, provided there is no clustering. To treat random

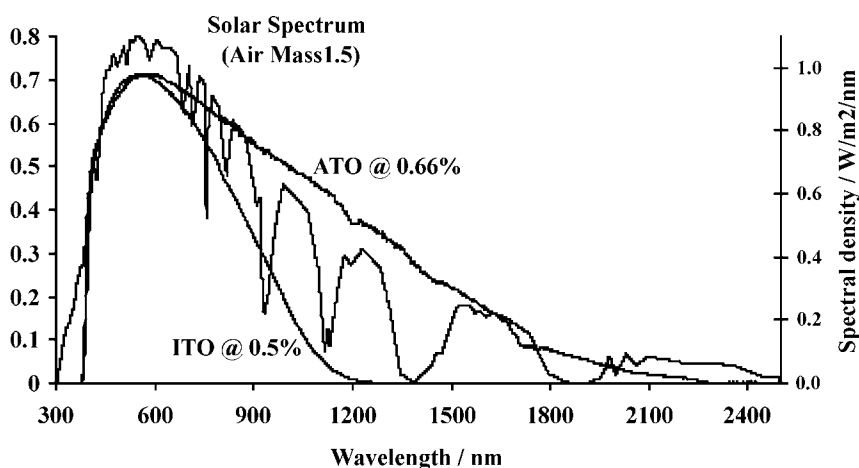


Figure 4 Spectral solar control with nanoparticles of ATO and ITO in a clear low-haze laminated glazing. Volume % is shown. The average solar spectrum at the earth's surface is shown for reference.

arrays of spheres in which only dipoles exist on each particle and interact, the concept of a *random unit cell* [22], [24] was introduced. This can include most commonly used types of effective medium models, including Bruggemann, Maxwell Garnett and several others. Embedded in the effective medium of relative permittivity ε^* , this cell produces no additional polarization. The resultant expression provides a self-consistent solution for ε^* .

Arrays of spheroids are often useful to model columnar structures arising in thin films. Spheroids require care since inappropriate effective-medium models are often applied to ellipsoid arrays. Columns are intrinsically anisotropic in polarization and tend to form with a preferred orientation. This can be useful but must be modeled carefully. It was recognized long ago [36] that some data on structures containing ellipsoids in effect required an effective-medium *cell* with a boundary that is not spherical. Only when the particles are randomly oriented or the density is very dilute does a spherical cell apply. Until the recent work by Skryabin *et al.* [37] there were various approaches, but no rigorous basis for choice. The approach adopted in some common software is physically incorrect. In common with many models, these programs use a cell of the same shape (and hence a depolarization factor L) as the particle. This actually creates multipoles. For a cell that only supports a dipole, the particle and shell boundaries must both be equipotentials, and that requires confocality for ellipsoids. Only then does a simple dipole solution arise for ε^* . Since the *shell* volume within the cell represents the volume fraction of the host, a procedure linking volume fraction f and cell shape had to be established [37].

The generalized expression that can be solved for the effective relative permittivity ε^{MG} for one polarization direction if ellipsoids are aligned is

$$\frac{\varepsilon^{\text{MG}} - \varepsilon_h}{\varepsilon_h + L^h(\varepsilon^{\text{MG}} - \varepsilon_h)} = f \frac{\varepsilon_m - \varepsilon_h}{\varepsilon_h + L^m(\varepsilon_m - \varepsilon_h)}, \quad (5)$$

where L^h is the depolarization factor of the cell or envelope and L^m that of the inclusion or particle. ε_h and ε_m denote host and particle relative permittivities, respectively. L^h is a function of f . L^m applies to a specific orientation of the particles, and each orientation has been used along with transforms to the film coordinate frame, to successfully model oblique columnar films [13]. If other orientations arise, then the right side of (5) must be averaged and the shell shape will change. When both L^m and L^h equal $1/3$, the classical MG result follows. The expression on the right side of (5) leads to a modification of (2) which was for the noninteracting particle's resonance position. The SP resonance frequency ω_{SP} for a generalized isolated spheroid is found from the frequency that satisfies

$$\varepsilon'_m = -\varepsilon_h \left(\frac{1}{L^m} - 1 \right). \quad (6)$$

An increasing shift of ω_{SP} to long wavelengths as L decreases is apparent. Upon solving (5) for ε^{MG} at a sufficiently large f so that interactions between

particles impact on the resonance frequency, it is then easy to define ω_{SP} as a function of f and $L^h(f)$. In this weak interaction dipole approximation, the isolated plasmon resonance is shifted in energy but does not split into symmetric (S) and antisymmetric (AS) modes, since the model only has one mode if the spheroids are oriented. Once interactions induce multiple resonances, S and AS modes can appear. In all exact solutions or higher-order approximations this splitting is apparent, with the symmetric solution at an energy level above that of the single-particle degenerate mode. It is then simple to qualitatively understand what occurs as particles cluster and move closer together. The lowest-order AS modes shift to increasingly lower energies as particles approach each other, since charges of opposite sign are then also closer together. All plasmon modes are incorporated in the general expressions [38] of (7), which has two common forms. We find the second form most useful for checking symmetry and when separating high- and low-energy modes; it is given by

$$\varepsilon^* - \varepsilon_h = \sum_n \frac{g_n}{[1/(\varepsilon - 1)] + L_n} = \sum_n \frac{g'_n}{v + t_n}, \quad (7)$$

with $\varepsilon = \varepsilon_m/\varepsilon_h$, $v = (1 + \varepsilon)/(1 - \varepsilon)$ being the materials related part, and $t_n = 1 - 2L_n$, $g'_n = g_n/2$ being the structural part. The $n = 1$ mode is the fundamental mode, but in some structures can have zero spectral weight. Examination of t_n is useful as the two signs it can take indicate which is symmetric and antisymmetric [28]. It embodies the broken symmetry that occurs once interactions lift the degeneracy in a single cylinder or single sphere. Examples in two dimensions [28] (long cylinders) show t_n comes in pairs with equal magnitudes but opposite signs; this grouping constitutes a symmetric and antisymmetric pair. In two dimensions, each t_n pair has two different values of L_n which add up to unity. A tendency to cluster at any density can occur with nanoparticles in solution or vacuum particle beam techniques, due to van der Waals forces. Such low-density clustering, and the close approach between particles intrinsic to a high-density distribution, both open up the need to include higher-order particle multipoles and multiple resonant modes in the calculation, since several g_n are finite for this situation.

The lowest-order modes have the largest energy shift from the isolated-particle SP energy, and the higher-order modes cluster around the isolated particle wavelengths. The eigenvalues v_n of each mode, which define the resonance frequencies, are given by the poles in (7), and their structure is also necessary to give the correct response when the field direction changes by 90° from any symmetry axis. An array of spheres or ellipsoids has groups of three modes that are degenerate when the spheres are sufficiently far apart. However, upon coupling they split, typically into one AS and two S modes, with the trio of t_n values again totalling 0 and the associated L_n 's again summing to 1. Cluster symmetry is clearly important.

Many two-dimensional (2D) arrays are open to efficient conformal-frame mathematical techniques that lead directly to the normal modes, and often to the analytic solution of potentials. For three-dimensional (3D) arrays (including 2D layers of spheres or ellipsoids), the equivalent approach has been limited to special

separable coordinate frames for spheres, including separate sphere pairs (bispherical) [39-40], touching spheres [41] (tangent sphere), and intersecting spheres [42] (toroidal frame). For other arrays (even though 2D analogous hypercomplex variable techniques are possible [43]), it is usual to resort to potential expansions based on the isolated particle. Spherical harmonic expansion such as (1) or other orthogonal basis functions are used. The experimental bispherical analysis of plasmons in aluminum particle arrays by Batson [40] shows that, for nanoparticle arrays of metal spheres, the lowest-order pairwise modes dominate. Having these exact quasi-static solutions is useful to establish if polaritons are present. If data diverge from exact quasi-static results, then polaritons or other extended current effects within clusters are probably occurring and altering the nature of the effective medium. Without exact models such conclusions were not possible, at least without a detailed analysis of induced charge dynamics. With today's computer power, multipole expansions are increasingly used; and for simple clusters one can use them to go directly to the dynamic field calculations for exploration of these issues [15]. It will, however, be a while before they become an everyday design tool for nanostructured layers or for use in the inverse least-squares fitting problem to find component optical constant and structural data from optical spectrums. That is, determining nanostructure, indexes, and concentrations from optical data the way we can now do with MG and BR effective-medium models in widely used ellipsometry and spectrophotometry software, requires new and better models to handle polariton-based effective mediums. *We can, however, now flag those structures where exact quasi-static models may fail.* One of these structures is discussed further in Section 4.

What is interesting to note is that experimental evidence for distinct multiple modes has rarely been seen to date. Two possible causes are (i) a wide mix of local structures produces many overlapping plasmon bands as observed in granular metal films, or (ii) metal particles that touch and so produce a continuum of modes (see Section 5). This situation may change as layer and particle-array nanoengineering techniques improve. For instance, self-assembly with precision monothickness dielectric coatings on monosize metal nanoparticles may yield the desired single class of modes. Some complex arrays do lead to apparently simple modal structures. Examples include collinear arrays of a few to a very large number of particles, and some other structures [44], where we have found that the $n = 1$ mode is usually well isolated. Such a situation could easily again lead to the mistaken belief that it is the MG dipole mode that is being observed if experimental wavelengths were limited in range.

4 Metal thin films on dielectric nanoparticles and nanostructures

Coating dielectric nanoparticles or microparticles with nanothickness metal layers creates interesting new plasmon features and, hence, spectral responses. The most

important of these is the ability to tune particle SP resonances across a wide spectral band, just by varying layer thickness. An *ideal* example in the quasi-static limit, with very dilute Ag-coated titania particles with Ag of varying thickness embedded in a polymer layer, is shown in Figure 5. *Ideal* means bulk metal relaxation rates, which may be difficult to approach except maybe for layers on larger particles (where quasi-static models will not apply). Much data reported to date has been for coated nanoparticles in a water matrix. These systems have been optically hazy due to extinction by scattering adding to that from absorption. For most glazing applications haze is not acceptable. Fortunately, scattering is much weaker in the higher-index polymer matrixes used in window applications (typically, clear polymers have a refractive index in the range 1.49 to 1.6, whereas that of water is 1.33). If scattering is present it can often be confined to the NIR when the resonance is located there. The polymer matrix also shifts the value of ω_{SP} to significantly longer wavelengths than its location for a water matrix. These coated particles were a theoretical curiosity until recent progress using chemical methods [19], [45]. Many applications beckon, but the ability of polymer layers containing such particles to compete with multilayer thin films, for instance, depends crucially on the quality of the metal layer. The chemical methods generally create nanorough surfaces and start with very small (~ 2 -nm) grains. The combination of a very thin metal shell thickness and its nanoscale roughness results in much enhanced carrier relaxation rates relative to the bulk metal. Hence, broad spectrums with very much weaker peak intensities compared with the ideal ones of Figure 5 have been found to date.

Progress in reducing the relaxation rate is crucial and should result from more deposition research and from the use of larger particles. Important advances in obtaining more uniform metal shells have recently occurred [46]. Theoretically, these

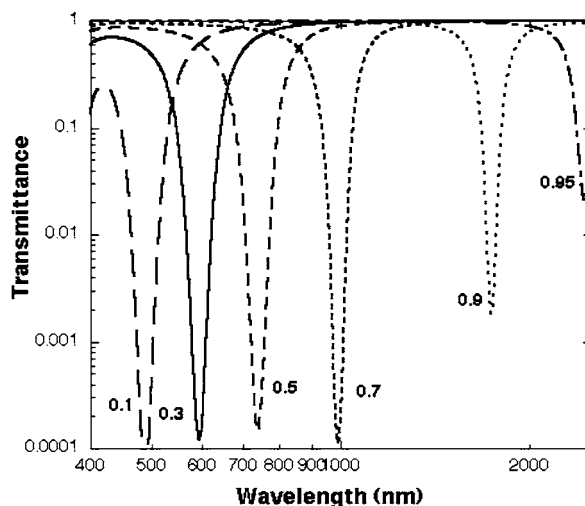


Figure 5 Modeled quasi-static transmittance of a laminated window with the 0.7-mm laminate layer containing 0.05% by volume of *ideal* Ag-coated titania nanoparticles for a range of silver thicknesses. Associated core volume fraction appears by each curve.

systems have three novel features, which are now briefly outlined. Firstly, the isolated particle has nondegenerate modes due to internal SP coupling across the thin metal layer. Another feature is that the dielectric core has a much enhanced field compared to that in a metal or bare dielectric particle. One key new feature is found in clusters [47] of such coated particles: they appear to behave very differently for most metal layer thicknesses compared to clusters of all-metal particles; in particular, the plasmon resonance wavelength and structure are far less sensitive to particle spacing.

Interesting spectral effects as functions of metal layer thickness on sub-100-nm dielectric nanofeatures, such as occur on nanoimprinted polymer surfaces or a densely packed layer of polymer nanoparticles, may be commonplace. For instance, when nano-thin metal films were deposited onto a continuous dielectric nanostructured surface formed by dense packing of dielectric nanospheres [48] into layers, different colors arose as the particle size in the dielectric layer was altered while metal thickness was fixed. These metal layers were vacuum deposited.

5 Dense arrays, clusters touching particles

Metal nanoparticles that touch or intersect have qualitatively different quasi-static spectral properties to those given by (7). Discrete spectrums become a continuum, so the sum of spectrums becomes an integral [41-42] and broad bands arise. Expansion techniques will thus not easily converge. The analytic result for touching pairs, which is mathematically correct and exact in the quasi-static limit despite recent assertions to the contrary¹, has not been matched in dynamic multiple multipole simulations [14]. This may be direct evidence of a situation where polariton effects are fundamentally altering the nature of the effective medium. The inability of quasi-static models to give correct results for some structures seems to be associated with the presence of sharp narrow features on a scale down to a few nanometers in metal surfaces, as occurs in touching spheres. One might expect nanoholes in metal films as in Figure 2(e) and in granular metals as in Figure 2(b) to be another situation where this occurs. We have found the optical properties of these two systems cannot be fitted with simple quasi-static models. In touching metal pairs, strong, highly localized circulating surface plasmon waves near contact points with SPP wavelengths as small as a few nanometers are seen in some models. Similar effects have been indicated for sharp indents in films or metal surfaces [25].

There is no good experimental evidence as yet to support these new models for touching spheres and sharp indents. This might be attributed to other effects such as spatial dispersion in the electron response on such scales. Dynamic modeling by my group on touching metal-coated spheres, however, shows no such anomalies; but, as noted previously, internal SP coupling may be responsible. Since great care

¹A footnote in Ref. 15 asserts that the exact analytic result in Ref. 41 is incorrect, based on an error in the appendix. The result is in fact correct and can be found by alternative routes (e.g., Ref. 42). The appendix in Ref. 41 unfortunately does have a misprint, with ν and μ interchanged in one equation, which was taken directly from the literature.

is needed to avoid numerical errors with multipole computational tools in full Mie treatments, much further experimental and theoretical work is needed.

Recent optical anomalies with nanoholes in metal and related effects in granular metals also indicate that quasi-static models may never work in such systems, but that effective-medium models can still be established. There were many unconvincing attempts to use classical quasi-static models for granular metals until recently. The granular film with oblique voids shown in Figure 2(b) made by depositing aluminum at typical low deposition pressures (but not ultralow vacuum) so as to grow obliquely has additional anomalies [49]. In particular, its effects on different polarizations of incident light were an unexpected surprise. Quasi-static models failed badly, whereas they had worked well for oblique isolated metal columns [50] in many studies. Surface effects such as SPPs can resolve the polarization dilemma, which relates to the internal symmetry of expected effective optical constants in an oblique void system. These films were also interesting in that they displayed another common feature of metal nanostructures with unusual responses; when overcoated with a dielectric (so as not to fill the voids), the unexpected effects were strongly enhanced.

What is clear in all such nanometal systems is that local field densities are very high, and these have been measured and observed [51]. The indications are that quasi-static models may be inappropriate whenever very large field enhancements are due to coupled surface plasmons between particles or across voids or thin films. Experimentally it means that weak or negligible scattering does not necessarily imply quasi-static models. Operationally, however, effective refractive indexes are indicated, and hence new types of effective-medium models are needed for granular metals and metals with random nanoholes. Some ideas such as the Generalized Ohm's Law noted earlier have recently been suggested (see chapter of Sarychev and Shalaev in this volume), but further development is needed on SPPs.

Quasi-static models of clusters of touching metal particles using standard sphere multipole expansions do not succeed because of the continuum of SP resonance modes. My group has recently looked at this problem with a new coupled multipole expansion technique [47], [52] and found that even with truncation of equations at pole orders of 500, one is still far from convergence. In dramatic contrast, the same cluster approximation with metal-coated nanoparticles shows complete convergence in under 10 pole orders and simple resonance structure. Convergence is directly related to the complexity and strength of external fields that are very strong near sharp contact points for metal. However, with coated metal the cross-coupling across the metal layer still dominates. Large fields are within particles, not between particles, so quasi-static models may still work here as unusual surface plasmons are less likely. Dynamic studies to explore this have now commenced. There is also a major shift in spectral behavior compared to close and touching metal particles. As metal particles approach each other, resonance wavelengths shift markedly, along with the advent of additional resonant poles, since plasmon coupling between particles continuously varies. For metal shells the resonance positions change marginally while spectral weights shift between S and

AS surface plasmon modes. Having an absorption band relatively insensitive to concentration is of practical and technological significance.

Films with tilting metal columns have an anisotropic nanostructure and hence display optical anisotropy in absorption. They were of interest to my own [50] and other research groups [53-54] as angular selective solar control films. Effective-medium models in the frame determined by ellipsoid axes as determined by $L^{h,m}$ in (4) can be transformed into the frame of the film normal and surface, to give the effective optical constants that are needed for modeling. Several groups have shown this procedure to work quite well. They also found these films to be an interesting testbed for fundamental effective-medium concepts because they have a semi-ordered structure with oblique metal columns that pack roughly in rows and can be approximated to nanoscale venetian blinds. As noted previously, a fascinating new variety of these films was recently discovered [49], which uses tilted voids that can form in some thin metal layers that have granular structure but are obliquely deposited. Quasi-static models fail here. SPPs with spatial character determined by the subwavelength voids seem to couple across the film and also provide unusual anisotropy. Standard normally deposited granular metals should also have interesting angular dependent properties, as surface polariton modes depend on the angle of incidence.

6 Conclusions

Nanostructured thin films containing conductors can provide a rich variety of spectral and directional properties at wavelengths where the conductor's relative permittivity is negative. Many applications are possible, some of which may replace conventional multilayer thin films. Detailed knowledge, and means of control of the nanostructure, as well as quantitative models of its influence on charge carrier response to radiation, help establish effective-medium models. These models then facilitate development of film production processes to achieve the many emerging applications, such as those listed in Table 1. Models provide an engineering basis for better exploitation of these phenomena. Dilute isolated particles, some lattice arrays, and some columnar thin films have such a basis, but many important structures cannot yet be routinely optically modeled. Three classes of effective-medium models have been identified, only two of which use quasi-static theories. The third is when extended currents or surface plasmon polaritons can form on a scale larger than the smallest nanofeatures, and it is more likely to cause major changes to response when very fine gaps or discontinuities occur in continuous metal or within clusters of metal nanoparticles. The study of thin coatings of metal on dielectric nanoparticles is a very promising new field, provided the quality of coating can be improved. It yields different surface plasmon and cluster character to metal particle arrays, and the optical response then involves up to three hierarchies of nanostructure that are not independent. Other emerging applications of these phenomena, not covered explicitly, include subwavelength lithography or imaging, surface light

guiding for nanophotonics, enhanced nonlinear response and possibly evanescent-wave amplification (or the superlens).

References

- [1] H. Ditlbacher, J.R. Krenn, G. Schider, A. Leitner and F.R. Aussenegg, Two-dimensional optics with surface plasmon polaritons, *Appl Phys Lett* **81** (2002), 1762–1764.
- [2] R. Messier, Toward quantification of thin film morphology, *J Vac Sci Technol A* **4** (1986), 490–495.
- [3] B.A. Movchan and A.V. Demchisin, Study of the structure and properties of thick vacuum condensates of nickel, titanium, tungsten, aluminium oxide and zirconium oxide, *Phys Met Metallogr* **28** (1969), 83–90.
- [4] J.M. Nieuwenhuizen and H.B. Haanstra, Microfractography of thin films, *Philips Tech Rev* **27** (1966), 87–91.
- [5] A.G. Dirks and H.J. Leamy, Columnar microstructure in vapor-deposited thin films, *Thin Solid Films* **47** (1977), 219–233.
- [6] P.J. Martin, R.P. Netterfield and T.J. Kinder, Ion-beam-deposited films by filtered arc evaporation, *Thin Solid Films* **193/194** (1990), 77–83.
- [7] S. Veprek, A. Niederhofer, K. Moto, T. Bolom, H.-D. Maennling, P. Nesladek, G. Dollinger and A. Bermaier, Composition, nanostructure and origin of the ultrahardness in nc-TiN/a-Si₃N₄/a- and nc-TiSi₂ nanocomposites with $H_v = 80$ to ≥ 105 GPa, *Surf Coat Technol* **133/134** (2000), 152–159.
- [8] G. Mbise, G.B. Smith, G.A. Niklasson and C.G. Granqvist, Angular-selective optical properties of Cr films made by oblique-angle evaporation, *Appl Phys Lett* **54** (1989), 987–989.
- [9] P.M. Valanju, R.M. Walser and A.P. Valanju, Wave refraction in negative index media: always positive and very inhomogeneous, *Phys Rev Lett* **88** (2002), 187401.
- [10] A. Lakhtakia, M.W. McCall and W.S. Weiglhofer, Negative phase-velocity mediums, *Introduction to complex mediums for optics and electromagnetics* (W.S. Weiglhofer and A. Lakhtakia, eds), SPIE Press, Bellingham, WA, USA, 2003.
- [11] N. Garcia and M. Nieto-Vesperinas, Left-handed materials do not make a perfect lens, *Phys Rev Lett* **88** (2002), 207403.
- [12] G.B. Smith, Effective medium theory and angular dispersion of optical constants in films with oblique columnar microstructure, *Opt Commun* **71** (1989), 279–284.
- [13] G.B. Smith, Theory of angular selective transmittance in oblique columnar thin film containing metal and voids, *Appl Opt* **29** (1990), 3685–3693.
- [14] T.W. Ebbesen, H.J. Lezec, H.F. Ghaemi, T. Thio and P.A. Wolff, Extraordinary optical transmission through sub-wavelength hole arrays, *Nature* **391** (1998), 667–669.

- [15] A. Pack, M. Hietschold and R. Wannemacher, Failure of local Mie theory: optical spectra of colloidal aggregates, *Opt Commun* **194** (2001), 277–287.
- [16] H.C. van de Hulst, *Light scattering by small particles*, Dover Press, New York, NY, USA, 1981.
- [17] A. Lakhtakia and R. Messier, The past, the present and the future of sculptured thin films, *Introduction to complex mediums for optics and electromagnetics* (W.S. Weiglhofer and A. Lakhtakia, eds), SPIE Press, Bellingham, WA, USA, 2003.
- [18] S.K. Andersson and C.G. Ribbing, Light extinction by bulk voids and surface irregularities in ceramic materials, *Phys Rev B* **49** (1994), 11336–11343.
- [19] S.J. Oldenburg, R.D. Averitt, S.L. Westcott and N.J. Halas, Nanoengineering of optical resonances, *Chem Phys Lett* **288** (1998), 243–247.
- [20] G.B. Smith, A. Gentle, P.D. Swift, A. Earp and N. Mronga, Coloured paints based on coated flakes of metal as the pigment, for enhanced solar reflectance and cooler interiors: description and theory, *Sol Energy Mater Sol Cells* **79** (2003), 163–177.
- [21] G.B. Smith, A. Gentle, P.D. Swift, A. Earp and N. Mronga, Coloured paints based on iron oxide and silicon oxide coated flakes of aluminum as the pigment, for energy efficient paint: optical and thermal experiments, *Sol Energy Mater Sol Cells* **79** (2003), 179–197.
- [22] G.B. Smith, Dielectric constants for mixed media, *J Phys D: Appl Phys* **10** (1977), L39–L42.
- [23] G.B. Smith, The scope of effective medium theory for fine metal particle solar absorbers, *Appl Phys Lett* **35** (1979), 668–670.
- [24] G.A. Niklasson, C.G. Granqvist and D. Hunderi, Effective medium models for the optical properties of inhomogenous materials, *Appl Opt* **20** (1981), 26–30.
- [25] J.R. Sambles, T.I.N. Preist, W.J. Tan and N.P. Wanstall, Surface plasmons and zero order metal gratings, *Proceedings of IUTAM symposium on mechanical and electromagnetic waves in structured media* (R.C. McPhedran, L.C. Botten and A. Nicorovici, eds), Kluwer, Dordrecht, The Netherlands, 2001, 285–296.
- [26] R. Levy-Nathaisson and D.J. Bergman, Studies of the generalized Ohm's Law, *Physica A* **241** (1997), 166–172.
- [27] V.A. Shubin, A.K. Sarychev, J.P. Clerc and V.M. Shalaev, Local electric and magnetic fields in semicontinuous metal films: Beyond the quasi-static approximation, *Phys Rev B* **62** (2000), 11230–11244.
- [28] A.J. Reuben and G.B. Smith, Normal-mode decomposition for the optical response of cylinder clusters, *Phys Rev E* **58** (1998), 1101–1111.
- [29] P. Taneja, P. Ayyub and R. Chandra, Size dependence of the optical spectrum in nanocrystalline silver, *Phys Rev B* **65** (2002), 245412.
- [30] G.A. Niklasson and C.G. Granqvist, Surfaces for selective absorption of solar energy: an annotated bibliography 1955–1981, *J Mater Sci* **18** (1983), 3475–3534.

- [31] Q.C. Zhang and D.R. Mills, New cermet film structures with much improved selectivity for solar thermal applications, *Appl Phys Lett* **60** (1992), 545–547.
- [32] L. Harris, R.T. McGinnies and B.M. Siegel, The preparation and optical properties of gold blacks, *J Opt Soc Am* **38** (1948), 582–589.
- [33] L. Harris, *The optical properties of metal black and carbon blacks*, The Eppley Foundation, Kingston, RI, USA, 1967.
- [34] G.B. Smith, C.A. Deller, P.D. Swift, A. Gentle, P.D. Garrett and W.K. Fisher, Nanoparticle-doped polymer foils for use in solar control glazing, *J Nanopart Res* **4** (2002), 157–165.
- [35] S. Schelm and G.B. Smith, Dilute LaB₆ nanoparticles in polymer as optimized clear solar control glazing, *Appl Phys Lett* **82** (2003), 4346–4348.
- [36] R.W. Cohen, G.D. Cody, M.D. Coutts and B. Abeles, Optical properties of granular silver and gold films, *Phys Rev B* **8** (1973), 3689–3701.
- [37] I.L. Skryabin, A.V. Radchik, P. Moses and G.B. Smith, The consistent application of Maxwell-Garnett effective medium theory to anisotropic composites, *Appl Phys Lett* **70** (1997), 2221–2223.
- [38] D.J. Bergman, The dielectric constant of a composite material—a problem in classical physics, *Phys Rept* **43** (1978), 377–407.
- [39] R. Ruppin, Surface modes of two spheres, *Phys Rev B* **26** (1982), 3440–3444.
- [40] P.E. Batson, Surface plasmon coupling in clusters of small spheres, *Phys Rev Lett* **49** (1982), 936–940.
- [41] A.V. Paley, A.V. Radchik and G.B. Smith, Quasistatic optical response of pairs of touching spheres with arbitrary dielectric permeability, *J Appl Phys* **73** (1993), 3446–3453.
- [42] A.V. Radchik, A.V. Paley, G.B. Smith and A.V. Vagov, Polarization and resonant absorption in intersecting cylinders and spheres, *J Appl Phys* **76** (1994), 4827–4835.
- [43] A.V. Vagov, A. Radchik and G.B. Smith, Optical response of arrays of spheres from the theory of hypercomplex variables, *Phys Rev Lett* **73** (1994), 1035–1038.
- [44] G.B. Smith and A.J. Reuben, Normal mode analysis of optical polarisation response in nanocomposite particles, *Nanostruct Mater* **12** (1999), 353–356.
- [45] S.J. Oldenburg, G.D. Hale, J.B. Jackson and N.J. Halas, Light scattering from dipole and quadrupole nanoshell antennas, *Appl Phys Lett* **75** (1999), 1063–1065.
- [46] C. Graf and A. van Blaaderen, Metallodielectric colloidal core-shell particles for photonic applications, *Langmuir* **18** (2002), 524–534.
- [47] G.B. Smith and V.N. Pustovit, Coupled multipolar interactions in clusters of nanoparticles with metal shells, *Opt Commun* **211** (2002), 197–204.
- [48] H. Takei, Biological sensor based on localized surface plasmons associated with surface-bound Au/polystyrene composite microparticles, *Microfluidic devices and systems* (A. Frazier and C. Ahn, eds), Proc SPIE **3515**, Bellingham, WA, USA, 1998, 278–283.

- [49] G.B. Smith, A.K.M. Hossain and A. Gentle, Near infra-red radiation squeezing through 20 nm voids in obliquely deposited metal films, *Appl Phys Lett* **78** (2001), 2143–2144.
- [50] G.B. Smith, S. Dligatch, R. Sullivan and M.G. Hutchins, Thin film angular selective glazing, *Solar Energy* **62** (1998), 229–244.
- [51] S. Grésillon, J.-Cl. Rivoal, P. Gadenne, X. Quélin, V. Shalaev and A. Sarychev, Nanoscale observation of enhanced electromagnetic field, *Phys Stat Sol (a)* **175** (1999), 337–343.
- [52] V.N. Pustovit, J.A. Sotelo and G.A. Niklasson, Coupled multipolar interactions in small-particle metallic clusters, *J Opt Soc Am A* **19** (2002), 513–518.
- [53] M. Suzuki and Y. Taga, Anisotropy in absorption of Ag-SiO₂ thin films with oblique columnar structures, *J Appl Phys* **71** (1992), 2848–2854.
- [54] D. Le Bellac, G.A. Niklasson and C.G. Granqvist, Angular-selective optical transmittance of anisotropic inhomogeneous Cr-based films made by sputtering, *J Appl Phys* **77** (1995), 6145–6151.



Geoff B. Smith is Professor of Applied Physics at the University of Technology, Sydney, Australia, and currently leads research programs in the physics area and at its Institute of Nanoscale Technology. He has worked on theory and practical aspects of materials for solar energy and energy efficiency for 26 years, recently receiving a Renewable Energy Pioneer award from the World Renewable Energy Network and a Doctoral degree (honoris causa) from the University of Uppsala in Sweden for his scientific and technological contributions to optical materials, solar energy, windows, lighting and daylighting.

He is a Fellow of the Australian Institute of Physics, and has led and contributed to IEA projects. He has more than 135 reviewed publications, two chapters in a recent textbook on nanotechnology, and 9 patents. His work has contributed to a number of lighting and glazing products, most recently laminated solar control windows using lanthanum hexaboride nanoparticles and supersidelight specially doped flexible polymer light guides for efficient continuous light delivery over long distances. He is also currently researching metal and metal shell nanoparticle arrays, nanohole arrays in metal, light mixing, novel skylights, new approaches to microstructured polymer optical fiber, and outdoor display systems.

The Past, the Present, and the Future of Sculptured Thin Films

Akhlesh Lakhtakia and Russell Messier

Abstract

Sculptured thin films are nanostructured inorganic materials with unidirectionally varying properties that can be designed and realized in a controllable manner using physical vapor deposition. We trace their development from their precursors, from chronological as well as morphological points of view. Their continuum electromagnetic constitutive relations are presented, along with a control model to link their nanostructure to their continuum properties. Both accomplished and emerging applications are reviewed, and directions for future research are suggested.

1 Introduction

In its decadal survey entitled *Physics in a New Era* conducted during the 1990s, the National Research Council (NRC) of the United States naturally explored research trends and requirements in the materials sciences [1]. A dominant theme that emerged is of nanosciences and nanotechnologies. The nanoscale is Janusian: matter at the 10–100-nm length scale exhibits continuum characteristics, but molecules and their microclusters can still display their individuality. For that reason, the U.S. National Science Foundation has begun to focus on material morphologies and architectures with at least one dimension smaller than 100 nm in its research initiatives.

Among the nanoengineered materials identified by the NRC are sculptured thin films (STFs) [1, p. 123]. This chapter is focused on these nanostructured inorganic materials with unidirectionally varying properties that can be designed and realized in a controllable manner using physical vapor deposition (PVD) [2]–[5]. The ability to virtually instantaneously change the growth direction of their columnar morphology, through simple variations in the direction of the incident vapor flux, leads to a wide spectrum of columnar forms. These forms can be

- i) two-dimensional, ranging from the simple slanted columns and chevrons to the more complex C- and S-shaped morphologies; as well as
- ii) three-dimensional, including simple helixes and superhelixes.

Several examples of STFs are presented in Figures 1 and 2.

For most envisioned applications, the column diameter and the column separation normal to the thickness direction of any STF should be constant. The column diameter can range from about 10 to 300 nm, while the density may lie between its theoretical maximum value to less than 20% thereof. The crystallinity must be at a scale smaller than the column diameter. The chemical composition is essentially unlimited, ranging from insulators to semiconductors to metals. Despite the fact that precursors of STFs have been made for over a century [6]–[13], systematic exploration of the science and technology of STFs began only during the mid-1990s.

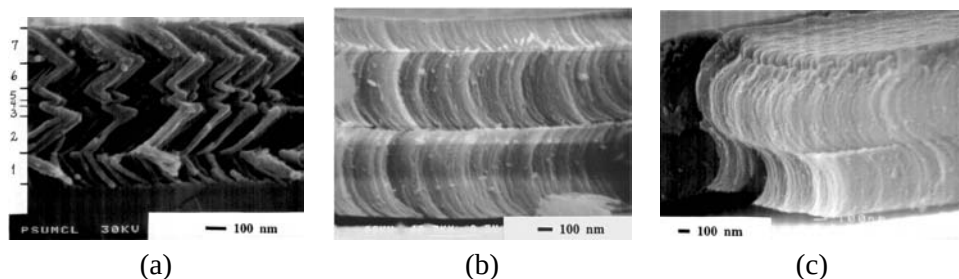


Figure 1 Scanning electron micrographs of sculptured thin films made of MgF₂ with two-dimensional morphologies: (a) seven-section zigzag; (b) C-shaped; (c) S-shaped.

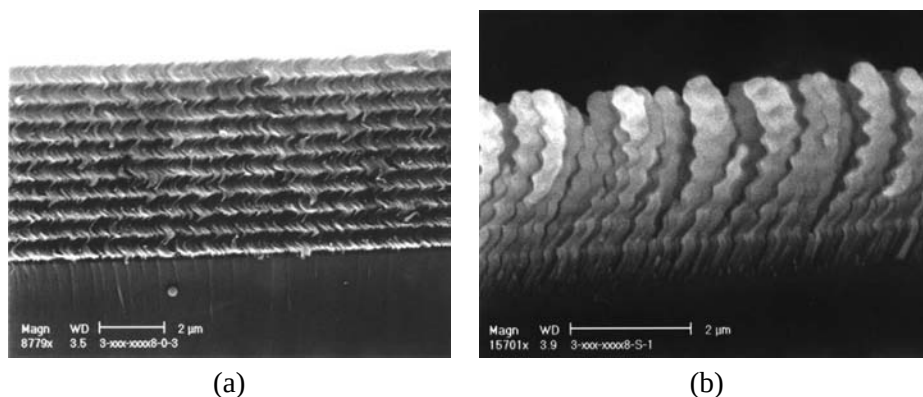


Figure 2 Scanning electron micrographs of sculptured thin films with three-dimensional morphologies: (a) helical (SiO); (b) superhelical (MgF_2).

What is emerging from recent research [14]–[19] is that STF technology has a bright future.

At visible and infrared wavelengths, a single-section STF is a unidirectionally nonhomogeneous continuum with direction-dependent properties. Several sections can be grown consecutively into a multisection STF, which can be conceived of as an optical circuit that can be integrated with electronic circuitry on a chip. Being porous, a STF can act as a sensor of fluids and can be impregnated with liquid crystals for switching applications too. Application as low-permittivity barrier layers in electronic chips as well as for solar cells has also been suggested. The first optical applications of STFs saw the light of the day in 1999.

This chapter is organized as follows: in Section 2, we trace the development of STFs from their precursors, with emphasis on two seminal events that occurred in 1959 and 1966, as well as on the sculpturing of the columnar morphology of the so-called columnar thin films (CTFs). The defining continuum electromagnetic constitutive relations of STFs are presented in Section 3, along with a control model to link the nanostructure of STFs to their continuum properties *via* a local homogenization formalism. Both accomplished and emerging applications are reviewed in Section 4, with special reference to optics; and we end in Section 5 with prognostications of research trends and requirements.

2 From columnar to sculptured thin films

2.1 Columnar thin films

Chronologically as well as morphologically, it is sensible to begin with the CTFs. These films are analogous to orthorhombic crystals because of their columnar morphology. The sculpturing of that morphology during growth underlies the STF concept.

The growth of CTFs by PVD is usually credited to Kundt in 1885 [6]. At low temperature and pressure, material in a source boat evaporates towards a substrate and the arriving adatoms settle on it to form a thin film, as shown in Figure 3. Isolated nucleation clusters approximately 1–3 nm in diameter initially form on the substrate. The clusters evolve into expanding and competing columns as the film thickness increases, if the film temperature is below approximately a third of its melting point [20]. The addition of ion bombardment during growth can eliminate columns, thereby yielding dense, smooth and stable thin films that meet the stringent requirements for laser-based applications of optical coatings. Scanning electron micrographs of a series of amorphous silicon CTFs are shown in Figure 4.

Significantly, an intermediate state occurs between columnar expansion and the elimination of the columns. In that state, competition between neighboring columns is frustrated [21] and stable columns grow. The CTF morphology is achieved either through intermediate levels of ion bombardment or by simply depositing the films at oblique angles [22]. The columns thus grow at a controllable angle $\chi \geq 25^\circ$ to the substrate, as shown in Figure 5, while the average direction of the incident vapor flux is delineated by the angle $\chi_v \leq \chi$.

2.1.1 Growth mechanics

Columnar thin films are the direct result of self-shadowing at the length scale of the arriving adatoms (which are atoms on the growing film surface before they attain their final free energy state). CTFs are fabricated under low adatom-mobility conditions—where the sticking coefficient for adatoms is essentially unity and when the vapor arrives from a specific direction, either totally (e.g., in evaporation from a point source at 10^{-5} Torr) or on average (e.g., in sputtering from a plane source at 10^{-3} to 10^{-1} Torr). Only two issues need to be addressed in order

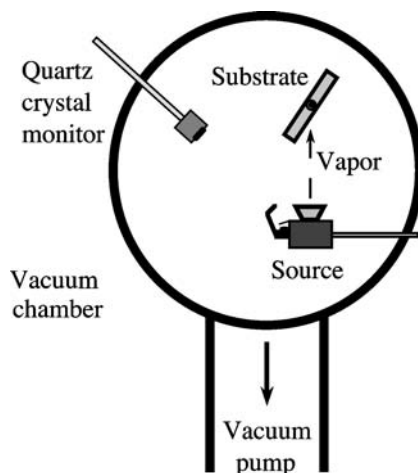


Figure 3 Schematic of the basic system for physical vapor deposition of columnar thin films on planar substrates. Although an electron-beam evaporation point source is shown, distributed directional sources—such as those used in sputter deposition—can be used to similar effect.

to understand the origin of CTFs: first, whether the arriving adatoms form a continuous network or do they cluster; and, second, if they form clusters, then do the clusters grow independently or do they compete for growth much like the leaves of a tree vying for sunlight. As it turns out, clusters with associated honeycomblike density deficit regions (commonly called void networks) are formed. Furthermore, the clusters do compete for growth, leading to power-law-shaped columns that are hierarchical cluster aggregates resembling a cauliflower when viewed from above their top surface. This has been seen experimentally [23]-[25] and has also been modeled *via* ballistic aggregation [26] and molecular dynamics [27] simulation studies.

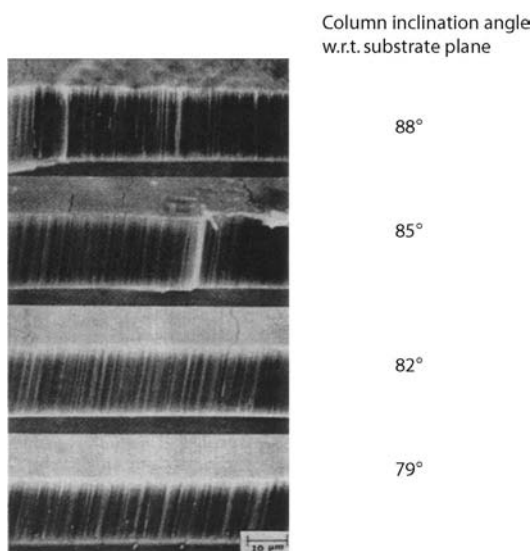


Figure 4 Scanning electron micrographs of four columnar thin films made of amorphous silicon. The column inclination angle χ , defined *via* Figure 5, is indicated for all four CTFs.

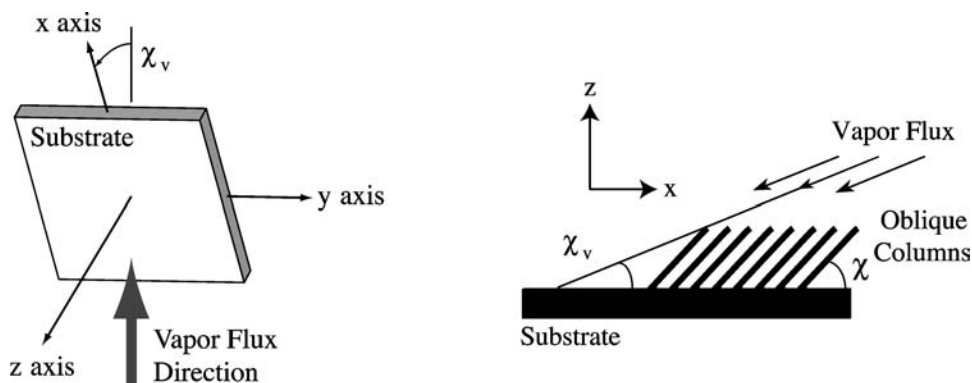


Figure 5 Coordinate system, the vapor incidence angle χ_v , and the column inclination angle χ of a CTF.

For deposition normal to the substrate surface (i.e., $\chi_v = 90^\circ$), the adatoms arrive on rectilinear trajectories and form 1–3-nm size clusters [20], [25]. The sizes of these clusters have a statistical variation, and the larger clusters tend to capture more of the arriving adatoms. But the clusters do not simply get larger; rather, more clusters of the same size distribution continue to form atop the clusters formed earlier, and the clusters tend to aggregate, which leads to competitive columnar growth [20]. The edges of the dominant columns have a capture radius that extends beyond the perimeter of the cluster/column boundary, thus making columnar expansion a possibility. This statistically favorable expansion of the tallest columns leads to shadowing of the smaller adjoining columns. As the top surface area is virtually fixed, at least one column must contract for every column that expands. In the absence of any additional moderating mechanism, the columns evolve eventually to yield a cauliflowerlike surface morphology observable with conventional scanning electron microscopes for films greater than $\sim 1 \mu\text{m}$ in thickness. This has been classified as the Zone 1 morphology in the well-known Structure Zone Model (SZM) of Thornton [28]. The column diameters expand according to a power-law dependence in which the power-law exponent is a function of the deposition conditions [23], [29].

The columns *can* become noncompetitive in their growth evolution [21]. These parallel columnar structures constitute the so-called matchstick morphology, since they resemble a bundle of wooden matchsticks when viewed from the top and the sides [22], [30]. The latest revision [31] of the SZM, shown in Figure 6, contains the Zone M for matchstick morphology. This zone has been observed in sputtered thin films of a-Si, a-Ge, SiC, and WO_3 , where the moderating mechanism appears to be ion bombardment at normal incidence to the substrate, near and above the sputtering threshold (~ 30 – 50 eV for most materials). Under such low-energy bombardment conditions, the most loosely bound adatoms, at the edges of the clusters and columns, are very likely forward-sputtered toward the underlying shadowed regions. Then, as the top surfaces of the favored columns become rougher and more convoluted with evolutionary development, the increase in sputtering yield as a function of χ_v [32] drives the top surface and column edges toward the steady-state morphology characteristic of Zone M. Furthermore, the interior morphology of the individual matchsticks is expected to resemble the noncompetitive morphology of the transition zone, termed Zone T by Thornton [28].

Matchstick morphology also appears when films are deposited at oblique incidence angles, especially when χ_v is small (i.e., at glancing angles). As χ_v increases, the cluster density also increases due to decreased self-shadowing. In consequence, the columns become less separated and, thus, more competitive for growth evolution. This has also been seen in ballistic aggregation simulations [22].

In an extensive review of both experimental and ballistic aggregation modeling studies of obliquely deposited CTFs, van Kranenburg and Lodder [33] concluded that elongated clusters and columns generally pointing in the direction of the incoming vapor flux are a direct consequence of the adatomic self-shadowing process; furthermore, the length of the long axis relative to the width of the cluster

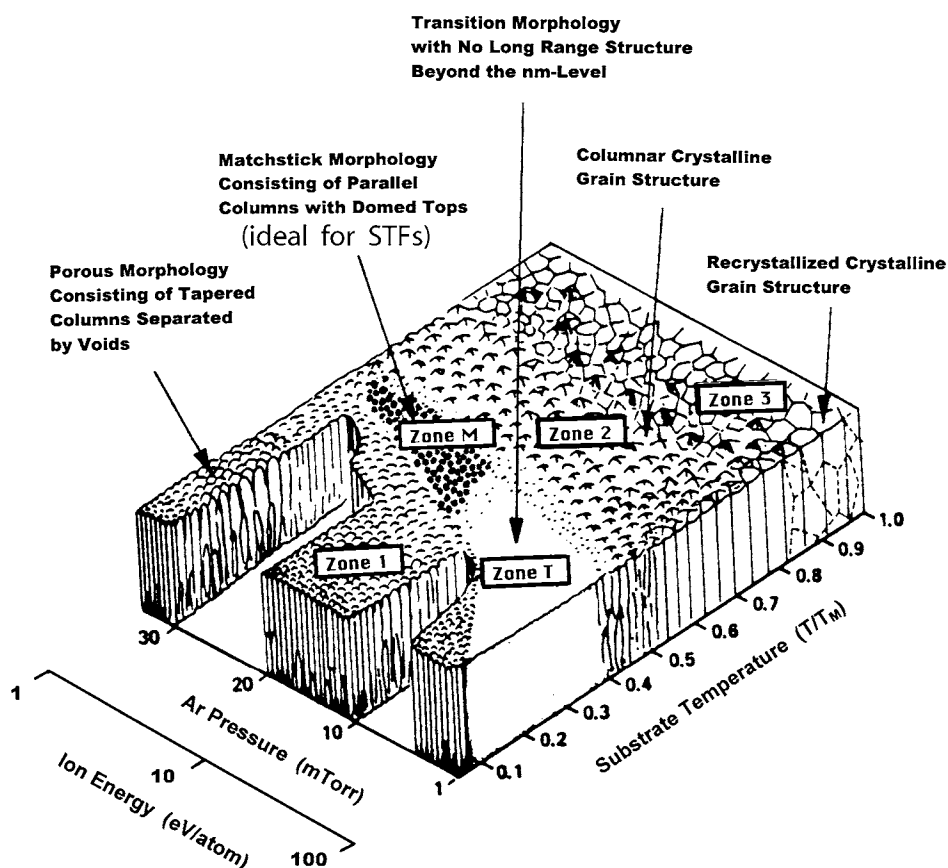


Figure 6 The latest version of the Structure Zone Model for columnar thin films [31]. The process variables are the reduced temperature T/T_M (where T_M is the melting point), the argon pressure in the PVD chamber, and the ion bombardment energy. This figure is adapted from Thornton's highly cited publication [28] in order to include the matchstick morphology of Zone M that lies between Zones 1 and T. Zones 2 and 3 are irrelevant for STFs, since high adatom-mobility eliminates the effects of self-shadowing and clustering.

increases markedly for $\chi_v < 30^\circ$. The columns become separated and noncompeting cylinders—with elliptical cross-sections due to anisotropy in self-shadowing [22], [33]—as χ_v is decreased further. The columns become more separated in the vapor incidence direction due to the increased shadowing effect in the longitudinal direction (parallel to the vapor incidence plane), while shadowing in the transverse direction is unaffected by changes in χ_v . This leads to a higher material density in the transverse direction, and to a density anisotropy that has been exploited commercially in magnetic thin films [33].

2.1.2 Dielectric continuum characterization

As the columnar cross-sectional radius is less than or equal to 150 nm for a large variety of CTFs, they can be considered effectively as homogeneous orthorhombic

continuums at visible and infrared frequencies, depending on the constitutive parameters of the deposited material [34]. Generally thought of as dielectric materials, their optical birefringence has long been appreciated and exploited [35-36].

If the plane $z = 0$ is a reference plane parallel to the planar substrate, the three unit vectors

$$\mathbf{u}_\tau = \mathbf{u}_x \cos \chi + \mathbf{u}_z \sin \chi, \quad (1)$$

$$\mathbf{u}_n = -\mathbf{u}_x \sin \chi + \mathbf{u}_z \cos \chi, \quad (2)$$

$$\mathbf{u}_b = -\mathbf{u}_y \quad (3)$$

assist in the electromagnetic characterization of both CTFs and STFs, with $\mathbf{u}_x$, $\mathbf{u}_y$ and $\mathbf{u}_z$ as the three Cartesian unit vectors. Whereas $\mathbf{u}_\tau$ is parallel to the columns of a CTF, $\mathbf{u}_n$ is perpendicular to $\mathbf{u}_\tau$, and both unit vectors lie in the plane of deposition (the xz plane). The unit vector $\mathbf{u}_b$ is normal to that plane.

The constitutive relations of a CTF may be set up as functions of position vector $\mathbf{r} = x\mathbf{u}_x + y\mathbf{u}_y + z\mathbf{u}_z$ and angular frequency ω as

$$\mathbf{D}(\mathbf{r}, \omega) = \epsilon_0 \underline{\underline{\epsilon}}_r(\omega) \cdot \mathbf{E}(\mathbf{r}, \omega), \quad (4)$$

$$\mathbf{B}(\mathbf{r}, \omega) = \mu_0 \mathbf{H}(\mathbf{r}, \omega). \quad (5)$$

In these equations, $\epsilon_0 = 8.854 \times 10^{-12}$ F/m and $\mu_0 = 4\pi \times 10^{-7}$ H/m are the permittivity and the permeability of free space (i.e., vacuum), respectively, and an $\exp(-i\omega t)$ time-dependence is implicit.

Extensive experimental research on CTFs [37]-[42] necessitates the orthorhombic prescription

$$\underline{\underline{\epsilon}}_r(\omega) = \epsilon_a(\omega) \mathbf{u}_n \mathbf{u}_n + \epsilon_b(\omega) \mathbf{u}_\tau \mathbf{u}_\tau + \epsilon_c(\omega) \mathbf{u}_b \mathbf{u}_b \quad (6)$$

for the relative permittivity dyadic $\underline{\underline{\epsilon}}_r(\omega)$.¹ This dyadic is biaxial, with the reduction $\epsilon_c = \epsilon_a$ appropriate for uniaxiality. Optical measurements indicate that some CTFs can be uniaxial [38-39], but biaxiality is probably the norm [40]-[42]. Biaxiality may arise from the columnar cross-sections being noncircular, as discussed in Section 2.1.1, and can have significant optical consequences relative to uniaxiality for CTFs as well as STFs [45]-[47].

Increased adatomic self-shadowing with decrease in χ_v creates a density anisotropy. This implies that not only should $\underline{\underline{\epsilon}}_r(\omega)$ have the orthorhombic form (6), but also that ϵ_a , ϵ_b and ϵ_c be dependent on the column inclination angle χ .

¹See [43, Ch. 1] for a lucid introduction to dyadics. A serviceable introduction has also been provided by Weiglhofer [44] in this volume. For all intents and purposes in this chapter, a dyadic is equivalent to a 3×3 matrix.

A series of optical characterization experiments on CTFs of the oxides of tantalum, titanium and zirconium [42] suggests the following parametric relationships:

$$\epsilon_a(\omega, \chi) = [a_0(\omega) + a_1(\omega)v + a_2(\omega)v^2]^2, \quad (7)$$

$$\epsilon_b(\omega, \chi) = [b_0(\omega) + b_1(\omega)v + b_2(\omega)v^2]^2, \quad (8)$$

$$\epsilon_c(\omega, \chi) = [c_0(\omega) + c_1(\omega)v + c_2(\omega)v^2]^2, \quad (9)$$

$$\tan \chi = m \tan \chi_v. \quad (10)$$

Here, the coefficient $m > 1$ and the functions $a_0(\omega)$, etc., vary with the deposited material species and the deposition conditions, while $v = \chi_v/(\pi/2)$ is the vapor incidence angle expressed as a fraction of a right angle. For vapor incidence perpendicular to the substrate, $v = 1$ and the columnar cross-section is circular; hence, $a_0 + a_1 + a_2 \equiv c_0 + c_1 + c_2$. Equations (7)–(10) should hold at frequencies substantially distant from material resonances [48]. A modification of (10) to $\tan \chi = k + m \tan \chi_v$, $k \neq 0$, may hold better for very low values of χ_v . Let us note that (7)–(10) may have to be modified for CTFs realized by techniques other than thermal evaporation.

Parenthetically, in order to enhance biaxiality and linear birefringence, Hodgkinson and Wu [49] modified the standard PVD technique by using two identical source boats. In their serial bideposition technique, both vapor fluxes are directed in the xz plane, but one flux is directed at an angle χ_v and the other at $180^\circ - \chi_v$. Furthermore, the sources are switched on and off alternately, so that each source is responsible for the deposition of ~ 3 nm thickness only. The resulting columns are upright so that $\chi = 90^\circ$. Uniaxiality should arise, however, if one of the two sources directs vapor in the xz plane and the other in the yz plane [50].

2.2 Primitive STF with nematic morphology

A seminal event occurred in 1966 that eventually led to the emergence of the STF concept in 1994 [2]. While a CTF was growing, Nieuwenhuizen and Haanstra deliberately altered the vapor flux angle χ_v to prove that columnar morphology “cannot be a result of the method of preparation itself” [12]. The resulting change in the column inclination angle χ was accomplished while the film thickness grew ~ 3 nm, the transition being practically abrupt with respect to visible wavelengths. Some two decades later, Motohiro and Taga demonstrated that χ can be abruptly altered many times during growth [13], which is the basis for realizing STF with bent nematic morphologies.

Thus, primitive STF with zigzag and chevronic morphologies came into existence. The similarity of CTFs to crystals had long been noticed in the optical literature [35], so that the primitive STF with nematic morphology could be seen as stacked crystalline plates. This was astutely exploited for designing, fabricating and testing various optical devices such as anti-reflection coatings, reflectors, linear polarization filters, phase retarders, and polarizers [36], [51].

Serial as well as simultaneous bideposition of CTFs and chevronic STF was systematically undertaken by Taga and colleagues [52]-[54] at Toyota R&D. More than one material species were deposited in order to fabricate waveplates with desirable optical retardance as well as uniformity and mechanical stability. Mass-produced, these waveplates are now routinely employed for glare-proof rearview and outer mirrors.

2.3 Chiral STFs

Another seminal event towards the emergence of the STF concept had already occurred in 1959, but had evidently been ignored. Coming to know of it in 1998, we feel strongly that all credit for periodic STFs with chiral (i.e., handed) morphology be accorded to Young and Kowal [9]. Without actually seeing the anisotropic morphology of CTFs through scanning electron microscopy or otherwise, these two pioneers consciously [55] rotated the substrate about the z axis constantly during growth to create thin films with morphology predicted to display transmission optical activity. They wrote as follows on their optically active fluorite films:

The possibility of evaporating in vacuum an optically active film of an otherwise isotropic material, by oblique deposition on to a flat glass substrate revolving about its normal as an axis, is confirmed. The helically deposited fluorite films ranged up to about $[18\ \mu\text{m}]$ thick, and turned the plane of polarization of a normally incident beam by as much as 2.25° at [the free-space wavelength of] $0.546\ \mu\text{m}$.

We speculate that Young and Kowal must have been the first researchers to deliberately engineer thin-film morphology for producing a nontrivial STF—one with a full three-dimensional morphology.

Although the concept of *voids* in thin films was to become common 12 years later, Young and Kowal actually used that word to infer that “the [optical] activity of a helically deposited film could be due to the cooperative action of a helically symmetrical arrangement of crystallites, crystal growth or voids.” Furthermore, they conjectured that the columnar direction could change virtually instantaneously and continuously with changes in the position and the orientation of the substrate.

Their pioneering work on chiral STFs apparently went unnoticed, possibly because scanning electron microscopes—routine tools nowadays for directly observing morphology—were then not available to lend veracity to their optical activity measurements, which also were perhaps clouded by doubt among contemporary researchers. The Young–Kowal technique of rotating the substrate, the helicoidal morphology realized thereby, and the transmission optical activity of chiral STFs were rediscovered a few years ago [56]-[58].

2.4 Sculptured thin films

During the 1990s [2-3] recognition finally came that (i) a very wide variety of columnar morphologies is possible through control of two fundamental axes of substrate rotation; (ii) multisections with different columnar morphologies and/or

different materials are possible; (iii) a wide range of potential applications exists; and (iv) preparation-property-application connections can be truly engineered through a coupling of theoretical and experimental results.

STFs are modifications of CTFs in which the column direction can be changed almost abruptly during their growth evolution. When CTFs are deposited obliquely, a wide variety of STF morphologies tailored at the nanometer scale are realizable by simple variations of two fundamental axes of rotation, either separately or concurrently [30], [56], [59]–[67]. These fundamental axes lead to two canonical classes of STFs that have been termed sculptured nematic thin films (SNTFs) [60] and thin-film helicoidal bianisotropic mediums (TFHBMs) [56]. More complex shapes and even multisections, in which either the material or the shape or both are changed from section to section along the z axis, can be easily conceived and have been executed [68–69].

SNTF morphologies include such simple two-dimensional shapes as slanted columns, chevrons, and zigzags as well as the more complex C- and S-shapes; see Figure 1. The substrate has to be rotated about the y axis, which lies in the substrate plane and is perpendicular to the vapor incidence direction, while χ_v is varied either episodically or continuously [60]. One concern with this approach is related to the fact that the density of a CTF is highly dependent upon χ_v [5], [60], [62] and, therefore, density variations are expected as a SNTF grows. The compensation of these variations is an area of future research.

TFHBMs are fabricated by tilting the substrate at some oblique angle to the incident vapor flux (i.e., $\chi_v \leq 90^\circ$), followed by substrate rotation about the z axis. Helicoidal morphologies result for constant rotational velocity about the z axis [9], [56]. By varying the rotational velocity in some prescribed manner throughout a rotational cycle, a slanted helicoidal structure occurs with the slant angle controllable over all χ above its minimum value for static glancing angle deposition. Furthermore, by making the rotational cycle $360^\circ + \Delta$, where $\Delta \leq 0^\circ$ is some chosen phase-shift angle, it is possible to engineer a wide range of superhelixes with controlled handedness [5], [30]. The mass density as a function of film thickness is expected to remain constant since χ_v is fixed for TFHBMs, so long as the columns attain a steady-state diameter in the early nucleation and growth stages. However, as seen in Figure 2, the column width of the superhelix increases with growth evolution, which begs the scientific question: why?

In many reports on STFs, the columns shown by cross-section scanning electron microscopy usually appear parallel with cross-sectional diameters of 10–300 nm and with a similar range in separation distances. Since most optical applications of STFs will require that the column diameter, the separation distance between columns, and the connectivity between columns all remain constant during growth, this steady-state column size condition has practical advantages. For instance, when an STF deposition-cycle is repeated for a multiple-period TFHBM prepared by both evaporation and sputtering at constant angular velocities of substrate rotation, it is important that the last turn has the same pitch and porosity as all previous turns. This morphological reproducibility leads to, for instance, predictable optical rotation spectrums [58], [65], since the column sizes and shapes

allow the STFs to effectively act as rotationally nonhomogeneous continua at optical frequencies [4], [14]. This can only occur if the films evolve from the initial clusters into steady-state, noncompetitive, parallel columns within a small fraction of the total film thickness, and constant cross-section as well as constant porosity are maintained for the remainder of the film growth.

3 Electromagnetic field equations

Although nonoptical applications of STFs are certainly possible [3], [70-71], optical applications have been the chief drivers of STF research [14-15], [69], [72-73]. Moreover, optical monitoring of STF growth appears necessary in a production setting. Therefore, the linear electromagnetic properties—mostly for application in the visible and the infrared regimes—of STFs are recounted in this section. The description of STF properties follows very naturally from that of CTF properties.

3.1 Linear constitutive equations

By definition, the morphology of a single-section STF in any plane $z = z_1$ can be made to coincide with the morphology in another plane $z = z_2$ with the help of a suitable rotation. In other words, the *local* morphology is spatially uniform, but the *global* morphology is unidirectionally nonhomogeneous. Naturally, this leads to the concept of local constitutive properties of a STF. The global constitutive properties of a STF can be connected to the local ones by means of rotation operators.

As discussed by Weiglhofer elsewhere in this volume [44], the most general linear medium is bianisotropic; i.e., both the electric displacement phasor $\mathbf{D}(\mathbf{r}, \omega)$ and the magnetic field phasor $\mathbf{B}(\mathbf{r}, \omega)$ depend linearly on both the electric field phasor $\mathbf{E}(\mathbf{r}, \omega)$ and the magnetic induction phasor $\mathbf{H}(\mathbf{r}, \omega)$.² Accordingly, the linear frequency-domain constitutive relations of a single-section STF are set up as

$$\begin{aligned} \mathbf{D}(\mathbf{r}, \omega) = \epsilon_0 \underline{\underline{\epsilon}}(z) \cdot \left[\underline{\underline{\epsilon}}_r(\omega) \cdot \underline{\underline{S}}^T(z) \cdot \mathbf{E}(\mathbf{r}, \omega) \right. \\ \left. + \underline{\underline{\alpha}}_r(\omega) \cdot \underline{\underline{S}}^T(z) \cdot \mathbf{H}(\mathbf{r}, \omega) \right], \end{aligned} \quad (11)$$

$$\begin{aligned} \mathbf{B}(\mathbf{r}, \omega) = \mu_0 \underline{\underline{\mu}}(z) \cdot \left[\underline{\underline{\beta}}_r(\omega) \cdot \underline{\underline{S}}^T(z) \cdot \mathbf{E}(\mathbf{r}, \omega) \right. \\ \left. + \underline{\underline{\mu}}_r(\omega) \cdot \underline{\underline{S}}^T(z) \cdot \mathbf{H}(\mathbf{r}, \omega) \right], \end{aligned} \quad (12)$$

wherein the superscript T denotes the transpose. These equations model the STF as a bianisotropic continuum. Whereas the 3×3 dyadics $\underline{\underline{\epsilon}}_r(\omega)$ and $\underline{\underline{\mu}}_r(\omega)$ represent the local electric and magnetic properties, respectively, the 3×3 dyadics $\underline{\underline{\alpha}}_r(\omega)$ and $\underline{\underline{\beta}}_r(\omega)$ delineate the local magnetoelectric properties. All four of these dyadics have

²The terminology of magnetic fields is historically confused [44]. The correct names are used in this paragraph.

the same form as that given in (6), to which gyrotropic terms can be added [74]. For a purely dielectric STF, both $\underline{\underline{\alpha}}_r(\omega)$ and $\underline{\underline{\beta}}_r(\omega)$ are equal to the null dyadic, and $\underline{\underline{\mu}}_r(\omega)$ is equal to the identity dyadic.

Nominally, all columns in a STF twist and bend identically as z changes, which feature is captured by the rotation dyadic $\underline{\underline{S}}(z)$. This 3×3 dyadic is some composition of the following three elementary rotation 3×3 dyadics:

$$\begin{aligned} \underline{\underline{S}}_x(z) = & \mathbf{u}_x \mathbf{u}_x + (\mathbf{u}_y \mathbf{u}_y + \mathbf{u}_z \mathbf{u}_z) \cos \xi(z) \\ & + (\mathbf{u}_z \mathbf{u}_y - \mathbf{u}_y \mathbf{u}_z) \sin \xi(z), \end{aligned} \quad (13)$$

$$\begin{aligned} \underline{\underline{S}}_y(z) = & \mathbf{u}_y \mathbf{u}_y + (\mathbf{u}_x \mathbf{u}_x + \mathbf{u}_z \mathbf{u}_z) \cos \tau(z) \\ & + (\mathbf{u}_z \mathbf{u}_x - \mathbf{u}_x \mathbf{u}_z) \sin \tau(z), \end{aligned} \quad (14)$$

$$\begin{aligned} \underline{\underline{S}}_z(z) = & \mathbf{u}_z \mathbf{u}_z + (\mathbf{u}_x \mathbf{u}_x + \mathbf{u}_y \mathbf{u}_y) \cos \zeta(z) \\ & + (\mathbf{u}_y \mathbf{u}_x - \mathbf{u}_x \mathbf{u}_y) \sin \zeta(z). \end{aligned} \quad (15)$$

The angular functions of z in these equations may be specified piecewise. Either $\underline{\underline{S}}(z) = \underline{\underline{S}}_x(z)$ or $\underline{\underline{S}}(z) = \underline{\underline{S}}_y(z)$ for SNTFs, while $\underline{\underline{S}}(z) = \underline{\underline{S}}_z(z)$ for TFHBMs. Although TFHBMs need not be periodically nonhomogeneous along the z axis, it is easy to fabricate them with periods chosen anywhere between 50 nm and 2000 nm. Chiral STFs are generally analyzed as periodic dielectric TFHBMs with $\zeta(z) = \pi z / \Omega$, with 2Ω as the structural period [14], [75].

Multisection STFs are cascades of single-section STFs fabricated in an integrated manner [3]. The five dyadics $\underline{\underline{S}}(z)$, $\underline{\underline{\epsilon}}_r(\omega)$, $\underline{\underline{\mu}}_r(\omega)$, $\underline{\underline{\alpha}}_r(\omega)$ and $\underline{\underline{\beta}}_r(\omega)$ are defined differently for each section, the transitions between the sections being virtually abrupt and, therefore, optically inconsequential [68-69].

3.2 Electromagnetic wave propagation

Electromagnetic wave propagation in a STF is best handled using 4×4 matrixes and column 4-vectors. At any given frequency, with the transverse wavenumber κ and the polar angle ψ_{inc} fixed by excitation conditions, the following spatial representation of the electric and the magnetic field phasors is useful:

$$\mathbf{E}(\mathbf{r}, \omega) = \mathbf{e}(z, \kappa, \psi_{inc}, \omega) \exp[i\kappa(x \cos \psi_{inc} + y \sin \psi_{inc})], \quad (16)$$

$$\mathbf{H}(\mathbf{r}, \omega) = \mathbf{h}(z, \kappa, \psi_{inc}, \omega) \exp[i\kappa(x \cos \psi_{inc} + y \sin \psi_{inc})]. \quad (17)$$

Substitution of the foregoing representation into the source-free Maxwell curl postulates, $\nabla \times \mathbf{E}(\mathbf{r}, \omega) = i\omega \mathbf{B}(\mathbf{r}, \omega)$ and $\nabla \times \mathbf{H}(\mathbf{r}, \omega) = -i\omega \mathbf{D}(\mathbf{r}, \omega)$, followed by the use of the constitutive relations (11) and (12), leads to four ordinary differential equations and two algebraic equations. The phasor components $e_z(z, \kappa, \psi_{inc}, \omega)$

and $h_z(z, \kappa, \psi_{inc}, \omega)$ are then eliminated to obtain the 4×4 matrix differential equation [74]

$$\frac{d}{dz}[\mathbf{f}(z, \kappa, \psi_{inc}, \omega)] = i[\mathbf{P}(z, \kappa, \psi_{inc}, \omega)][\mathbf{f}(z, \kappa, \psi_{inc}, \omega)]. \quad (18)$$

In this equation,

$$[\mathbf{f}(z, \kappa, \psi_{inc}, \omega)] = [e_x(z, \kappa, \psi_{inc}, \omega), e_y(z, \kappa, \psi_{inc}, \omega), h_x(z, \kappa, \psi_{inc}, \omega), h_y(z, \kappa, \psi_{inc}, \omega)]^T \quad (19)$$

is a column 4-vector, and $[\mathbf{P}(z, \kappa, \psi_{inc}, \omega)]$ is a 4×4 matrix function of z that can be easily obtained using symbolic manipulation programs. The 4×4 system reduces to two autonomous 2×2 systems for single-section SNTFs, in certain instances [76].

Analytic solution of (18) can be obtained, provided $[\mathbf{P}(z, \kappa, \psi_{inc}, \omega)]$ is not a function of z (i.e., $[\mathbf{P}(z, \kappa, \psi_{inc}, \omega)] = [\mathbf{P}_{con}(\kappa, \psi_{inc}, \omega)]$). This happens, of course, for CTFs, and the solution procedure is described by Hochstadt [77]. Exact analytic solution of (18) has been obtained also for axial propagation (i.e., $\kappa = 0$) in periodic TFHBMs [75], [78-79]. A solution in terms of a convergent matrix polynomial series is available for oblique propagation (i.e., $\kappa \neq 0$) in periodic TFHBMs [80]-[83].

More generally, only a numerical solution of (18) can be obtained. Suppose that $[\mathbf{P}(z, \kappa, \psi_{inc}, \omega)]$ is a periodic function of z . Then, a perturbative approach [84] can be used to obtain simple results for weakly anisotropic STFs [80], [85], while a coupled-mode approach may come handy if otherwise [86]-[88]. But if $[\mathbf{P}(z, \kappa, \psi_{inc}, \omega)]$ is not periodic, the constitutive dyadics can be assumed as piecewise constant—i.e., constant over slices of thickness Δz —and the approximate transfer equation [74]

$$[\mathbf{f}(z + \Delta z, \kappa, \psi_{inc}, \omega)] \simeq \exp\left\{i\left[\mathbf{P}\left(z + \frac{\Delta z}{2}, \kappa, \psi_{inc}, \omega\right)\right]\Delta z\right\}[\mathbf{f}(z, \kappa, \psi_{inc}, \omega)] \quad (20)$$

can be suitably manipulated with appropriately small values of Δz . Specialization to chiral STFs is also available [81].

Finally, quasi-static solutions of (18) can be obtained in the same ways, after taking the limit $\omega \rightarrow 0$ *ab initio* [89]. These are useful if applications of STFs in the microwave and lower-frequency regimes are desired—for examples, as interlayer dielectrics in integrated electronic circuits [90-91] and for humidity sensors that rely on capacitance change induced by altered humidity [92].

3.3 Structure-property relationships

Howsoever is it carried out, the solution of (18) paves the way for predicting electromagnetic phenomenons as well as for designing STF devices that are compatible

with planar technology. There is one more theoretical issue to be cleared up, before optical and infrared applications of STF can be discussed.

Equations (11) and (12) incorporate the assumption of a STF as a unidirectionally nonhomogeneous continuum. This is valid in a macroscopic sense, i.e., in the visible and the infrared frequency regimes, because STFs with appropriate morphological length scales can be fabricated. However, the relationship of the nanostructure to the macroscopic properties must be understood—or at least modeled—for intelligent design and fabrication of STF devices.

A simple model was proposed by Lakhtakia *et al.* [93], based on the concept of local homogenization of electrically small, ellipsoidal particles. The deposited material as well as the void regions are nominally thought of as parallel ellipsoids in any xy plane. The Bruggeman formalism [34] is then used to estimate $\underline{\epsilon}_r(\omega)$, etc., in terms of the shape factors of the ellipsoids, the bulk constitutive properties of the deposited material, and the porosity of the STF. The possibility of the void regions being infiltrated by some material can also be handled [94–95]. Furthermore, this model can incorporate the frequency-dependence of the constitutive properties of the deposited materials [96].

Sherwin and Lakhtakia [96] used the devised model to extensively study the planewave responses of dispersive chiral STFs on axial excitation, studying in particular the spectrums of various measures of transmission optical activity. The dependencies of these quantities on the column inclination angle, the periodicity, porosity, and the ellipsoidal shape factors were deduced. The model requires calibration against experimentally obtained reflectance/transmittance data [97].

The identified functional relationships should assist in design of STF devices as well as the on-line monitoring of STF fabrication processes. When fully coupled with (6) as well as (7)–(10), and after calibration against experimental data, the ellipsoidal model is expected to evolve into a powerful design tool and process-control paradigm. It has already been applied to assess the piezoelectric tunability of lasers and filters made of chiral STFs [98–99].

4 Applications of STFs

4.1 Accomplishments

Although many applications were forecast early on [3], the potential of STFs has been most successfully actualized in linear optics thus far. Several types of optical filters, sensors, photonic bandgap (PBG) materials, and electrically addressable displays are in various stages of development, but are definitely past their embryonic stages.

4.1.1 Optical filters

Chiral STFs must display the circular Bragg phenomenon in accordance with their periodic nonhomogeneity along the z axis [14], [81]. Briefly, a structurally right/left-handed chiral STF only a few periods thick almost completely reflects

normally incident right/left circularly polarized (RCP/LCP) light with wavelength lying in the so-called Bragg regime; while the reflection of normally incident LCP/RCP light in the same regime is very little. Figure 7 presents the measured and the predicted transmittances of a structurally left-handed chiral STF made of titanium oxide, showing the almost complete blockage of incident LCP light and the high transmission of incident RCP light at free-space wavelengths in the neighborhood of 620 nm.

The bandwidth of the Bragg regime and the peak reflectivity therein first increase with the thickness of the chiral STF, and then saturate. Once this saturation has occurred, further thickening of the film has negligible effects on the reflection spectrum [14]. The Bragg regime is also marked by high levels of optical activity [14], [57], [65] which, however, does not scale with the film thickness and is also highly dependent on the orientation of the incident electric field phasor [101].

More than one Bragg regime is possible when light is obliquely incident (i.e., $\kappa \neq 0$) [81], but it is the normal-incidence case that appears to be of the greatest value in the context of planar technology. The major successes reported are as follows:

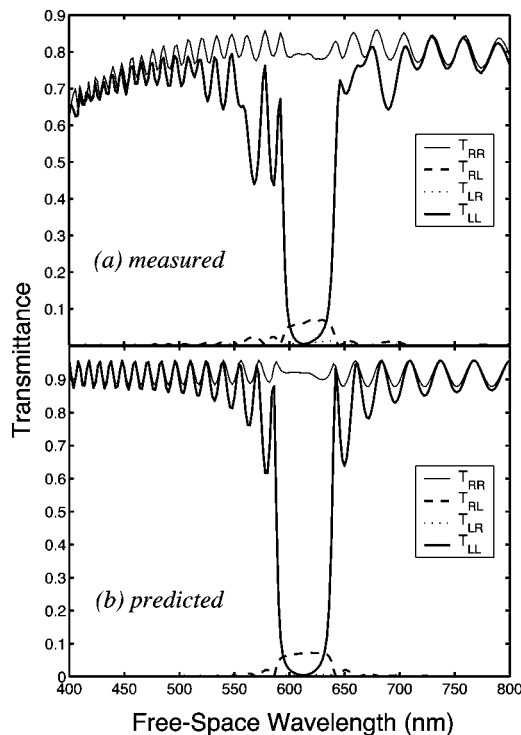


Figure 7 Measured and predicted transmittance spectra of a structurally left-handed chiral STF for normal incidence ($\kappa = 0$). The transmittance T_{LR} is the intensity of the LCP component of the transmitted plane wave relative to the intensity of the RCP component of the incident plane wave, etc. Dispersion was not taken into account when predicting the transmittances from the solution of (18). (Adapted from Wu *et al.* [100].)

- *Circular polarization filters and laser mirrors.* The circular Bragg phenomenon can be employed to realize circular polarization filters. Normally incident, circularly polarized light of one handedness can be reflected almost completely, while that of the other handedness is substantially transmitted, if absorption is small enough and the film is sufficiently thick, in the Bragg regime. This has been experimentally demonstrated [100]. As of now, the Bragg regime can be positioned at virtually any free-space wavelength between 450 and 1700 nm. Calculations show that polarization-insensitivity can be realized with a cascade of two otherwise identical chiral STF's but of opposite structural handedness for laser mirrors [102–103]. Furthermore, stepwise chirping can widen the bandwidth [104], and tightly interlaced chiral STF's may be attractive for bandwidth engineering [105]. Finally, dispersive characteristics can allow for more than one Bragg regime to be possible [106], as exemplified by the calculated reflectance spectrums shown in Figure 8.
- *Polarization-discriminatory handedness-inverters.* A handedness-inverter for light of only one of the two circular polarization states was designed [107], and then fabricated as well as tested [108]. As the first two-section STF device, it comprises a chiral STF and a CTF functioning as a half-wave plate. Basically, it almost completely reflects, say, LCP light, while it substantially transmits incident RCP light after transforming it into LCP light, in the Bragg regime.

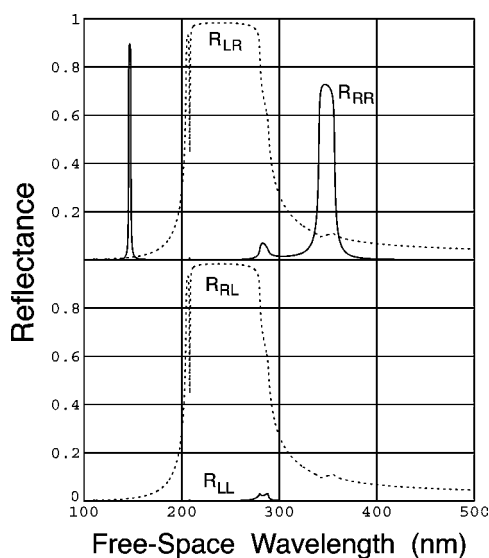


Figure 8 Calculated reflectance spectrums of a structurally right-handed chiral STF half-space for normal incidence ($\kappa = 0$). The reflectance R_{LR} is the intensity of the LCP component of the reflected plane wave relative to the intensity of the RCP component of the incident plane wave, etc. Dispersion is responsible for the circular-polarization-sensitive Bragg regimes centered at 147 and 349 nm wavelengths. (Adapted from Wang *et al.* [106].)

- *Spectral hole filters.* In a further bid to illustrate the potential of the STF a three-section STF was proposed as a spectral hole filter. Its first and third layers are identical chiral STFs, while the thin middle layer is homogeneous [109-110]. The middle layer is supposed to act as a phase defect. This design was actually implemented to obtain a 11-nm-wide spectral hole centered at a free-space wavelength of 580 nm [111]. A better design became available shortly thereafter, wherein the middle layer was eliminated, but the lower chiral STF was twisted by 90° with respect to the upper chiral STF about the z axis. The twist performed satisfactorily as the required phase defect [68].

Most recently, slanted chiral STFs have been introduced [112] to couple the circular Bragg phenomenon to the Rayleigh–Wood anomalies exhibited by surface-relief gratings [113]. This coupling occurs when the helicoidal axis is inclined with respect to the z axis, and suggests the use of these new types of STFs as narrowband circular polarization beamsplitters.

SNTFs can also be pressed into service as optical filters—for linearly polarized plane waves. Rugate filters have been realized as piecewise uniform SNTFs to function as narrowband reflectors [15]. Solc filters of the *fan* and the *folded* types are also possible with the same technology [36], [114]. The major issue for further research and development is the control of mass density and, hence, $\underline{\epsilon}_r(\omega)$ with χ_v when fabricating continuously nonhomogeneous SNTFs.

The future of multisection STF devices in optics appears bright, because of the recent feat of Suzuki and Taga [69] in being able to deposit a cascade of six different sections of combined thickness $\sim 2 \mu\text{m}$.

4.1.2 Optical fluid sensors

The porosity of STFs makes them attractive for fluid concentration sensing applications [115-116] because their optical response properties must change in accordance with the number density of infiltrant molecules. In particular, theoretical research has shown that the Bragg regime of a chiral STF must shift accordingly, thereby providing a measure of the fluid concentration [115]. Qualitative support for this finding is provided by experiments on wet and dry chiral STFs [117]. Furthermore, STF spectral hole filters can function as highly sensitive fluid concentration sensors; and proof-of-concept experiments with both circularly polarized and unpolarized incident light have confirmed the red-shift of spectral holes upon exposure to moisture [95].

4.1.3 Chiral PBG materials

Chiral STFs have been grown on regular lattices by lithographically patterning the substrates [66], [118]. Whereas slow substrate rotation results in the growth of arrays of nano/micro-helices spaced as close as 20 nm from their nearest neighbors, faster rotation yields arrays of increasingly denser pillars [30]. Such STFs are essentially PBG materials in the visible and the infrared regimes [119-120], and

the possibility of fabricating them on cheap polymeric substrates is very attractive [121].

4.1.4 Electrically addressable displays

Liquid crystals (LCs) can be electrically addressed and are therefore widely used these days for displays [122-123]. Although STFs are not electronically addressable, the alignment of nematic LCs forced into the void regions of chiral STFs has been shown to respond to applied voltages [124]. Thus, STF-LC composites may have a future as robust displays.

Another interesting possibility, in the same vein, is to grow carbon (and other) nanotubes by chemical reactions involving fluid catalysts and precursors [125] inside highly porous STFs. The growing nanotubes would have to conform to the structure imposed by the STF skeleton, and the nanotube-STF composite thus formed could be useful for field emission devices.

4.2 Emerging applications

The chief emerging applications are mostly optical, but electronic, acoustic, thermal, chemical and biological applications have also been conceived and examined to some extent.

4.2.1 Optical interconnects

Efficient use of optoelectronic devices is facilitated by optical interconnects which, in addition to providing effective signal transmission, must be simple to fabricate on integrated circuitry. STF technology is compatible with the planar technology of electronic chips. Guided wave propagation in chiral STFs turns out to yield the space-guide concept: the capacity to simultaneously support propagation modes with different phase velocities in different directions [126-127]. This could result in efficient use of the available *real estate* in electronic chips. Furthermore, the helicoidal structure of chiral STFs would resist vertical cleavage and fracture commonly seen to occur in regular crystals along planes of weakness [128], in contrast to CTFs which can also function as space-guides. Simultaneous microrefrigeration enabled by the porous STFs to remove thermal energy generated within electronic chips would be a bonus.

4.2.2 Optical pulse-shapers

The current explosive growth of digital optics communication has provided impetus for time-domain research on novel materials. As chiral STFs are very attractive for optical applications, the circular Bragg phenomenon is now being studied in the time domain. A pulse-bleeding phenomenon has been identified as the underlying mechanism, which can drastically affect the shapes, amplitudes and spectral components of femtosecond pulses [129]. However, narrowband rectangular pulses can pass through without significant loss of information [130]. Application of STFs to shape optical pulses appears to be waiting in the wings.

4.2.3 Biochips

Endowed with porosity of engineered texture, STF's can function as biochips as follows: Suppose that identical single-stranded DNA molecules—with a particular genomic sequence matched to, say, *E. coli*—are dispersed in a STF. A drop of contaminated water containing analyte DNA molecules from exploded *E. coli* is put on the thus-prepared STF, followed by a drop of an appropriate ruthenium complex. The resulting bioluminescence signal from the STF can be optically sensed by photon counters in order to measure the degree of contamination.

Bioluminescent emission is bound to be affected by the reactor characteristics. If the reactor is a chiral STF, its helicoidal periodicity can be exploited. Indeed, the structural handedness as well as the periodicity of chiral STF's have been theoretically shown to critically control the emission spectrum and intensity, while the polarization state of the emitted light is strongly correlated with the structural handedness of the embedded source filaments [131]. Most recently, a pseudo-isotropic point and a zone of high emission efficiencies have been identified with respect to χ_v [132–133]. Bioluminescence STF sensors therefore merit closer attention.

4.2.4 Interlayer dielectrics

With the microelectronics industry moving relentlessly towards decreasing feature sizes and increasingly stringent tolerance levels, an urgent need exists for the use of low-permittivity materials as interlayer dielectrics. Silicon dioxide, the current material of choice, has too high a permittivity. The porosity of STF's and nanoporous silica makes them attractive low-permittivity materials for microelectronic and electronic packaging applications [91]. However, chiral STF's are likely to have significant thermal, mechanical, as well as electrical advantages over nanoporous silica—because of (i) porosity with controllable texture and (ii) helicoidal morphology. Also, STF's can be impregnated with various kinds of polymers.

4.2.5 Ultrasonic applications

The sciences of electromagnetics and elastodynamics have an underlying mathematical unity. For that reason, all optical and electronic applications described thus far in Section 4 possess ultrasonic analogs. Indeed, ultrasonic wave propagation in chiral STF's is theoretically well established [134]–[136], as also is the potential for its applications [137–138]. Actual implementation would, however, require the fabrication of chiral STF's with periods $\sim 20\ \mu\text{m}$ [139], which development is still awaited.

4.2.6 Other applications

From their inception [3], STF's were expected to have a wide range of applications, implementable after their properties came to be better understood. Their optical applications came to be investigated first, as detailed in Section 4. However, their high porosity—in combination with optical anisotropy and possible two-dimensional electron confinement in the nanostructure—makes STF's potential candidates also

as (i) electroluminescent devices; (ii) high-speed, high-efficiency electrochromic films; (iii) optically transparent conducting films sculptured from pure metals; and (iv) multistate electronic switches based on filamentary conduction. That same porosity can be harnessed in microreactors and thermal barriers, as it is accompanied by high surface area [70-71], [140]. For the same reason, STFs may be useful as nano/micro-sieves for the entrapment of viruses or for growing biological tissues on surfaces of diverse provenances.

5 Future research directions

The emerging applications mentioned in Section 4.2 are all in their incipient stages at best, with several barely past conceptualization. Considerable research on them is warranted, before they become commercially viable. Just a few of the accomplishments discussed in Section 4.1 have crossed the threshold of academic research, and now require several issues to be addressed.

A key issue is that of environmental degradability of STFs. The chemical stability of STFs has not been examined in any detail yet, although the susceptibility of porous thin films to moisture is known [117]. An indentation experiment on a chiral STF [141] as well as the successful deposition of six-section STFs [69] strongly indicate that mechanical stability must be investigated in depth. However, only a preliminary model for the mechanical loading of STFs exists at this time [142-143]. Due to porosity, internal stresses and the morphological stability of STFs in the absence of external loads have to be examined carefully as well [144-145].

Another key issue is that of efficiency. The vapor incidence angle χ_v , the bulk properties of the deposited material (responsible, e.g., for ϵ_a , ϵ_b and ϵ_c) and the substrate rotation parameters appearing in $\underline{S}(z)$ must be optimized to achieve desired performance characteristics. As examples, the photocatalytic efficiency of chiral STFs of tantalum oxide is known to be optimal when $\chi_v = 20^\circ$ [140], efficient bioluminescent emission has been shown to require $\chi_v \lesssim 15^\circ$ [132], while χ_v could be manipulated in order to maximize the bandwidth of a Bragg regime [133]. A study on second harmonic generation in uniaxial chiral STFs has underscored the criticality of χ (and therefore of χ_v) for efficiency [146]. Cross-polarized remittances are drastically reduced and the diversity in their copolarized remittances is enhanced by the incorporation of index-matched layers at the entry and the exit pupils of circular polarization filters [100]. Further improvements may require the simultaneous deposition of different types of materials in order to reduce absorption and dispersion in desired wavelength ranges in optical filters based on the STF concept.

Nonlinear optics with STFs is practically virgin territory, despite two reported forays into second harmonic generation [146-147]. Due to the numerous classes of nonlinearity [148], the delineation of nanocrystallinity in STFs will be of primary importance. Likewise, understanding of nanodomains in magnetic STFs, as well as of magnetoelectric effects in bianisotropic STFs, are topics of future research.

Although the demonstrated successes of the STF concept and technology are few as yet, the electromagnetic and elastodynamic frameworks for STFs are reasonably mature. They have sufficient allure that committed researchers shall continue, and new researchers will be attracted by their novelty and challenge. Section 4 and the foregoing paragraphs, as well as another chapter in this book [73], point to many avenues for further research.

But in order for STF research and use to be truly widespread, economical production must be enabled. Any satisfactory production technique must be rapid and deliver high yields, so that large-scale fabrication must become possible. The latter appears feasible with the adaptation of ion-thruster technology [149]. Furthermore, the films will have to be laterally uniform with growth evolution (i.e., possess Zone M morphology), and χ may have to be lower than 20° . If PVD (or any variant) is to be industrially successful, then new architectures for the evaporant flux source—whether discrete or continuous, single or multiple—must be developed to deposit STFs on large substrates. Reliability of deposition uniformity would be facilitated by computer-controlled source architectures. In turn, they will require the development of *in situ* monitoring of the deposition process and appropriate control models. These and related avenues for manufacturing research need to be opened up.

Acknowledgments

We dedicate this chapter to N.O. Young and J. Kowal, prescient STF researchers who came about 30 years before the time was ripe for them. Our thanks are due to Dr. V.C. Venugopal for many constructive suggestions to improve this chapter.

References

- [1] National Research Council, *Condensed-matter and materials physics: Basic research for tomorrow's technology*, National Academies Press, Washington, DC, USA, 1999.
- [2] A. Lakhtakia and R. Messier, The key to a thin film HBM: the Motohiro–Taga interface, *Proceedings of Chiral '94: 3rd international workshop on chiral, bi-isotropic and bi-anisotropic media* (F. Mariotte and J.-P. Parneix, eds), Périgueux, France, 1994, 125–130.
- [3] A. Lakhtakia, R. Messier, M.J. Brett and K. Robbie, Sculptured thin films (STFs) for optical, chemical and biological applications, *Innovat Mater Res* **1** (1996), 165–176.
- [4] A. Lakhtakia and R. Messier, Sculptured thin films—I. Concepts, *Mater Res Innovat* **1** (1997), 145–148.
- [5] R. Messier and A. Lakhtakia, Sculptured thin films—II. Experiments and applications, *Mater Res Innovat* **2** (1999), 217–222.

- [6] A. Kundt, Ueber Doppelbrechung des Lichtes in Metallschichten, welche durch Zerstäuben einer Kathode hergestellt sind, *Ann Phys Chem Lpz* **27** (1886), 59–71.
- [7] H. König and G. Helwig, Über die Struktur schräg aufgedampfter Schichten und ihr Einfluß auf die Entwicklung submikroskopischer Oberflächenrauigkeiten, *Optik* **6** (1950), 111–124.
- [8] L. Holland, The effect of vapor incidence on the structure of evaporated aluminum films, *J Opt Soc Am* **43** (1953), 376–380.
- [9] N.O. Young and J. Kowal, Optically active fluorite films, *Nature* **183** (1959), 104–105.
- [10] D.O. Smith, M.S. Cohen and G.P. Weiss, Oblique-incidence anisotropy in evaporated permalloy films, *J Appl Phys* **31** (1960), 1755–1762.
- [11] W. Metzdorf and H.E. Wiehl, Negative oblique-incidence anisotropy in magnetostriction-free permalloy films, *Phys Stat Sol* **17** (1966), 285–294.
- [12] J.M. Nieuwenhuizen and H.B. Haanstra, Microfractography of thin films, *Philips Tech Rev* **27** (1966), 87–91.
- [13] T. Motohiro and Y. Taga, Thin film retardation plate by oblique deposition, *Appl Opt* **28** (1989), 2466–2482.
- [14] V.C. Venugopal and A. Lakhtakia, Sculptured thin films: conception, optical properties and applications, *Electromagnetic fields in unconventional materials and structures* (O.N. Singh and A. Lakhtakia, eds), Wiley, New York, NY, USA, 2000.
- [15] I.J. Hodgkinson and Q.h. Wu, Inorganic chiral optical materials, *Adv Mater* **13** (2001), 889–897.
- [16] A. Lakhtakia and R.F. Messier (eds), *Engineered nanostructural films and materials*, Proc SPIE **3790**, Bellingham, WA, USA, 1999.
- [17] A. Lakhtakia, W.S. Weiglhofer and R.F. Messier (eds), *Complex mediums*, Proc SPIE **4097**, Bellingham, WA, USA, 2000.
- [18] A. Lakhtakia, W.S. Weiglhofer and I.J. Hodgkinson (eds), *Complex mediums II: Beyond linear isotropic dielectrics*, Proc SPIE **4467**, Bellingham, WA, USA, 2001.
- [19] A. Lakhtakia, G. Dewar and M.W. McCall (eds), *Complex mediums III: Beyond linear isotropic dielectrics*, Proc SPIE **4806**, Bellingham, WA, USA, 2002.
- [20] R. Messier, A.P. Giri and R.A. Roy, Revised structure zone model for thin film physical structure, *J Vac Sci Technol A* **2** (1984), 500–503.
- [21] J.M. García-Ruiz, A. Lakhtakia and R. Messier, Does competition between growth elements eventually eliminate self-affinity? *Speculat Sci Technol* **15** (1992), 60–71.
- [22] R.F. Messier, P. Sunal and V.C. Venugopal, Evolution of sculptured thin films, *Engineered nanostructural films and materials* (A. Lakhtakia and R.F. Messier, eds), Proc SPIE **3790**, Bellingham, WA, USA, 1999, 133–141.
- [23] R.A. Roy and R. Messier, Evolutionary growth development in SiC sputtered films, *Mater Res Soc Symp Proc* **38** (1985), 363–370.

- [24] R.A. Roy and R. Messier, Quantitative analysis of thin film morphology evolution, *High performance ceramic films and coatings* (P. Vincenzini, ed), Elsevier, Amsterdam, The Netherlands, 1991.
- [25] R. Messier, Toward quantification of thin film morphology, *J Vac Sci Technol A* **4** (1986), 490–495.
- [26] P. Meakin, Fractal scaling in thin film condensation and material surfaces, *CRC Crit Rev Solid State Mater Sci* **13** (2) (1987), 143–189.
- [27] K.-H. Muller, Stress and microstructure of sputter-deposited thin films—molecular dynamics investigation, *J Appl Phys* **62** (1987), 1796–1799.
- [28] J.A. Thornton, High rate thick film growth, *Annu Rev Mater Sci* **7** (1977), 239–260.
- [29] B. Yang, B.L. Walden, R. Messier and W.B. White, Computer simulation of the cross-sectional morphology of thin films, *Modeling of optical thin films* (M.R. Jacobson, ed), Proc SPIE **821**, Bellingham, WA, USA, 1987, 68–76.
- [30] R. Messier, V.C. Venugopal and P.D. Sunal, Origin and evolution of sculptured thin films, *J Vac Sci Technol A* **18** (2000), 1538–1545.
- [31] R. Messier and S. Trolier-McKinstry, Processing of ceramic thin films, *Encyclopedia of materials: Science and technology* (K.H.I. Buschow, R.W. Cahir, M.C. Flemings, B. Ilschner, E.J. Kramer and S. Mahajan, eds), Elsevier, Amsterdam, The Netherlands, 2001.
- [32] S. Hamaguchi and S.M. Rossnagel, Surface-topography simulations of ionized sputter metal deposition, *Mater Res Soc Symp Proc* **389** (1995), 113–117.
- [33] H. van Kranenberg and C. Lodder, Tailoring growth and local composition by oblique-incidence deposition: a review and new experimental data, *Mater Sci Engg R* **11** (1994), 295–354.
- [34] A. Lakhtakia (ed), *Selected papers on linear optical composite materials*, SPIE Press, Bellingham, WA, USA, 1996.
- [35] O.S. Heavens, *Optical properties of thin solid films*, Butterworths, London, UK, 1955.
- [36] I.J. Hodgkinson and Q.-h. Wu, *Birefringent thin films and polarizing elements*, World Scientific, Singapore, 1997.
- [37] G.B. Smith, Effective medium theory and angular dispersion of optical constants in films with oblique columnar structure, *Opt Commun* **71** (1989), 279–284.
- [38] H. Wang, Assessment of optical constants of multilayer thin films with columnar-structure-induced anisotropy, *J Phys D: Appl Phys* **28** (1995), 571–575.
- [39] H. Wang, Spectrophotometric measurements on a titanium dioxide film with columnar-structure-induced anisotropy, *Fiber Integrat Opt* **15** (1996), 37–44.
- [40] H. Wang, Reflection/transmission measurements of anisotropic films with one of the principal axes in the direction of columnar growth, *J Modern Opt* **42** (1995), 497–505.

- [41] I.J. Hodgkinson, F. Horowitz, H.A. Macleod, M. Sikkens and J.J. Wharton, Measurement of the principal refractive indices of thin films deposited at oblique incidence, *J Opt Soc Am A* **2** (1985), 1693–1697.
- [42] I.J. Hodgkinson, Q.h. Wu and J. Hazel, Empirical equations for the principal refractive indices and column angle of obliquely deposited films of tantalum oxide, titanium oxide, and zirconium oxide, *Appl Opt* **37** (1998), 2653–2659.
- [43] H.C. Chen, *Theory of electromagnetic waves*, TechBooks, Fairfax, VA, USA, 1992.
- [44] W.S. Weiglhofer, Constitutive characterization of simple and complex mediums, *Introduction to complex mediums for optics and electromagnetics* (W.S. Weiglhofer and A. Lakhtakia, eds), SPIE Press, Bellingham, WA, USA, 2003.
- [45] A. Lakhtakia and W.S. Weiglhofer, Significance of cross-sectional morphology for Motohiro–Taga interfaces, *Optik* **110** (1999), 33–36.
- [46] I. Abdulhalim, Point of ultra-sensitivity to perturbations for axial propagation in helicoidal bianisotropic structures, *Europhys Lett* **48** (1999), 177–181.
- [47] A. Lakhtakia and W.S. Weiglhofer, A comparative study of planewave propagation in helicoidal bianisotropic mediums and isotropic chiral mediums, *J Opt A: Pure Appl Opt* **2** (2000), 107–111.
- [48] C. Kittel, *Introduction to solid state physics*, Wiley Eastern, New Delhi, India, 1971.
- [49] I.J. Hodgkinson and Q.h. Wu, Serial bideposition of anisotropic thin films with enhanced linear birefringence, *Appl Opt* **38** (1999), 3621–3625.
- [50] A. Lakhtakia, Spectral signatures of axially excited slabs of dielectric thin-film helicoidal bianisotropic mediums, *Eur Phys J Appl Phys* **8** (1999), 129–137.
- [51] H.A. Macleod, *Thin-film optical filters*, Institute of Physics Publishing, Bristol, UK, 2001.
- [52] M. Suzuki, S. Tokito and Y. Taga, Review of thin film technology in automobile industry, *Mater Sci Engg B* **51** (1998), 66–71.
- [53] Y. Taga, Review of plasma thin-film technology in automobile industry, *Surf Coat Technol* **112** (1999), 339–346.
- [54] M. Suzuki, T. Ito and Y. Taga, Recent progress of obliquely deposited thin films for industrial applications, *Engineered nanostructural films and materials* (A. Lakhtakia and R.F. Messier, eds), Proc SPIE **3790**, Bellingham, WA, USA, 1999, 94–105.
- [55] E.F. Dawson and N.O. Young, Helical Kerr cell, *J Opt Soc Am* **50** (1960), 170–171.
- [56] K. Robbie, M.J. Brett and A. Lakhtakia, First thin film realization of a helicoidal bianisotropic medium, *J Vac Sci Technol A* **13** (1995), 2991–2993.
- [57] K. Robbie, M.J. Brett and A. Lakhtakia, Chiral sculptured thin films, *Nature* **384** (1996), 616.

- [58] P.I. Rovira, R.A. Yarussi, R.W. Collins, R. Messier, V.C. Venugopal, A. Lakhtakia, K. Robbie and M.J. Brett, Transmission ellipsometry of a thin-film helicoidal bianisotropic medium, *Appl Phys Lett* **71** (1997), 1180–1182.
- [59] K. Robbie and M.J. Brett, Sculptured thin films and glancing angle deposition: Growth mechanics and applications, *J Vac Sci Technol A* **15** (1997), 1460–1465.
- [60] R. Messier, T. Gehrke, C. Frankel, V.C. Venugopal, W. Otaño and A. Lakhtakia, Engineered sculptured nematic thin films, *J Vac Sci Technol A* **15** (1997), 2148–2152.
- [61] J.C. Sit, D. Vick, K. Robbie and M.J. Brett, Thin film microstructure control using glancing angle deposition by sputtering, *J Mater Res* **14** (1999), 1197–1199.
- [62] K. Robbie, J.C. Sit and M.J. Brett, Advanced techniques for glancing angle deposition, *J Vac Sci Technol B* **16** (1998), 1115–1122.
- [63] O.R. Monteiro, A. Vizir and I.G. Brown, Multilayer thin-films with chevron-like microstructure, *J Phys D: Appl Phys* **31** (1998), 3188–3196.
- [64] F. Liu, M.T. Umlor, L. Shen, W. Eads, J.A. Barnard and G.J. Mankey, The growth of nanoscale structured iron films by glancing angle deposition, *J Appl Phys* **85** (1999), 5486–5488.
- [65] I. Hodgkinson, Q.h. Wu, B. Knight, A. Lakhtakia and K. Robbie, Vacuum deposition of chiral sculptured thin films with high optical activity, *Appl Opt* **39** (2000), 642–649.
- [66] M. Malac and R.F. Egerton, Observations of the microscopic growth mechanism of pillars and helices formed by glancing-angle thin-film deposition, *J Vac Sci Technol A* **19** (2001), 158–166.
- [67] Y.-P. Zhao, D.-X. Ye, G.-C. Wang and T.-M. Lu, Novel nano-column and nano-flower arrays by glancing angle deposition, *Nano Lett* **2** (2002), 351–354.
- [68] I.J. Hodgkinson, Q.H. Wu, K.E. Thorn, A. Lakhtakia and M.W. McCall, Spacerless circular-polarization spectral-hole filters using chiral sculptured thin films: theory and experiment, *Opt Commun* **184** (2000), 57–66.
- [69] M. Suzuki and Y. Taga, Integrated sculptured thin films, *Jap J Appl Phys Part 2* **40** (2001), L358–L359.
- [70] K.D. Harris, M.J. Brett, T.J. Smy and C. Backhouse, Microchannel surface area enhancement using porous thin films, *J Electrochem Soc* **147** (2000), 2002–2006.
- [71] K.D. Harris, D. Vick, E.J. Gonzalez, T. Smy, K. Robbie and M.J. Brett, Porous thin films for thermal barrier coatings, *Surf Coat Technol* **138** (2001), 185–191.
- [72] Y. Taga and M. Suzuki, Automobile applications of optical thin films modified by nanotechnology, *Complex mediums II: Beyond linear isotropic dielectrics* (A. Lakhtakia, W.S. Weiglhofer and I.J. Hodgkinson, eds), Proc SPIE **4467**, Bellingham, WA, USA, 2001, 102–108.

- [73] M.W. McCall, Towards optoelectronic applications of chiral sculptured thin films, *Introduction to complex mediums for optics and electromagnetics* (W.S. Weiglhofer and A. Lakhtakia, eds), SPIE Press, Bellingham, WA, USA, 2003.
- [74] A. Lakhtakia, Director-based theory for the optics of sculptured thin films, *Optik* **107** (1997), 57–61.
- [75] A. Lakhtakia and W.S. Weiglhofer, Axial propagation in general helicoidal bianisotropic media, *Microw Opt Technol Lett* **6** (1993), 804–806.
- [76] A. Lakhtakia, Linear optical responses of sculptured thin films (SNTFs), *Optik* **106** (1997), 45–52.
- [77] H. Hochstadt, *Differential equations—A modern approach*, Dover Press, New York, NY, USA, 1975.
- [78] A. Lakhtakia and W.S. Weiglhofer, On light propagation in helicoidal bianisotropic mediums, *Proc Roy Soc Lond A* **448** (1995), 419–437; erratum: **454** (1998), 3275.
- [79] A. Lakhtakia, Anomalous axial propagation in helicoidal bianisotropic media, *Opt Commun* **157** (1998), 193–201.
- [80] A. Lakhtakia and W.S. Weiglhofer, Further results on light propagation in helicoidal bianisotropic mediums: oblique propagation, *Proc Roy Soc Lond A* **453** (1997), 93–105; erratum: **454** (1998), 3275.
- [81] V.C. Venugopal and A. Lakhtakia, Electromagnetic plane-wave response characteristics of non-axially excited slabs of dielectric thin-film helicoidal bianisotropic mediums, *Proc Roy Soc Lond A* **456** (2000), 125–161.
- [82] M. Schubert and C.M. Herzinger, Ellipsometry on anisotropic materials: Bragg conditions and phonons in dielectric helical thin films, *Phys Stat Sol (a)* **188** (2001), 1563–1575.
- [83] J.A. Polo, Jr., and A. Lakhtakia, Numerical implementation of exact analytical solution for oblique propagation in a cholesteric liquid crystal, *Microw Opt Technol Lett* **35** (2002), 397–400. [The two factors on the right side of Eq. (16) of this paper should be interchanged.]
- [84] V.A. Yakubovich and V.M. Starzhinskii, *Linear differential equations with periodic coefficients*, Wiley, New York, NY, USA, 1975.
- [85] W.S. Weiglhofer and A. Lakhtakia, Oblique propagation in a cholesteric liquid crystal: 4×4 matrix perturbational solution, *Optik* **102** (1996), 111–114.
- [86] K. Rokushima and J. Yamakita, Analysis of diffraction in periodic liquid crystals: the optics of the chiral smectic C phase, *J Opt Soc Am A* **4** (1987), 27–33.
- [87] M.W. McCall and A. Lakhtakia, Development and assessment of coupled wave theory of axial propagation in thin-film helicoidal bianisotropic media. Part 1: reflectances and transmittances, *J Modern Opt* **47** (2000), 973–991.
- [88] M.W. McCall and A. Lakhtakia, Development and assessment of coupled wave theory of axial propagation in thin-film helicoidal bi-anisotropic media. Part 2: dichroisms, ellipticity transformation and optical rotation, *J Modern Opt* **48** (2001), 143–158.

- [89] A. Lakhtakia, On the quasistatic approximation for helicoidal bianisotropic mediums, *Electromagnetics* **19** (1999), 513–525.
- [90] A. Lakhtakia, Capacitance of a slab of a dielectric thin-film helicoidal bianisotropic medium, *Microw Opt Technol Lett* **21** (1999), 286–288.
- [91] V.C. Venugopal, A. Lakhtakia, R. Messier and J.-P. Kucera, Low-permittivity materials using sculptured thin film technology, *J Vac Sci Technol B* **18** (2000), 32–36.
- [92] A.T. Wu and M.J. Brett, Sensing humidity using nanostructured SiO posts: mechanism and optimization, *Sens Mater* **13** (2001), 399–431.
- [93] A. Lakhtakia, P.D. Sunal, V.C. Venugopal and E. Ertekin, Homogenization and optical response properties of sculptured thin films, *Engineered nanostructural films and materials* (A. Lakhtakia and R.F. Messier, eds), Proc SPIE **3790**, Bellingham, WA, USA, 1999, 77–83.
- [94] A. Lakhtakia, Enhancement of optical activity of chiral sculptured thin films by suitable infiltration of void regions, *Optik* **112** (2001), 145–148; erratum: **112** (2001), 544.
- [95] A. Lakhtakia, M.W. McCall, J.A. Sherwin, Q.H. Wu and I.J. Hodgkinson, Sculptured-thin-film spectral holes for optical sensing of fluids, *Opt Commun* **194** (2001), 33–46.
- [96] J.A. Sherwin and A. Lakhtakia, Nominal model for structure-property relations of chiral dielectric sculptured thin films, *Math Comput Modelling* **34** (2001), 1499–1514; erratum: **35** (2002), 1355–1363.
- [97] J.A. Sherwin, A. Lakhtakia and I.J. Hodgkinson, On calibration of a nominal structure-property relationship model for chiral sculptured thin films by axial transmittance measurements, *Opt Commun* **209** (2002), 369–375.
- [98] F. Wang, A. Lakhtakia and R. Messier, On piezoelectric control of the optical response of sculptured thin films, *J Modern Opt* **50** (2003), 239–249.
- [99] F. Wang, A. Lakhtakia and R. Messier, Towards piezoelectrically tunable chiral sculptured thin film lasers, *Sens Actuat A: Phys* **102** (2002), 31–35.
- [100] Q. Wu, I.J. Hodgkinson and A. Lakhtakia, Circular polarization filters made of chiral sculptured thin films: experimental and simulation results, *Opt Engg* **39** (2000), 1863–1868.
- [101] V.C. Venugopal and A. Lakhtakia, On optical rotation and ellipticity transformation by axially excited slabs of dielectric thin-film helicoidal bianisotropic mediums (TFHBMs), *Int J Appl Electromag Mech* **9** (1998), 201–210.
- [102] A. Lakhtakia and V.C. Venugopal, Dielectric thin-film helicoidal bianisotropic medium bilayers as tunable polarization-independent laser mirrors and notch filters, *Microw Opt Technol Lett* **17** (1998), 135–140.
- [103] A. Lakhtakia and I.J. Hodgkinson, Spectral response of dielectric thin-film helicoidal bianisotropic medium bilayer, *Opt Commun* **167** (1999), 191–202.
- [104] A. Lakhtakia, Stepwise chirping of chiral sculptured thin films for Bragg bandwidth enhancement, *Microw Opt Technol Lett* **28** (2001), 323–326.

- [105] A. Lakhtakia, Axial excitation of tightly interlaced chiral sculptured thin films: “averaged” circular Bragg phenomenon, *Optik* **112** (2001), 119–124.
- [106] J. Wang, A. Lakhtakia and J.B. Geddes III, Multiple Bragg regimes exhibited by a chiral sculptured thin film half-space on axial excitation, *Optik* **113** (2002), 213–221.
- [107] A. Lakhtakia, Dielectric sculptured thin films for polarization-discriminatory handedness-inversion of circularly polarized light, *Opt Engg* **38** (1999), 1596–1602.
- [108] I.J. Hodgkinson, A. Lakhtakia and Q.h. Wu, Experimental realization of sculptured-thin-film polarization-discriminatory light-handedness inverters, *Opt Engg* **39** (2000), 2831–2834.
- [109] A. Lakhtakia and M. McCall, Sculptured thin films as ultranarrow-bandpass circular-polarization filters, *Opt Commun* **168** (1999), 457–465.
- [110] A. Lakhtakia, V.C. Venugopal and M.W. McCall, Spectral holes in Bragg reflection from chiral sculptured thin films: circular polarization filters, *Opt Commun* **177** (2000), 57–68.
- [111] I.J. Hodgkinson, Q.h. Wu, A. Lakhtakia and M.W. McCall, Spectral-hole filter fabricated using sculptured thin-film technology, *Opt Commun* **177** (2000), 79–84.
- [112] F. Wang, A. Lakhtakia and R. Messier, Coupling of Rayleigh–Wood anomalies and the circular Bragg phenomenon in slanted chiral sculptured thin films, *Eur Phys J Appl Phys* **20** (2002), 91–104.
- [113] D. Maystre (ed), *Selected papers on diffraction gratings*, SPIE Press, Bellingham, WA, USA, 1993.
- [114] A. Lakhtakia, Dielectric sculptured thin films as Šolc filters, *Opt Engg* **37** (1998), 1870–1875.
- [115] A. Lakhtakia, On determining gas concentrations using thin-film helicoidal bianisotropic medium bilayers, *Sens Actuat B: Chem* **52** (1998), 243–250.
- [116] E. Ertekin and A. Lakhtakia, Sculptured thin film Šolc filters for optical sensing of gas concentration, *Eur Phys J Appl Phys* **5** (1999), 45–50.
- [117] I.J. Hodgkinson, Q.H. Wu and K.M. McGrath, Moisture adsorption effects in biaxial and chiral optical thin film coatings, *Engineered nanostructural films and materials* (A. Lakhtakia and R.F. Messier, eds), Proc SPIE **3790**, Bellingham, WA, USA, 1999, 184–194.
- [118] M. Malac and R.F. Egerton, Thin-film regular-array structures with 10–100 nm repeat distance, *Nanotechnology* **12** (2001), 11–13.
- [119] O. Toader and S. John, Proposed square spiral microfabrication architecture for large three-dimensional photonic band gap crystals, *Science* **292** (2001), 1133–1135.
- [120] S.R. Kennedy, M.J. Brett, O. Toader and S. John, Fabrication of tetragonal square spiral photonic crystals, *Nano Lett* **2** (2002), 59–62.
- [121] B. Dick, J.C. Sit, M.J. Brett, I.M.N. Votte and C.W.M. Bastiaansen, Embossed polymeric relief structures as a template for the growth of periodic inorganic microstructures, *Nano Lett* **1** (2001), 71–73.

- [122] S.D. Jacobs (ed), *Selected papers on liquid crystals for optics*, SPIE Press, Bellingham, WA, USA, 1992.
- [123] H. Kawamoto, The history of liquid-crystal displays, *Proc IEEE* **90** (2002), 460–500.
- [124] J.C. Sit, D.J. Broer and M.J. Brett, Liquid crystal alignment and switching in porous chiral thin films, *Adv Mater* **12** (2000), 371–373.
- [125] R. Vajtai, B.Q. Wei, Z.J. Zhang, Y. Jung, G. Ramanath and P.M. Ajayan, Building carbon nanotubes and their smart architectures, *Smart Mater Struct* **11** (2002), 691–698.
- [126] A. Lakhtakia, Towards sculptured thin films (STFs) as optical interconnects, *Optik* **110** (1999), 289–293.
- [127] E. Ertekin and A. Lakhtakia, Optical interconnects realizable with thin-film helicoidal bianisotropic mediums, *Proc Roy Soc Lond A* **457** (2001), 817–836.
- [128] C.D. Gribble and A.J. Hall, *Optical mineralogy: Principles and practice*, UCL Press, London, UK, 1992.
- [129] J.B. Geddes III and A. Lakhtakia, Reflection and transmission of optical narrow-extent pulses by axially excited chiral sculptured thin films, *Eur Phys J Appl Phys* **13** (2001), 3–14; erratum: **16** (2001), 247.
- [130] J.B. Geddes III and A. Lakhtakia, Pulse-coded information transmission across an axially excited chiral sculptured thin film in the Bragg regime, *Microw Opt Technol Lett* **28** (2001), 59–62.
- [131] A. Lakhtakia, On bioluminescent emission from chiral sculptured thin films, *Opt Commun* **188** (2001), 313–320.
- [132] A. Lakhtakia, Local inclination angle: a key structural factor in emission from chiral sculptured thin films, *Opt Commun* **202** (2002), 103–112; erratum: **203** (2002), 447.
- [133] A. Lakhtakia, Pseudo-isotropic and maximum-bandwidth points for axially excited chiral sculptured thin films, *Microw Opt Technol Lett* **34** (2002), 367–371.
- [134] A. Lakhtakia, Wave propagation in a piezoelectric, continuously twisted, structurally chiral medium along the axis of spirality, *Appl Acoust* **44** (1995), 25–37; erratum: **44** (1995), 385.
- [135] A. Lakhtakia, Exact analytic solution for oblique propagation in a piezoelectric, continuously twisted, structurally chiral medium, *Appl Acoust* **49** (1996), 225–236.
- [136] C. Oldano and S. Ponti, Acoustic wave propagation in structurally helical media, *Phys Rev E* **63** (2000), 011703.
- [137] A. Lakhtakia and M.W. Meredith, Shear axial modes in a PCTSCM. Part IV: bandstop and notch filters, *Sens Actuat A: Phys* **73** (1999), 193–200.
- [138] A. Lakhtakia, Shear axial modes in a PCTSCM. Part VI: simpler transmission spectral holes, *Sens Actuat A: Phys* **87** (2000), 78–80.
- [139] A. Lakhtakia, K. Robbie and M.J. Brett, Spectral Green's function for wave excitation and propagation in a piezoelectric, continuously twisted, structurally chiral medium, *J Acoust Soc Am* **101** (1997), 2052–2059.

- [140] M. Suzuki, T. Ito and Y. Taga, Photocatalysis of sculptured thin films of TiO_2 , *Appl Phys Lett* **78** (2001), 3968–3970.
- [141] M.W. Seto, K. Robbie, D. Vick, M.J. Brett and L. Kuhn, Mechanical response of thin films with helical microstructures, *J Vac Sci Technol B* **17** (1999), 2172–2177.
- [142] A. Lakhtakia, Axial loading of a chiral sculptured thin film, *Modelling Simul Mater Sci Engg* **8** (2000), 677–686.
- [143] A. Lakhtakia, Microscopic model for elastostatic and elastodynamic excitation of chiral sculptured thin films, *J Compos Mater* **36** (2002), 1277–1298.
- [144] R. Knepper and R. Messier, Morphology and mechanical properties of oblique angle columnar thin films, *Complex mediums* (A. Lakhtakia, W.S. Weiglhofer and R.F. Messier, eds), Proc SPIE **4097**, Bellingham, WA, USA, 291–298.
- [145] M. Suzuki, T. Ito and Y. Taga, Morphological stability of TiO_2 thin films with isolated columns, *Jap J Appl Phys Part 2* **40** (2001), L398–L400.
- [146] V.C. Venugopal and A. Lakhtakia, Second harmonic emission from an axially excited slab of a dielectric thin-film helicoidal bianisotropic medium, *Proc Roy Soc Lond A* **454** (1998), 1535–1571; erratums: **455** (1999), 4383.
- [147] A. Lakhtakia, On second harmonic generation in sculptured nematic thin films (SNTFs), *Optik* **105** (1997), 115–120.
- [148] R.W. Boyd, *Nonlinear optics*, Academic Press, San Diego, CA, USA, 1992.
- [149] S.G. Bilén, M.T. Domonkos and A.D. Gallimore, Simulating ionospheric plasma with a hollow cathode in a large vacuum chamber, *J Spacecraft Rockets* **38** (2001), 617–621.



Akhlesh Lakhtakia is a Professor of Engineering Science and Mechanics at The Pennsylvania State University. He has published more than 440 journal articles; has contributed chapters to eight research books; has edited, co-edited, authored or co-authored eight books and four conference proceedings; has reviewed for 65 journals; and was the Editor-in-Chief of the international journal *Speculations in Science and Technology* from 1993 to 1995. He headed the IEEE EMC Technical Committee on Non-sinusoidal Fields from 1992 to 1994, and served as the 1995 Scottish Amicable Visiting Lecturer at the University of Glasgow. He is a Fellow of the Optical Society of America, SPIE—The International Society for Optical Engineering, and the Institute of Physics (UK). He was awarded the PSES Outstanding Research Award in 1996. Since 1999, he has organized four annual SPIE conferences on nanostructured materials and complex mediums. His current research interests lie in the electromagnetics of complex mediums, sculptured thin films, and chiral nanotubes.

Towards Optoelectronic Applications of Chiral Sculptured Thin Films

Martin W. McCall

Abstract

Axial propagation in nonhomogeneous dielectric material forms a canonical class of problems for which the Maxwell equations can be significantly simplified. Appropriate projection of the relative permittivity tensor reduces the curl relations to a set of four coupled ordinary differential equations, which can then be further reduced to a Helmholtz equation for the transverse components of the electric field phasor. Axial propagation through a chiral sculptured thin film (CSTF) layer is chosen as an example in which the reduced relations can be integrated exactly, while the Helmholtz equation can be solved approximately, using coupled-wave techniques. The electromagnetic response is thus shown to relate closely to Bragg grating physics, which is briefly reviewed. CSTFs are a burgeoning area of thin-film research lying somewhere between the topical fields of photonic materials and Bragg grating filters. The relationship to Bragg gratings evinces many potential application areas of CSTFs in optoelectronics research that exploit the circular polarization specificity of a CSTFs electromagnetic response. The examples reviewed here include narrow-band filtering, multi-pass filtering, and polarization routing.

1 Introduction

Deposit a dielectric thin film in a vacuum chamber using standard techniques, but during deposition modulate some growth parameter (e.g., incident azimuthal angle, incident vapor flux direction, substrate rotation rate, etc.). Examine the resultant structure under a scanning electron microscope and find that a forest of material columns has formed interspersed with voids. On closer examination, each column has a periodic wiggle or curl (the detail depending on which parameter was modulated), which, on account of its period of hundreds of nanometers, will interact strongly with visible and infrared radiation. Such sculptured thin films (STFs) [1] are now becoming established as a novel category of passive optical structures with the potential for a range of potential applications. The evolution and trends of STFs in general are described elsewhere in this volume [2].

In the mid 1990s, Lakhtakia and Weiglhofer [3] examined electromagnetic fields in a particular class of anisotropic STFs that have a chiral morphology, or chiral sculptured thin films (CSTFs). In these materials the aciculate microstructure takes the form of a columnar helix, and is produced via either incremental or continuous substrate rotation during the deposition process [4]. The resultant material is similar to a cholesteric liquid crystal [5], although being in the solid phase it is “hard wired,” and no orientational reordering is possible after fabrication. The interest in CSTFs stems from their distinct responses to orthogonal circular polarizations. Within a certain wavelength regime, the Bragg zone, light of the same handedness as the material is strongly reflected, while the other polarization is mainly transmitted. This so-called circular Bragg phenomenon (CBP) gives rise to many potential applications that are chiral analogs of those stemming from standard Bragg filters.

Our purpose in this chapter is on the one hand pedagogical, and on the other directional. As a paradigm for a complex medium which yet possesses analytically tractable descriptions, the CSTF is hard to beat. We examine thoroughly two methods: a full electromagnetic analysis based on an analytic solution to the Maxwell equations for propagation along the helicoidal axis of a CSTF, and an alternative method based on coupled-wave theory. The coupled wave approach in particular highlights the underlying Bragg physics of axial propagation in CSTFs. Guided by research directions in the field of Bragg gratings, we suggest potential longer-term applications of CSTFs that utilize the CBP.

In Section 2 we define precisely what is meant by axial propagation through a nonhomogeneous dielectric medium. We show that for such mediums the Maxwell equations may be simplified into equations for the transverse field phasor components. The example of an isotropic medium with a region of periodic nonhomogeneity is reviewed as the simplest case giving rise to Bragg phenomena. In Section 3, we introduce the relative permittivity tensor for the less trivial case of a CSTF. The electromagnetic analysis of axial propagation through a CSTF layer is carried out in Section 4, and the resultant reflectivities and transmissivities presented in Section 5. The alternative approach, based on coupled-wave techniques is

presented in Section 6, and is shown to describe the handedness-preserving nature of reflection from the CSTF layer. It is through this approach, when combined with handedness-reversing Fresnel reflections from the layer interfaces, that the physical picture developed in Section 7 emerges. In Section 8, the connection with Bragg grating physics is used to suggest applications of CSTFs for use in optoelectronic devices and systems.

2 Preliminaries

To prepare the ground, we review in the following two subsections two key aspects that feature throughout the rest of this chapter: (i) we show how the Maxwell equations for axial propagation may be simplified to a second-order ordinary differential equation for the projection of the electric field onto the plane orthogonal to the axial direction, and (ii) we review the physics of Bragg gratings.

2.1 Projected Helmholtz equation

Let us recall the time-domain form of the Maxwell equations in a dielectric medium where free current and charge sources are absent:

$$\nabla \times \tilde{\mathbf{E}}(\mathbf{r}, t) = -\frac{\partial}{\partial t} \tilde{\mathbf{B}}(\mathbf{r}, t), \quad (1)$$

$$\nabla \times \tilde{\mathbf{H}}(\mathbf{r}, t) = \frac{\partial}{\partial t} \tilde{\mathbf{D}}(\mathbf{r}, t), \quad (2)$$

$$\nabla \cdot \tilde{\mathbf{D}}(\mathbf{r}, t) = 0, \quad (3)$$

$$\nabla \cdot \tilde{\mathbf{B}}(\mathbf{r}, t) = 0. \quad (4)$$

In (1)–(4), $\tilde{\mathbf{D}}(\mathbf{r}, t)$ and $\tilde{\mathbf{E}}(\mathbf{r}, t)$ represent the dielectric displacement and the electric field, while $\tilde{\mathbf{B}}(\mathbf{r}, t)$ and $\tilde{\mathbf{H}}(\mathbf{r}, t)$ represent the magnetic induction and the magnetic field.¹ After Fourier transforming $\tilde{\mathbf{D}}(\mathbf{r}, t)$, $\tilde{\mathbf{E}}(\mathbf{r}, t)$, $\tilde{\mathbf{B}}(\mathbf{r}, t)$ and $\tilde{\mathbf{H}}(\mathbf{r}, t)$ into their frequency-domain counterparts $\mathbf{D}(\mathbf{r}, \omega)$, $\mathbf{E}(\mathbf{r}, \omega)$, $\mathbf{B}(\mathbf{r}, \omega)$ and $\mathbf{H}(\mathbf{r}, \omega)$, we obtain

$$\nabla \times \mathbf{E}(\mathbf{r}, \omega) = i\omega \mathbf{B}(\mathbf{r}, \omega), \quad (5)$$

$$\nabla \times \mathbf{H}(\mathbf{r}, \omega) = -i\omega \mathbf{D}(\mathbf{r}, \omega), \quad (6)$$

$$\nabla \cdot \mathbf{D}(\mathbf{r}, \omega) = 0, \quad (7)$$

$$\nabla \cdot \mathbf{B}(\mathbf{r}, \omega) = 0. \quad (8)$$

As such, (5)–(8) must be supplemented by constitutive algebraic relations connecting, for example, $\mathbf{D}$ and $\mathbf{B}$, to $\mathbf{E}$ and $\mathbf{H}$. The bianisotropic mediums community will

¹It is an unfortunate historical accident that out of $\tilde{\mathbf{B}}$ and $\tilde{\mathbf{H}}$, it is the former that has been burdened with the appellation *induction*, since it is this “field” that appears in the Lorentz force law.

want to express these relations in terms of all possible linear coefficients; nonlinear optics won't rest unless nonlinear terms are included; but those preparing an undergraduate course will probably just write $\mathbf{D} = \epsilon \mathbf{E}$ and $\mathbf{B} = \mu \mathbf{H}$. For further details on the constitutive relations, see the chapter by Weiglhofer in this volume [6]. Our focus in this chapter is on the following case: magnetic effects are absent ($\mu \rightarrow \mu_0$), and complexity in the medium arises just in the permittivity, which we decree may be anisotropic and dependent on just one spatial coordinate:

$$\mathbf{D}(\mathbf{r}, \omega) = \epsilon_0 \underline{\underline{\epsilon}}(z, \omega) \cdot \mathbf{E}(\mathbf{r}, \omega), \quad (9)$$

where $\underline{\underline{\epsilon}}(z, \omega)$ is the frequency-dependent relative permittivity tensor. Among the mediums described by (9), we may count sculptured thin films, liquid crystals and fiber Bragg gratings. Finally, we restrict attention to the propagation of plane waves precisely along the direction of nonhomogeneity.

After rescaling the magnetic field to $\mathbf{H}' = (\epsilon_0/\mu_0)^{1/2} \mathbf{H}$, carrying out a little vector manipulation, and dropping the (z, ω) dependence, the curl relations (5)–(6) reduce to

$$\hat{\mathbf{z}} \times \frac{d\mathbf{E}_\perp}{dz} = ik_0 \mathbf{H}'_\perp \quad (10)$$

and

$$\hat{\mathbf{z}} \times \frac{d\mathbf{H}'_\perp}{dz} = -ik_0 \left[\left(\underline{\underline{\epsilon}}^{-1} \right)_\perp \right]^{-1} \cdot \mathbf{E}_\perp, \quad (11)$$

where the subscript $\perp$ denotes restriction to the plane orthogonal to the z axis, and k_0 is the free-space wavenumber. What about the longitudinal component of the electric field, E_z ? This is calculated from the solution for $\mathbf{E}_\perp = [E_x, E_y]^T$ (with the superscript T denoting the transpose) according to

$$E_z = -\frac{\epsilon_{zx}E_x + \epsilon_{zy}E_y}{\epsilon_{zz}}, \quad (12)$$

which is just the algebraic relation resulting from the nullity of the longitudinal component of $\underline{\underline{\epsilon}} \cdot \mathbf{E}$. By eliminating the longitudinal component of $\mathbf{E}$, (10)–(11) project (5)–(6) onto the plane orthogonal to the axial direction. Eliminating $\mathbf{H}_\perp$ from (10) and (11), we obtain the Helmholtz equation for $\mathbf{E}_\perp$ as follows:

$$\frac{d^2 \mathbf{E}_\perp}{dz^2} + k_0^2 \left[\left(\underline{\underline{\epsilon}}^{-1} \right)_\perp \right]^{-1} \cdot \mathbf{E}_\perp = \mathbf{0}. \quad (13)$$

A simple example applied to a homogeneous anisotropic medium illustrates the calculation of $\left[\left(\underline{\underline{\epsilon}}^{-1} \right)_\perp \right]^{-1}$. Suppose the principal dielectric axes are labeled a, b, c , with the associated principal refractive indexes denoted $n_{a,b,c}$. Orient axis c to be

parallel to the y axis, with axis b lying in the xz plane at an angle χ with respect to the x axis. In the Cartesian basis we then have

$$\underline{\underline{\epsilon}} = \begin{bmatrix} \cos \chi & 0 & -\sin \chi \\ 0 & 1 & 0 \\ \sin \chi & 0 & \cos \chi \end{bmatrix} \begin{bmatrix} n_b^2 & 0 & 0 \\ 0 & n_c^2 & 0 \\ 0 & 0 & n_a^2 \end{bmatrix} \begin{bmatrix} \cos \chi & 0 & \sin \chi \\ 0 & 1 & 0 \\ -\sin \chi & 0 & \cos \chi \end{bmatrix}, \quad (14)$$

so that the restriction of $\underline{\underline{\epsilon}}^{-1}$ onto the xy plane is then

$$\left(\underline{\underline{\epsilon}}^{-1}\right)_{\perp} = \begin{bmatrix} \cos^2 \chi n_b^{-2} + \sin^2 \chi n_a^{-2} & 0 \\ 0 & n_c^{-2} \end{bmatrix}. \quad (15)$$

Finally, we find

$$\left[\left(\underline{\underline{\epsilon}}^{-1}\right)_{\perp}\right]^{-1} = \begin{bmatrix} \tilde{n}_d^2 & 0 \\ 0 & n_c^2 \end{bmatrix}, \quad (16)$$

where $\tilde{n}_d = (\cos^2 \chi n_b^{-2} + \sin^2 \chi n_a^{-2})^{-1/2}$. Although the so-called angle of rise χ is significant for oblique (i.e., nonaxial) propagation in CSTFs [7], its only role in axial propagation is to determine $\tilde{n}_d$.

2.2 Bragg grating physics

The simplest example to which (13) may be applied is an isotropic lossless medium wherein the refractive index varies sinusoidally between the planes $z = 0$ and $z = L$. The relative permittivity tensor reduces to the real-valued scalar function $\epsilon(z)$ given by

$$\epsilon(z) = \begin{cases} n^2 & 0 \geq z; z \geq L \\ n^2 \left(1 + 2\frac{\delta n}{n}\right) \cos\left(\frac{2\pi z}{\Lambda}\right) & 0 < z < L \end{cases}. \quad (17)$$

Here n is the average refractive index of the medium, δn is the amplitude of the index modulation, and Λ is the structural period. The modulated region constitutes a phase grating, commonly referred to as a Bragg grating.

For axial propagation within the grating, (13) reduces to

$$\frac{d^2 \psi}{dz^2} + \epsilon(z) k_0^2 \psi = 0, \quad (18)$$

where ψ represents either component of $\mathbf{E}_{\perp}$. Within the modulated region, the index variation will give rise to a wave propagating in both the $+z$ and $-z$ directions

so that

$$\psi(z) = \psi^+(z)e^{ink_0z} + \psi^-(z)e^{-ink_0z}, \quad (19)$$

where the amplitudes $\psi^\pm(z)$ quantify the energy exchange between the forward and backward propagating waves. Provided $\delta n \ll n$, $\psi^\pm(z)$ may be calculated using coupled wave techniques, as described for example in Ref. 10. The reflectivity R of a wave incident from $z = -\infty$ is then calculated as

$$R = \left| \frac{\psi^-(0)}{\psi^+(0)} \right|^2 = \left| \frac{\kappa_s \sinh \Delta_s L}{\Delta_s \cosh \Delta_s L - i \frac{\delta k_s}{2} \sinh \Delta_s L} \right|^2, \quad (20)$$

where

$$\left. \begin{aligned} \kappa_s &= \frac{\pi \delta n}{\lambda_0} \\ \delta k_s &= 2nk_0 - \frac{2\pi}{\Lambda} \\ \Delta_s &= +[\kappa^2 - (\delta k_s/2)^2]^{1/2} \end{aligned} \right\}. \quad (21)$$

The key features of the spectrum associated with a Bragg grating are illustrated in Figure 1. The reflectivity is a maximum at the resonant Bragg wavelength $\lambda_0^{(Br)}$ occurring when $\delta k_s = 0$, or

$$\lambda_0^{(Br)} = 2n\Lambda. \quad (22)$$

At this wavelength the reflectivity reduces to $|\tanh \kappa L|^2$, which tends to unity as the grating thickness increases, as shown in Figure 1. The bandwidth of the central zone $(\Delta\lambda_0)^{(Br)}$, within which the reflectivity is close to its value at the Bragg wavelength, may be estimated as the detuning δk_s required to achieve $\Delta_s = 0$. The result is that

$$(\Delta\lambda_0)^{(Br)} = 2\delta n\Lambda. \quad (23)$$

With suitable modification to take account of transverse dielectric variation, the above analysis may be applied to fiber Bragg gratings [8], which are currently of considerable technological significance [9].

3 Chiral sculptured thin films

The microstructure of a typical CSTF is shown in Figure 2. A schematic of the optical configuration for axial propagation in CSTFs is shown in Figure 3, which

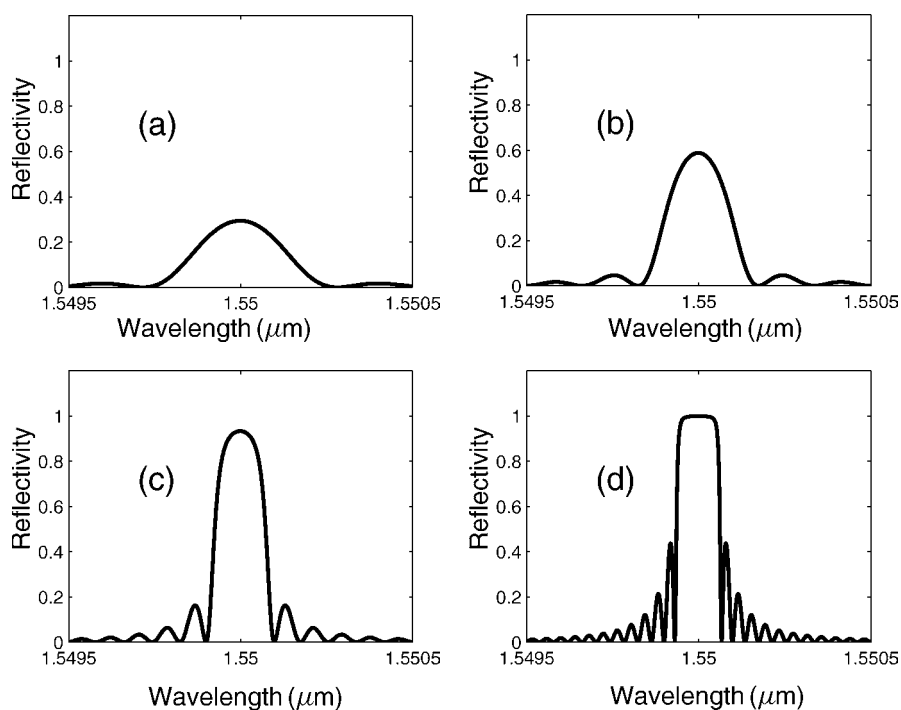


Figure 1 Axial reflectivities for the Bragg grating described by (17). Grating parameters: $n = 1.44$, $\delta n = 10^{-4}$, $\Lambda = 0.538 \mu\text{m}$, (a) $L = 3 \text{ mm}$, (b) $L = 5 \text{ mm}$, (c) $L = 10 \text{ mm}$, (d) $L = 20 \text{ mm}$.

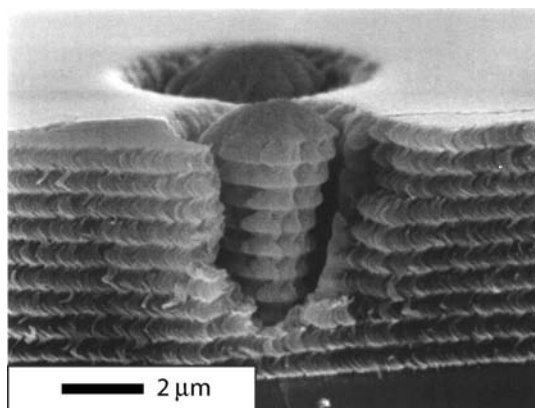


Figure 2 Scanning electron micrograph of a CSTF. A defect in the substrate caused an incidental hole to appear in the film. The chiral structure is thus exposed. Courtesy: R. Messier.

depicts a half-period twist of the sectional ellipse defined via

$$\left\langle \mathbf{r}_\perp, \left(\underline{\underline{\epsilon}}^{-1} \right)_\perp \mathbf{r}_\perp \right\rangle = 1, \quad (24)$$

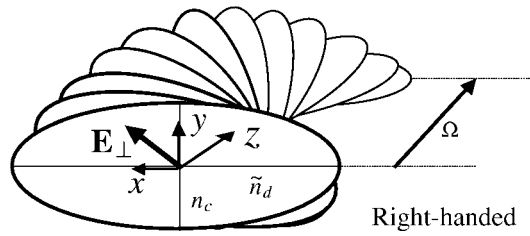


Figure 3 General configuration for axial propagation through a structurally right-handed CSTF.

where $\mathbf{r}_\perp \in \mathbb{R}^2$ and $\langle \cdot, \cdot \rangle$ denotes the usual inner product in $\mathbb{R}^2$. The twisting of the structure is contained in the z -dependence of $(\underline{\epsilon}^{-1})_\perp$. If the eigenvalues of $(\underline{\epsilon}^{-1})_\perp$ are denoted as $\tilde{n}_d^{-2}$ and n_c^{-2} , then the semi-axes of the ellipse described by (24) are $\tilde{n}_d$ and n_c . Although the distance from the center of the ellipse to a point on the ellipse has the physical significance of being proportional to the energy density stored by $\mathbf{E}_\perp$, the pedagogical significance of the ellipse here is to depict the rotation of the principal axes. It is also clear from Figure 3 that the optical period is exactly half of the structural period 2Ω for axial propagation. After setting $\underline{\epsilon}_{\text{perp}} \equiv [(\underline{\epsilon}^{-1})_\perp]^{-1}|_{z=0}$, the twisting anisotropy of the medium is described by

$$\left[(\underline{\epsilon}^{-1})_\perp \right]^{-1} = \mathcal{R}^{-1}(z) \cdot \underline{\epsilon}_{\text{perp}} \cdot \mathcal{R}(z), \quad (25)$$

where $\mathcal{R}(z) \in SO(2)$ is the rotation operator with the Cartesian representation

$$\mathcal{R}(z) = \begin{bmatrix} \cos pz & -\sin pz \\ -\sin pz & \cos pz \end{bmatrix} \quad (26)$$

with $p = \pi/\Omega$. The structural handedness can be described by bestowing a sign on Ω . Here, a positive value of Ω corresponds to a *right-handed* CSTF.

4 Full electromagnetic analysis

A remarkable feature of the electromagnetic analysis of axial propagation through CSTFs is that the Maxwell equations are *exactly* soluble. The analytic solution is not particularly convenient to write down, but it is very easy to code computationally. An alternative method that does yield explicit analytic formulas is given in Section 6. However, the exact solution is sufficiently enticing that it is worth deriving it here.

The two-component vectors $\mathbf{E}_\perp$ and $\mathbf{H}'_\perp$ are subjected to the so-called Oseen transformation [11]

$$\mathbf{e} = \mathcal{R}\mathbf{E}_\perp, \quad \mathbf{h}' = \mathcal{R}\mathbf{H}'_\perp, \quad (27)$$

and the four-component vector $[\mathbf{e}, \mathbf{h}']^T$ is then formed. This vector satisfies the matrix differential equation.

$$\frac{d}{dz} \begin{bmatrix} e_x \\ e_y \\ h'_x \\ h'_y \end{bmatrix} = i \begin{bmatrix} 0 & -ip & 0 & k_0 \\ ip & 0 & -k_0 & 0 \\ 0 & -n_c^2 k_0 & 0 & -ip \\ \tilde{n}_d^2 k_0 & 0 & ip & 0 \end{bmatrix} \begin{bmatrix} e_x \\ e_y \\ h'_x \\ h'_y \end{bmatrix} \equiv i \mathbf{M} \cdot \begin{bmatrix} \mathbf{e} \\ \mathbf{h}' \end{bmatrix}. \quad (28)$$

The effect of the transformation (27) is to render the operator $\mathbf{M}$ independent of z and thus for the solution to be found using standard methods [12-13]. The phasors $\mathbf{E}_\perp$ and $\mathbf{H}'_\perp$ can be recovered by inverting (27).

If we consider axial excitation of a CSTF of thickness L by arbitrarily polarized light incident from a homogeneous isotropic medium of refractive index n , the electric field phasor up to the incident boundary $z = 0$ can be written as

$$\mathbf{E}_\perp = \begin{bmatrix} a_x \\ a_y \end{bmatrix} e^{ink_0 z} + \begin{bmatrix} r_x \\ r_y \end{bmatrix} e^{-ink_0 z}, \quad (29)$$

while the transmitted field into a similar medium beyond the exit boundary at $z = L$ can be written as

$$\mathbf{E}_\perp = \begin{bmatrix} t_x \\ t_y \end{bmatrix} e^{ink_0(z-L)}. \quad (30)$$

Since $\mathbf{E}_\perp$ is continuous, the phasors in (29) and (30) also represent the fields just inside the boundaries. Application of this together with the continuity of $\mathbf{H}'_\perp$, which is determined by (10), leads to the algebraic relation

$$\begin{bmatrix} t_x \\ t_y \\ -nt_y \\ nt_x \end{bmatrix} = [\tilde{\mathbf{R}}^{-1}(L) e^{iLM}] \cdot \begin{bmatrix} a_x + r_x \\ a_y + r_y \\ n(r_y - a_y) \\ n(a_x - r_x) \end{bmatrix}, \quad (31)$$

where $\tilde{\mathbf{R}} = \mathcal{R} \otimes \underline{\underline{I}}$, with $\underline{\underline{I}}$ being the two-dimensional identity. Equation (31) is a set of four linear algebraic relations from which the unknown amplitudes t_x, t_y, r_x and r_y may be found in terms of the known incident amplitudes a_x and a_y . The helical morphology of CSTFs suggests the expression of all amplitudes in a circular basis so that, for example,

$$\begin{bmatrix} a_x \\ a_y \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ i & -i \end{bmatrix} \begin{bmatrix} a_L \\ a_R \end{bmatrix}, \quad (32)$$

etc. The solution of (31) may then be recast as

$$\begin{bmatrix} r_L \\ r_R \end{bmatrix} = \begin{bmatrix} r_{LL} & r_{LR} \\ r_{RL} & r_{RR} \end{bmatrix} \begin{bmatrix} a_L \\ a_R \end{bmatrix}, \quad (33)$$

$$\begin{bmatrix} t_L \\ t_R \end{bmatrix} = \begin{bmatrix} t_{LL} & t_{LR} \\ t_{RL} & t_{RR} \end{bmatrix} \begin{bmatrix} a_L \\ a_R \end{bmatrix}. \quad (34)$$

The reflection and transmission coefficients (r_{LL} , etc.) are completely analytically determined, but the formulas are much too cumbersome for reproduction here.

5 The optical response of a CSTF to axial excitation

The basic characteristics of a CSTF are well illustrated by their transmission and reflection properties, calculated from (33) and (34). Suppose we take $\Omega = +200$ nm (i.e., right-handed), $\chi = 30^\circ$, $n_a = 1.64$, $n_b = 1.73$, $n_c = 1.65$, $L = 30|\Omega|$ and surround the CSTF with vacuum ($n = 1$). The reflection and transmission characteristics are quantified by $|r_{jl}|^2$ and $|t_{jl}|^2$, where j and l correspond to either *Left* or *Right* circular polarization (respectively LCP or RCP). For example, $|r_{LR}|^2$ represents the intensity of light reflected into the LCP state for incident RCP state. Figures 4 and 5 illustrate these quantities for the CSTF layer. For wave-

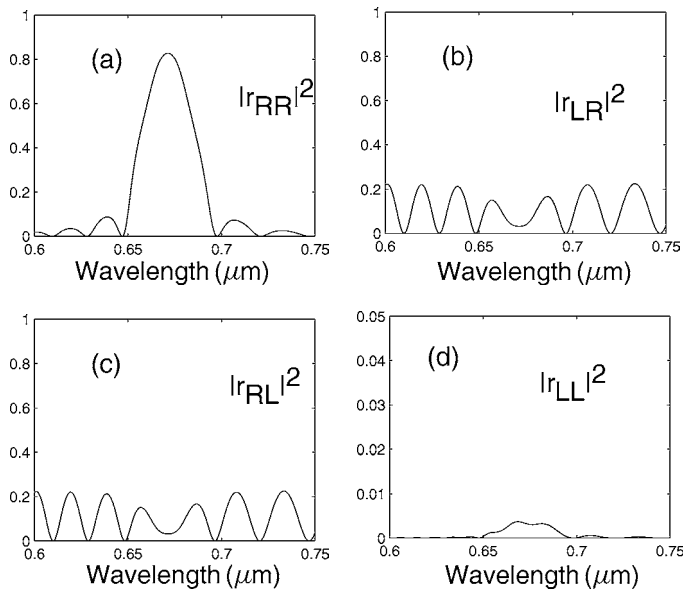


Figure 4 Axial reflectivities for a structurally right-handed CSTF: (a) $|r_{RR}|^2$, (b) $|r_{LR}|^2$, (c) $|r_{RL}|^2$ and (d) $|r_{LL}|^2$. Calculation parameters are given at the beginning of Section 5.

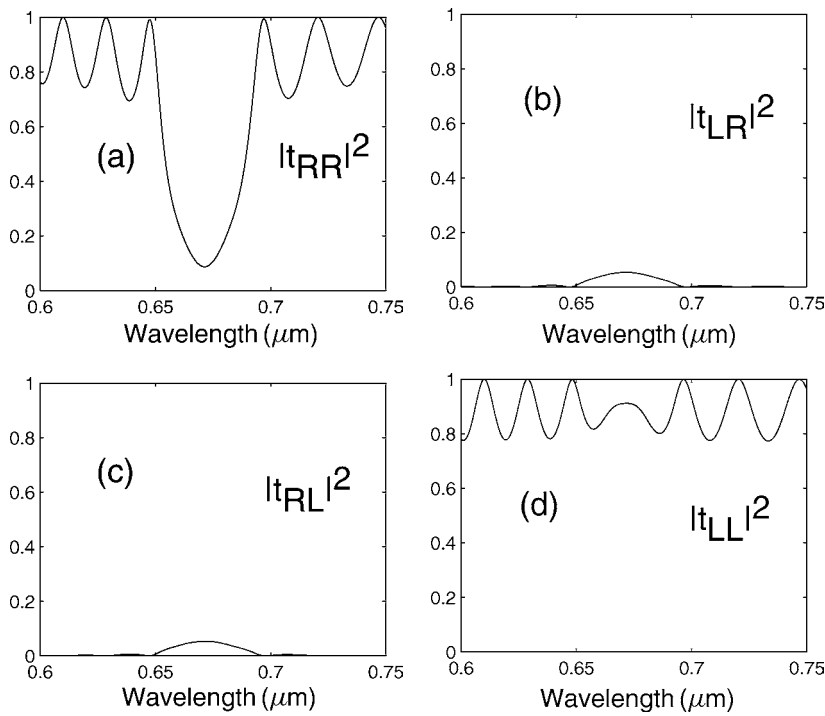


Figure 5 Axial transmissivities of a structurally right-handed CSTF: (a) $|t_{RR}|^2$, (b) $|t_{LR}|^2$, (c) $|t_{RL}|^2$ and (d) $|t_{LL}|^2$. Calculation parameters are given at the beginning of Section 5.

lengths in the range $660 \text{ nm} < \lambda_0^{(Br)} < 680 \text{ nm}$ the reflectivity $|r_{RR}|^2$ increases significantly while the corresponding transmissivity $|t_{RR}|^2$ decreases. This is indicative of the so-called circular Bragg phenomenon [14-15], wherein circularly polarized light whose handedness coincides with the material's structural handedness is strongly reflected from the film, while cross-handed light is transmitted. Within the same wavelength regime, the cross-polarized remittances $|r_{RL}|^2$, $|r_{LR}|^2$, $|t_{RL}|^2$, and $|t_{LR}|^2$ are all relatively small. Finally, the transmissivity $|t_{LL}|^2$ and reflectivity $|r_{LL}|^2$ are close to unity and zero respectively, indicating that the circular polarization with handedness opposite to the helicoidal medium propagates largely as if in a homogeneous dielectric medium. These features are all explained in physical terms in the context of the coupled wave description in Section 6. For a lossless CSTF, energy conservation requires that

$$|t_{LL}|^2 + |t_{RL}|^2 + |r_{LL}|^2 + |r_{RL}|^2 = 1 \quad (35)$$

and

$$|t_{LR}|^2 + |t_{RR}|^2 + |r_{LR}|^2 + |r_{RR}|^2 = 1. \quad (36)$$

Detailed analysis of the eigenvalues of the 4×4 matrix appearing in (28) shows that the CBP occurs in a regime centered at (free-space) wavelength [16]

$$\lambda_0^{Br} = |\Omega| (\tilde{n}_d + n_c) \approx 672 \text{ nm}, \quad (37)$$

and is of bandwidth

$$(\Delta\lambda_0)^{Br} = 2 |\Omega| (\tilde{n}_d - n_c) \approx 24 \text{ nm} \quad (38)$$

for the chosen parameters.

Of particular interest in optoelectronic applications are the characteristics of the CSTF as the layer thickness is increased. The key reflectivities are $|r_{RR}|^2$, which exhibits the Bragg resonance, and $|r_{LR}|^2$, which should usually be minimized in applications. These are shown for different layer thicknesses in Figure 6. As the layer thickness L is increased the response is desirably flatter within the Bragg regime, although the sidebands increase relative to the maximum reflectivity. The appended calculations of $|r_{LR}|^2$ in each case show that the polarization crosstalk can be quantitatively assessed for applications. In fact, simple strategies for minimizing $|r_{LR}|^2$ emerge, once the full physical picture of the CSTF as a handed-reflector is established, as discussed in the next section.

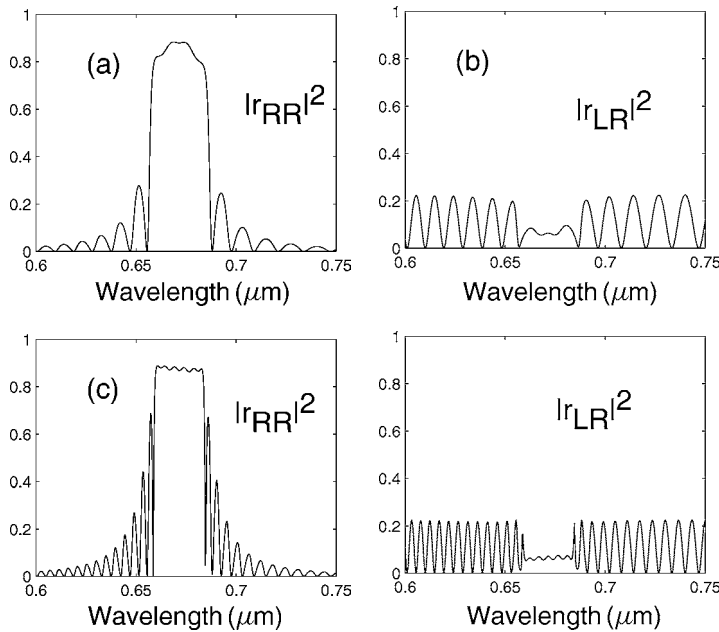


Figure 6 Axial reflectivities for different CSTF thicknesses: (a) $|r_{RR}|^2$, $L = 60\Omega$, (b) $|r_{LR}|^2$, $L = 60\Omega$ (c) $|r_{RR}|^2$, $L = 120\Omega$, and (d) $|r_{LR}|^2$, $L = 120\Omega$. Other parameters are as for Figure 4.

6 Coupled-wave techniques

The same problem of axial propagation through a CSTF may be treated in a different way using coupled-wave techniques [17-18]. Using the twisting anisotropic form for $[(\underline{\underline{\epsilon}}^{-1})_{\perp}]^{-1}$ given by (25) in (13), we arrive at

$$\frac{d^2 \mathbf{E}_{\perp}}{dz^2} + k_0^2 [\mathcal{R}^{-1} \underline{\underline{\epsilon}}_{\text{perp}} \mathcal{R}] \cdot \mathbf{E}_{\perp} = \mathbf{0}. \quad (39)$$

The transverse field inside the CSTF is then expanded in terms of forward and backward propagating modes as

$$\mathbf{E}_{\perp} = \mathbf{A}^+ e^{ikz} + \mathbf{A}^- e^{-ikz}, \quad (40)$$

where $k = \bar{n}k_0$. The average refractive index $\bar{n} = [(n_c^2 + \tilde{n}_d^2)/2]^{1/2} \approx (n_c + \tilde{n}_d)/2$ for small birefringence. $\mathbf{A}^{\pm}$ represent the forward and reverse propagating phasor amplitudes. The rotation $\mathcal{R}(z)$ from (26) is usefully expanded as

$$\mathcal{R}(z) = \frac{1}{2} (\sigma^* e^{ipz} + \sigma e^{-ipz}), \quad (41)$$

where

$$\sigma \equiv \begin{pmatrix} 1 & i \\ -i & 1 \end{pmatrix}. \quad (42)$$

After inserting (40)–(42) into (39) and equating phase-matched terms, we find, in the circular basis, for a right-handed CSTF that

$$\frac{d}{dz} \begin{bmatrix} A_L^+ \\ A_R^+ \end{bmatrix} = i\kappa \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} A_R^- \\ A_L^- \end{bmatrix} e^{-i\delta kz}, \quad (43)$$

$$\frac{d}{dz} \begin{bmatrix} A_R^- \\ A_L^- \end{bmatrix} = -i\kappa \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} A_L^+ \\ A_R^+ \end{bmatrix} e^{i\delta kz}, \quad (44)$$

where the coupling constant $\kappa = \pi(\tilde{n}_d - n_c)/\lambda_0$, and the detuning $\delta k = 2(k - \pi/|\Omega|)$, λ_0 being the free-space wavelength. The solutions of (43) and (44) enable the field components $E_{L,R}^{\pm} = A_{L,R}^{\pm} e^{\pm ikz}$ to be propagated from $z = 0$ to $z = L$:

$$\begin{bmatrix} E_L^+ \\ E_R^+ \\ E_L^- \\ E_R^- \end{bmatrix}_{z=L^-} = \begin{bmatrix} e^{ikL} & 0 & 0 & 0 \\ 0 & \mathcal{P} & 0 & \mathcal{Q} \\ 0 & 0 & e^{-ikL} & 0 \\ 0 & \mathcal{Q}^* & 0 & \mathcal{P}^* \end{bmatrix} \begin{bmatrix} E_L^+ \\ E_R^+ \\ E_L^- \\ E_R^- \end{bmatrix}_{z=0^+} = \mathcal{S} \cdot \mathcal{E}|_{z=0^+}, \quad (45)$$

where $\mathcal{E}$ denotes the column vector $[E_L^+, E_R^+, E_L^-, E_R^-]^T$. The notations $z = 0^+$ and $z = L^-$ indicate the location of fields at the boundaries of the CSTF, just inside the medium. Also,

$$\mathcal{P} = e^{ipL} \left[\cosh(\Delta L) + i \frac{\delta k}{2\Delta} \sinh(\Delta L) \right], \quad (46)$$

$$\mathcal{Q} = \frac{i e^{ipL} \kappa}{\Delta} \sinh(\Delta L), \quad (47)$$

and

$$\Delta = + \left[|\kappa|^2 - (\delta k/2)^2 \right]^{1/2}. \quad (48)$$

Equations (43) and (44), together with (45), yield significant insights into axial propagation in CSTFs. LCP light propagates as if in a homogeneous, isotropic medium with refractive index $\bar{n}$. Forward and reverse propagating RCP modes are coupled to each other exactly as for a Bragg grating [21], responding to the periodic dielectric perturbation $|n_c - \tilde{n}_d|$. Equation (45) may be solved for the reflectivity

$$R_{RR}^{(CWT)} \equiv |r_{RR}^{(CWT)}|^2 = \left| \frac{\kappa \sinh(\Delta L)}{\Delta \cosh(\Delta L) - i(\delta k/2) \sinh(\Delta L)} \right|^2. \quad (49)$$

Also,

$$T_{RR}^{(CWT)} = 1 - R_{RR}^{(CWT)} = |\cosh(\Delta L) - i(\delta k/2\Delta) \sinh(\Delta L)|^{-2}. \quad (50)$$

Equations (49) and (50) are well known from Bragg grating physics [21]. Equation (49), for example, should be compared with (20). Since the amplitude of the effective sinusoidal refractive index variation for the CSTF case is $\delta n = |\tilde{n}_d - n_c|/2$, the coupling constant is twice that for the scalar grating considered in Section 2.2: $\kappa = 2\kappa_s$. The coupled-wave theoretical approximation for the reflectivity $R_{RR}^{(CWT)}$ and transmissivity $T_{RR}^{(CWT)}$ for the exemplar CSTF discussed in Section 5 are illustrated in Figure 7. Comparison with Figures 4 and 5 shows that the coupled-wave interpretation is validated. The slight overestimate of $R_{RR}^{(CWT)}$ with respect to R_{RR} in the Bragg regime is accounted for through the neglect of the cross-polarized reflectivities $R_{LR} = R_{RL}$ which, in this simple form of the coupled-wave technique, are both zero. Under the approximation that $|n_c - \tilde{n}_d| \ll \bar{n}$, the coupled-wave estimates of the Bragg wavelength λ_0^{Br} and the Bragg zone bandwidth $(\Delta\lambda_0)^{Br}$ coincide with (37) and (38) [19].

For optoelectronic applications, the Bragg wavelength can, if desired, be controlled *solely* through the half-pitch Ω . Thus, provided the imaginary parts of the principal refractive indexes n_a , n_b and n_c are not significant at the wavelength of interest, a polarization-sensitive filter can be fabricated at that wavelength simply

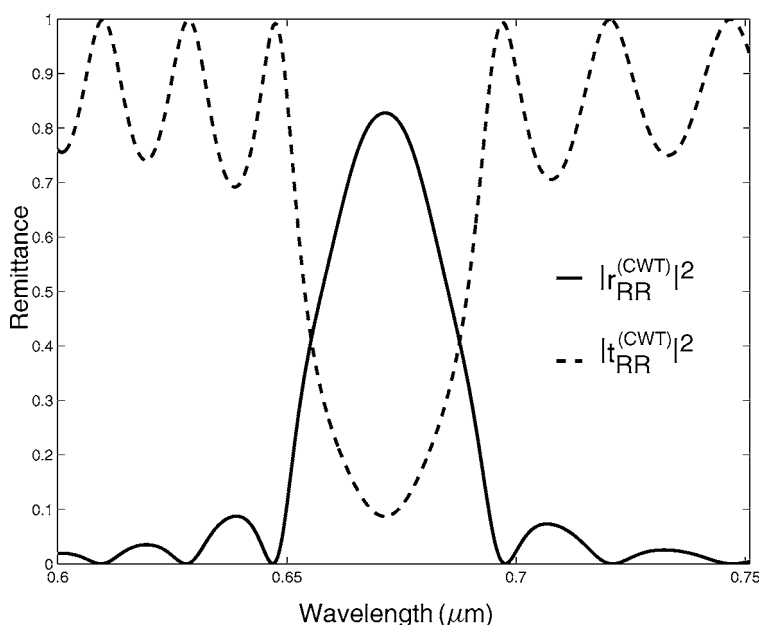


Figure 7 Axial reflectivity $R_{RR}^{(CWT)} = |r_{RR}^{(CWT)}|^2$ and transmissivity $T_{RR}^{(CWT)} = |t_{RR}^{(CWT)}|^2$ for a typical CSTF. Computational parameters are the same as for Figures 4 and 5.

by tailoring $|\Omega|$ according to (37). Thus, CSTFs provide considerable flexibility for device design.

The coupled-wave description of CSTFs has the advantage that simple explicit formulas for the co-handed spectral reflectivities and transmissivities, collectively called remittances, can be derived. It also provides a simple physical picture for why the CBP occurs; namely, that the co-handed polarization experiences a periodic modulation of index via the medium birefringence $|n_c - \tilde{n}_d|$. Its limitation, at least in the foregoing derivation, is that it does not provide a mechanism for coupling LCP and RCP states; but that coupling is evidently present from Figures 4 and 5. There turns out, however, to be a very simple physical explanation of this also.

7 The multireflectivity model of CSTFs

The coupled-wave technique yielded the propagation of the electromagnetic field *inside* the CSTF but took no account of reflection/transmission at the two faces of the thin film. Now the CSTF provides a mechanism for *co*-reflection of light of a given handedness. Reflection at the interfaces provides the mechanism for cross-polarized fields, leading to the simple physical picture for a right-handed CSTF shown in Figure 8. Around the Bragg wavelength, providing the CSTF is reasonably thick, co-handed light will experience a reflection at the entrance interface, and a reflection of the transmitted portion by the bulk of the CSTF. There is no interference between these two reflected contributions, as they are oppositely handed.

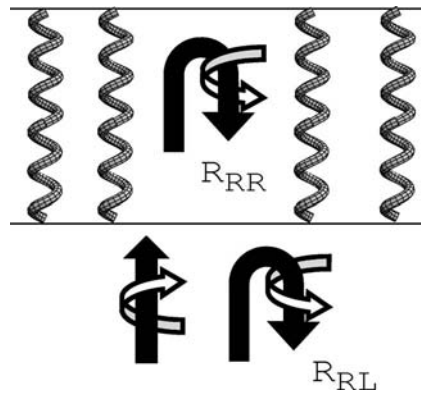


Figure 8 Physical model of propagation through a structurally right-handed CSTF.

These considerations of the various Fresnel reflections/transmissions combining with the chiral reflection lead to very simple but accurate estimates of the remittances at the Bragg wavelength for a right-handed CSTF, as follows (calculations for when the CBP is saturated, i.e., $\kappa L \rightarrow \infty$, were presented in [20]):

$$R_{LL} \approx 0, \quad (51)$$

$$R_{RL,LR} \approx \left(\frac{n - \bar{n}}{n + \bar{n}} \right)^2, \quad (52)$$

$$R_{RR} \approx \left[\frac{4n\bar{n}}{(n + \bar{n})^2} \right]^2 R_{RR}^{(CWT)}, \quad (53)$$

$$T_{LL} \approx \left[\frac{4n\bar{n}}{(n + \bar{n})^2} \right]^2, \quad (54)$$

$$T_{LR,RL} \approx \left[\frac{4n\bar{n}}{(n + \bar{n})^2} \right]^2 \frac{(n - \bar{n})^2}{(n + \bar{n})^2} R_{RR}^{(CWT)}, \quad (55)$$

$$T_{RR} \approx T_{RR}^{(CWT)}, \quad (56)$$

where $R_{RR}^{(CWT)}$ and $T_{RR}^{(CWT)}$ are provided by (49) and (50). Outside the Bragg regime, light of either handedness is readily transmitted, so that the layer acts essentially as a Fabry–Pérot resonator. This accounts for the ringing outside the Bragg regime in the cross reflectivities and transmissivities in Figures 4 and 5, and which become more finely spaced as the thickness increases (Figure 6).

Taking account of field continuity at the boundaries of a CSTF surrounded by a medium of index n yields the relationship

$$\mathcal{E}|_{z=L^+} = \mathcal{N}(\bar{n}, n) \cdot \mathcal{S} \cdot \mathcal{N}(n, \bar{n}) \mathcal{E}|_{z=0^-}, \quad (57)$$

where

$$\mathcal{N}(\sigma_a, \sigma_b) = \frac{1}{2} \begin{bmatrix} 1 + \frac{\sigma_a}{\sigma_b} & 0 & 0 & 1 - \frac{\sigma_a}{\sigma_b} \\ 0 & 1 + \frac{\sigma_a}{\sigma_b} & 1 - \frac{\sigma_a}{\sigma_b} & 0 \\ 0 & 1 - \frac{\sigma_a}{\sigma_b} & 1 + \frac{\sigma_a}{\sigma_b} & 0 \\ 1 - \frac{\sigma_a}{\sigma_b} & 0 & 0 & 1 + \frac{\sigma_a}{\sigma_b} \end{bmatrix}. \quad (58)$$

The spectral remittances of (33) and (34) can then be readily computed from (57); and, in fact, calculations using this method yield results very close to the full electromagnetic analysis, when the local birefringence is not high.

The physical insights arising from the model developed suggest various methods of virtually eliminating the cross-reflectivities. Index-matching the CSTF by surrounding it with a homogeneous isotropic dielectric of refractive index $\bar{n}$ is an obvious possibility, although Hodgkinson *et al.* [20] have also analyzed methods based on simultaneous phase/amplitude matching at the boundaries, and using moth-eye coatings. With careful adjustment of parameters the cross-contamination can be as little as 0.000003%.

Viewing the CSTF as a chiral Bragg reflector opens up many potential applications, to which we now turn.

8 Applications

We have seen that the basic CSTF provides a reflector that is specific to a particular circular polarization state. Proposals have been made for a pair of such mirrors in a resonator configuration [20] to create a novel means of coherent generation of circularly polarized light. Other applications include Šolc filters and gas sensors [22]. Our attention in this chapter is focused on the possibilities that CSTFs may offer optoelectronic networks.

For many years, optical scientists have endeavored to make polarization-flat components for integrated optics and telecommunications, fabricating, for example, polarization-maintaining optical fibers. Of course, this need not be the case; indeed, if suitable methods of polarization management could be developed, then there is a potential doubling of the available optical bandwidth. It is trivial to imagine any device, a coupler for example, to be sensitive to polarization by simply placing a polarizer before the structure. However, this method of introducing polarization sensitivity is hardly useful, since the polarization that is not passed is simply rejected and cannot be used. More promising would be nonrejective schemes, which are possible with combinations of retardation plates and polarizers; but such multielement schemes are inappropriate for applications with requirements for rigidity and compactness. Moreover, many simple polarization-selective devices send the different polarizations off in orthogonal directions. Thus, despite

the potential factor-of-two increase in bandwidth, device technology has not focused much on utilizing polarization effects. One difficulty is in reliably maintaining polarization through a fiber over significant distances (i.e., kilometers). However, once the change of the polarization state over a given fixed link length has been measured (i.e., from the source to the point of interest), repeatable restoration of the polarization to its state at the source is comparatively straightforward.

Much recent research has been done, both theoretically and experimentally, on the electromagnetic response of CSTFs, though for optoelectronic applications they are immature. They do, however, open up the possibility of polarization-multiplexing in which passed and rejected polarizations are filtered without loss. However, there are considerable technical challenges to be overcome before even the simplest optoelectronic devices can be fabricated. Of course, with ever greater demands on telecommunication subsystems to provide increased bandwidth, such possibilities, even at the level of speculation, merit serious attention.

We here describe a selection of applications that are suitable for integration into optoelectronic networks that are principally inspired by the Bragg nature of the phenomenon and are circularly-polarization-specific analogs of structures already widely employed.

8.1 Narrow-band polarization filter

Many applications in conventional integrated optics and optoelectronics require a filter that is narrower than that provided by the Bragg phenomenon alone. Unfortunately, strategies for reducing the bandwidth suggested by (38) are not effective. Reduction of Ω changes the Bragg wavelength according to (37); and reduction of $|n_c - \tilde{n}_d|$ decreases the cohered reflectivity since $\kappa \propto |n_c - \tilde{n}_d|$. Other methods are possible, however, based on creating a resonance between two distinct reflector regions, the chiral analog of techniques known from fiber optics and integrated optoelectronics [8]. The two such schemes studied to date are illustrated in Figure 9.

In the first scheme, an intermediate quarter-wave spacer layer between two CSTF layers sets up an effective Fabry–Pérot response for the strongly interacting polarization, whereas the orthogonal polarization again sees effectively a single medium of average refractive index $\bar{n}$. The index-matching requirement for the noninteracting polarization is that the spacer layer have this index also. This may be achieved either by fabricating an isotropic spacer layer between the two CSTF layers as in [23], or by simply halting the substrate rotation during fabrication for the same underlying material to yield an *anisotropic* spacer layer of the appropriate $\bar{n}$ as in [24]. Either way, the spectral response of such a filter is modeled adequately using standard Fabry–Pérot theory according to

$$T_{RR} = \frac{1}{1 + \frac{4|r_{RR}|^2}{(1 - |r_{RR}|^2)^2} \sin^2(\delta_{RR}/2)}, \quad (59)$$

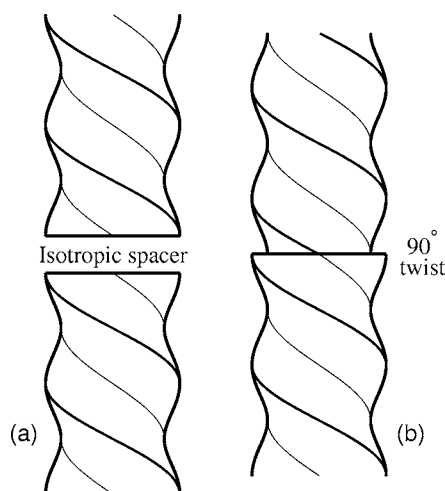


Figure 9 (a) Two CSTF layers with intermediate spacer, (b) two CSTF layers with intermediate twist discontinuity.

where δ_{RR} is the total phase difference between RCP light reflected at the first CSTF-spacer interface and light reflected at the second spacer-CSTF interface. Here, T_{RR} represents the reflectivity of the composite structure. The reflectivity $|r_{RR}|^2$ can be calculated using either the full electromagnetic analysis of Section 4, or via (49). A typical intra-Bragg-zone spectral hole using this technique has recently been fabricated [23].

The second scheme was proposed by Hodgkinson *et al.* [23] and is illustrated in Figure 9(b). Here, two identical CSTF layers are fabricated serially, the only distinction between the two being that a discrete axial rotation is applied to the second layer with respect to the first. This is in every respect the same type of discontinuity known in scalar Bragg gratings within which exists a phase discontinuity. The CSTF case is particularly simple (conceptually) to fabricate. Evaporation is ceased, the substrate is rotated, and then resumed. This scheme has been recently implemented experimentally [23], as shown for a left-handed CSTF in Figure 10.

The former scheme produces similar spectral characteristics [23], and both schemes offer the possibility of narrow-band polarization-dependent filtering. The twist-discontinuity technique in particular offers many possible configurations which are chiral analogs of known scalar filter designs.

8.2 Multipass narrow-band filters

In general, if there are N phase discontinuities along a Bragg grating, then there will be N spectral holes within the Bragg zone. With multiple twist discontinuities, therefore, it should be possible to design a multiply transmissive filter. Figure 11 achieves this through the use of three twist discontinuities effectively defining four separate CSTF layers. Each peak is about 2 nm wide, and the distribution of the twist discontinuities is such that they are all separated within the Bragg regime.

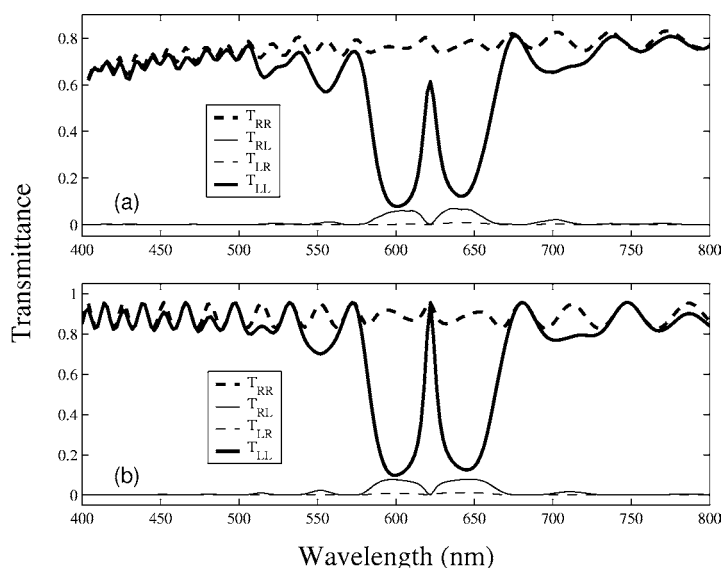


Figure 10 (a) Measured spectrums of the transmissivities, T_{RR} , T_{RL} , T_{LR} and T_{LL} of a spacerless filter with a twist. (b) Transmissivity spectrums simulated for the above filter using the parameters $\Omega = -163$ nm (left-handed), $L = 12|\Omega|$, $\bar{n} = 1.914$ and $|n_c - \tilde{n}_d| = 0.116$. The calculated Bragg wavelength is $\lambda_0^{(Br)} = 622$ nm. Taken from Ref. 23.

8.3 Tailored polarization-specific filters

Many techniques exist for improving the spectral response of fiber Bragg grating filters, based around applying an envelope function to the nominally sinusoidal modulation [25]. Appropriate apodization of the sinusoidal modulation can suppress the undesirable sidebands [26]. Similar techniques can be applied to the CSTF filters discussed here, to diminish the sidebands present in Figure 6. As the 90° twist discontinuity has been identified as the chiral analog of a quarter-wave shift in a scalar Bragg grating, further device concepts are possible. It is well known, for example, that the inclusion of quarter-period spacer layers in stacks consisting of alternating layers of isotropic dielectric layers with refractive indexes n_h and n_l , can be used to achieve filters with desirable shapes, such as a top-hat response [27]. The chiral analog, along with its polarization-dependant response is shown in Figure 12. Two technical considerations emerge from the comparison between propagation through CSTFs and isotropic dielectric stacks:

- For a given step $n_h - n_l$ in the standard stack filter, the equivalent birefringence for the CSTF filter is $|n_c - \tilde{n}_d| = 2|n_h - n_l|/\pi$.
- The equivalent CSTF filter requires exactly one-half the structural periods of the corresponding scalar filter. This originates from the fact that the optical repeat distance of the CSTF filter is half the structural period.

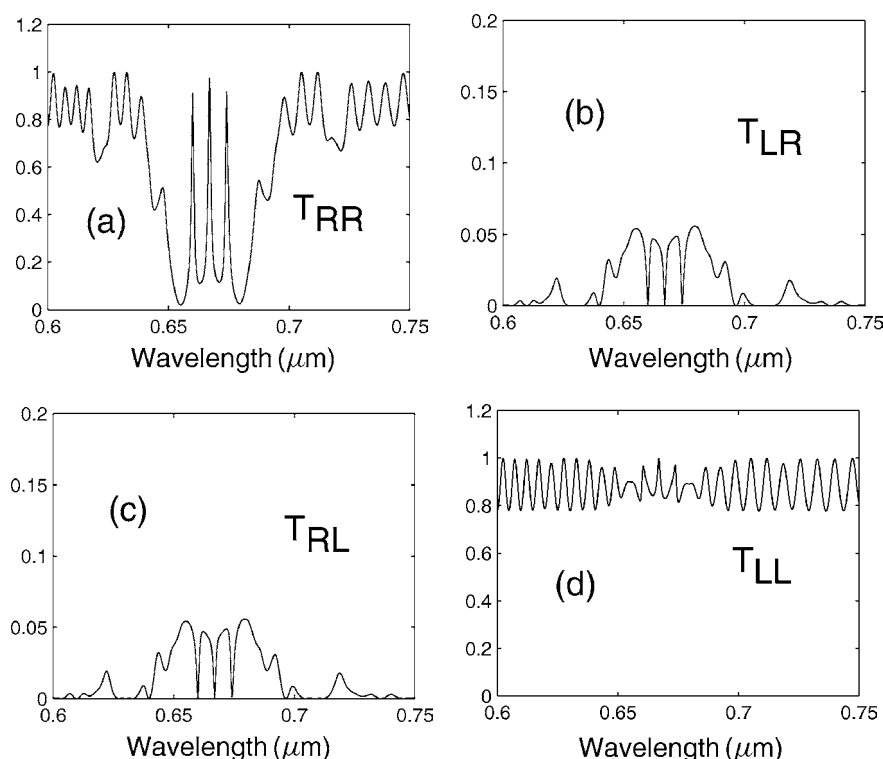


Figure 11 Composite transmissivities, T_{RL} , etc., for a CSTF multilayer filter. Layer thicknesses are $36|\Omega|$, $20|\Omega|$, $20|\Omega|$ and $36|\Omega|$, each separated by a twist discontinuity of 90° . Other CSTF parameters as in Figure 4. Each graph shows the axial transmissivity the entire four-layer structure.

8.4 Polarization routing

One area in which polarization filters may make a significant impact is in telecommunications, in which beam polarization has played only a relatively minor role with respect to the exploitation and encoding of information via other attributes of light such as intensity, wavelength and phase. Why is this? One reason is that many devices that exploit polarization are *active* requiring the application of an external electric or magnetic field, such as in isolators or circulators. Passive components, such as polarizers and polarizing beamsplitters, have inherent disadvantages. Polarizers are necessarily *rejective*, the polarization orthogonal to the transmission axis being simply absorbed. Polarizing beamsplitters preserve both components, but require *spatial* orthogonality, for either beamsplitting or beam combination. Integrated optoelectronics applications normally require such routing operations to be carried out *in-line* for compatibility with optical fibers. In this section we discuss a device design [28] which achieves passive in-line polarization routing using CSTF technology combined with a standard directional coupler, as shown in Figure 13.

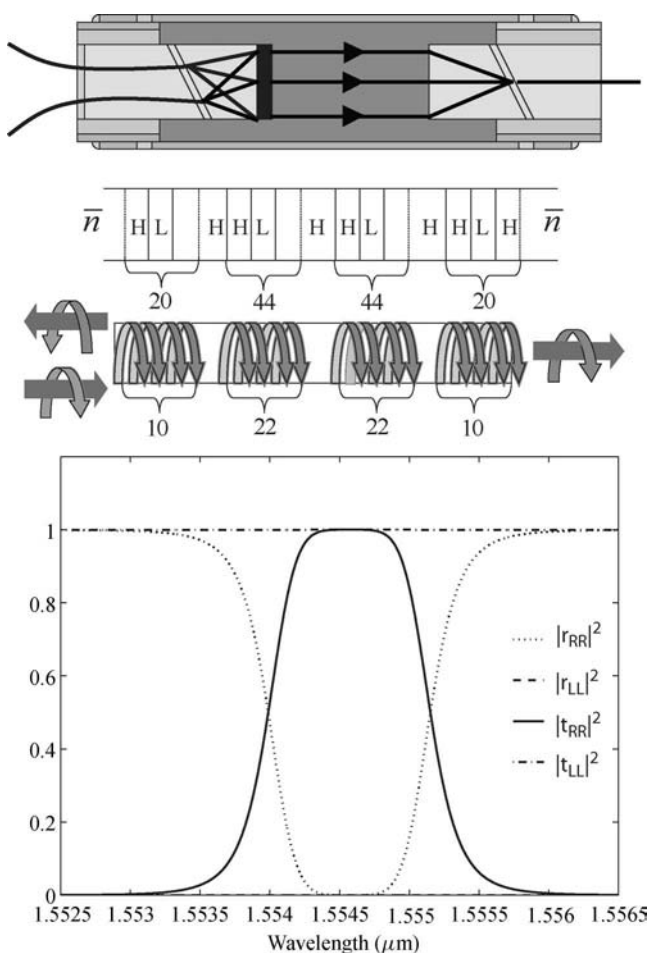


Figure 12 CSTF-equivalent wavelength division multiplexing filter.

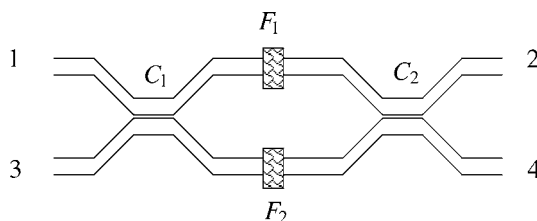


Figure 13 Schematic beam router/combiner based on CSTF technology.

The couplers C_1 and C_2 are 50–50 splitters at $\lambda_0 = 1.55 \mu\text{m}$. In the absence of the CSTF filters F_1 and F_2 , the device is a standard directional coupler which routes light entering port 1 to exit at port 4. With the CSTF filters present, the device becomes polarization-sensitive. For simplicity, we regard the fibers comprising the device as being polarization-maintaining. This is not a necessary restriction

but is imposed for ease of description. Suppose that the CSTFs are structurally left-handed, passing RCP but reflecting LCP. Thus, incident RCP at port 1 exits at port 4. Incident LCP, however is first split at C_1 , the equal components being reflected at F_1 and F_2 before combining again at C_1 to exit at port 3. In this way, the two polarizations incident at port 1 are filtered in the desired fashion.

Theoretical assessment of this device requires the application of standard coupled mode theory to the couplers C_1 and C_2 . The response of either $3dB$ couplers is quantified by calculating the amplitudes (A , B) exiting the coupler in terms of the amplitudes before the coupler (A_0 , B_0) as

$$\begin{aligned} A &= A_0 \cos\left(\frac{\pi \lambda_c}{4\lambda_0}\right) + i B_0 \sin\left(\frac{\pi \lambda_c}{4\lambda_0}\right), \\ B &= B_0 \cos\left(\frac{\pi \lambda_c}{4\lambda_0}\right) + i A_0 \sin\left(\frac{\pi \lambda_c}{4\lambda_0}\right), \end{aligned} \quad (60)$$

where λ_c is the coupler's optimal wavelength. The coupler response being polarization-independent, (60) can be applied to both LCP and RCP amplitudes. Let us apply (60) to describe propagation through C_1 and C_2 , while the full electromagnetic description is used for the passage of the polarized beams through the CSTF filters F_1 and F_2 . We presume that the beams entering the CSTF filters are collimated so that over a sufficiently small region, they may be regarded as plane waves.

The following parameters were selected for the CSTF filters: $n_c^2 = 2.72$ and $\tilde{n}_d^2 = 2.92$, $\Omega = -461.6$ nm (left-handed, and set to yield a Bragg wavelength $\lambda_0^{(Br)} = 1550$ nm), and $L = 60|\Omega|$. Suppose LCP is incident at port 1. The output intensities at ports 3 and 4 are shown in Figure 14. The LCP intensities $I_L^{(3)}$ and $I_L^{(4)}$ are evidently subjected to the polarization selectivity of the CSTF within the Bragg regime of the filter. The RCP intensities $I_R^{(3)}$ and $I_R^{(4)}$ are due to the small cross-scattered components in the CSTFs and are seen to be minimal.

The response to incident RCP at port 1 is shown in Figure 15. The device acts as a pure directional coupler for this polarization, corrupted slightly by the presence of the CSTF filters. The light is principally directed to port 4 and is wavelength-sensitive only through the minimal cross-scattering of the filters and through the mild wavelength-dependence of the couplers.

Another application of the device is as a polarization-sensitive beam combiner. LCP incident at port 1 and RCP at port 2 will yield a combined output at port 3. For equal input intensities, the LCP and RCP outputs at port 3 are given in Figure 16.

8.5 Issues for optical communications

To date, CSTFs have been fabricated to operate at optical wavelengths. As we have seen, the Bragg wavelength determines the central wavelength at which the

polarization-sensitive filter operates. Since fabrication of CSTFs involves controlled substrate rotation, it would appear to be relatively simple to fabricate CSTFs for telecommunication wavelengths: simply rotate the substrate at a rate appropriate to produce the desired pitch (and hence a Bragg wavelength of $1.55 \mu\text{m}$). However, simply taking established fabrication technology for CSTFs and extrapolating to the telecommunications wavelength regime may not be so simple. The key parameters for any material used in telecommunications are the absorption and dispersion characteristics, and the current materials used for CSTFs at visible wavelengths (e.g., TiO_2) are more dispersive at telecommunications wavelengths. In this regard, materials for CSTFs at telecommunications wavelengths should be sought that have minimal loss and dispersion at $1.55 \mu\text{m}$, but also have sufficient birefringence for CSTFs.

9 Conclusion

CSTFs form the most studied example of a class of complex mediums known as sculptured thin films. They exhibit the circular Bragg phenomenon wherein light co-handed with the material helicity is reflected, while the orthogonal state

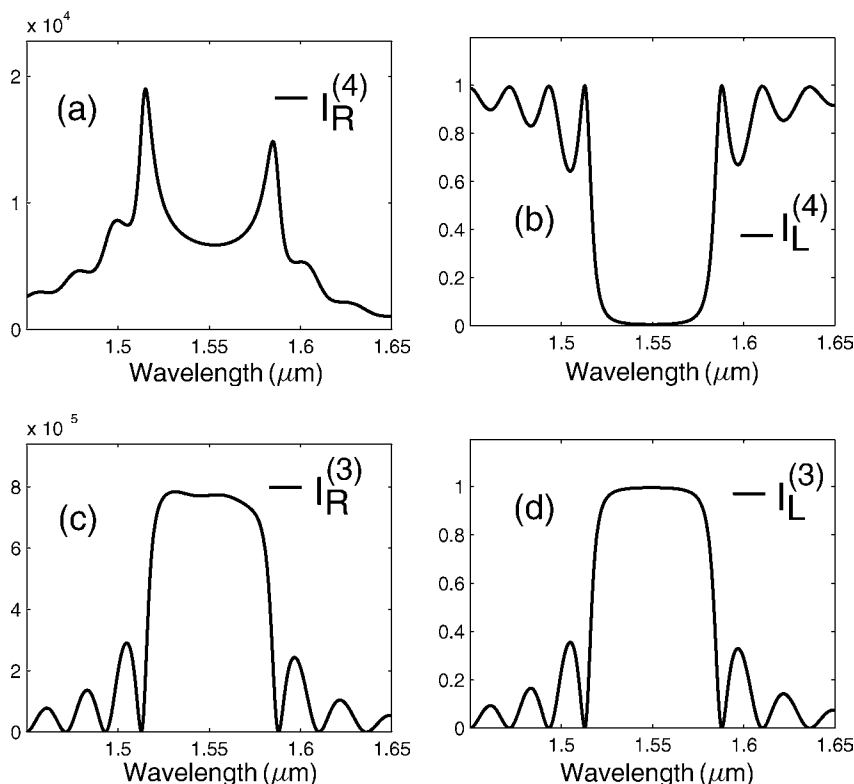


Figure 14 Output LCP and RCP intensities at ports 3 and 4 for incident LCP light at port 1. The CSTF filters are structurally left-handed.

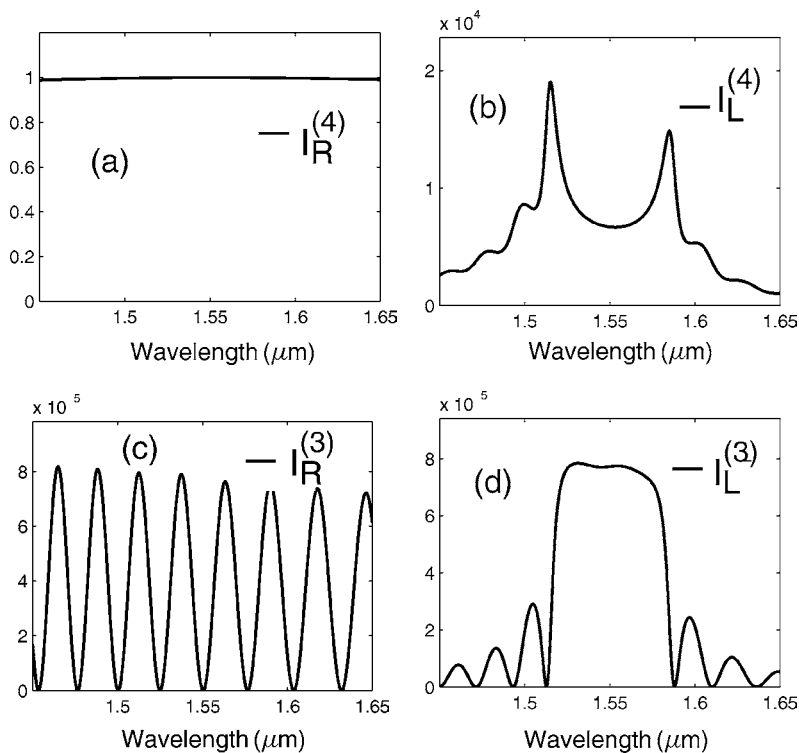


Figure 15 Output LCP and RCP intensities at ports 3 and 4 for incident RCP light at port 1. The CSTF filters are structurally left-handed.

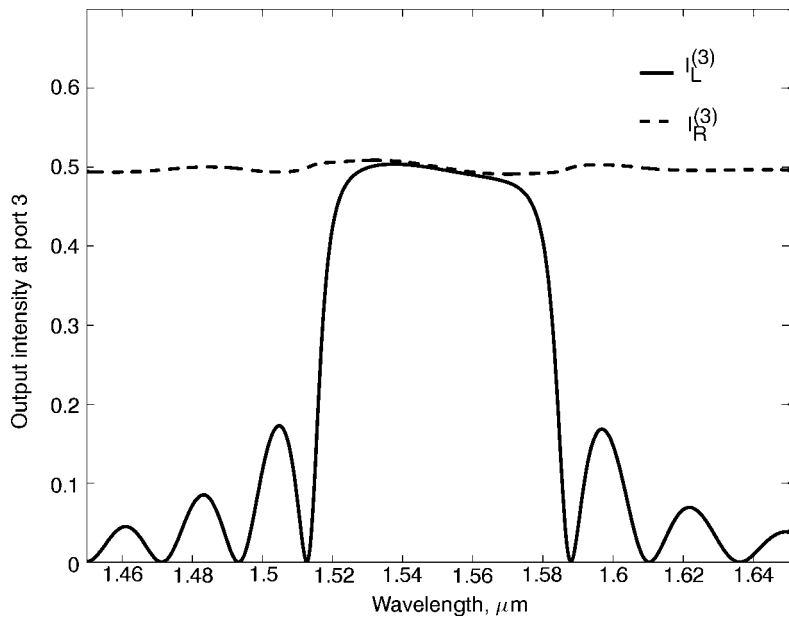


Figure 16 Output LCP and RCP intensities at port 3 for incident LCP light at port 1, and equal incident RCP at port 2. The filters are structurally left-handed.

is transmitted. A unique feature of CSTFs is that the Maxwell equations are exactly soluble for electromagnetic propagation along the helicoidal axis, enabling benchmark calculations to be made on proposed structures. For analytic evaluation, the coupled-wave approach is preferred. Various filter designs have been proposed which are a step towards lossless, in-line polarization management in optoelectronic systems.

References

- [1] R. Messier, V.C. Venugopal and P.D. Sunal, Origin and evolution of sculptured thin films, *J Vac Sci Technol A* **18** (2000), 1538–1545.
- [2] A. Lakhtakia and R. Messier, The past, the present, and the future of sculptured thin films, *Introduction to complex mediums for optics and electromagnetics* (W.S. Weiglhofer and A. Lakhtakia, eds), SPIE Press, Bellingham, WA, USA, 2003.
- [3] A. Lakhtakia and W.S. Weiglhofer, On light propagation in helicoidal bianisotropic mediums, *Proc Roy Soc Lond A* **448** (1995), 419–437; erratum: **454** (1998), 3275.
- [4] I. Hodgkinson and Q.h. Wu, Serial bideposition of anisotropic thin films with enhanced linear birefringence, *Appl Opt* **38** (1999), 3621–3625.
- [5] P. Yeh and C. Gu, *Optics of liquid crystal displays*, Wiley, New York, NY, USA, 1999.
- [6] W.S. Weiglhofer, Constitutive characterization of simple and complex mediums, *Introduction to complex mediums for optics and electromagnetics* (W.S. Weiglhofer and A. Lakhtakia, eds), SPIE Press, Bellingham, WA, USA, 2003.
- [7] A. Lakhtakia, Local inclination angle: a key structural factor in emission from chiral sculptured thin films, *Opt Commun* **202** (2002), 103–112; erratum: **203** (2002), 447.
- [8] T. Erdogan, Fiber grating spectra, *J Lightwave Technol* **15** (1997), 1277–1294.
- [9] A. Othonos, Fiber Bragg gratings, *Rev Sci Instrum* **68** (1997), 4309–4341.
- [10] M.W. McCall, Axial electromagnetic wave propagation in inhomogeneous dielectrics, *Math Comput Modelling* **34** (2001), 1483–1497.
- [11] A. Lakhtakia and W.S. Weiglhofer, Further results on light propagation in helicoidal bianisotropic mediums: oblique propagation, *Proc Roy Soc Lond A* **453** (1997), 93–105; erratum: **454** (1998), 3275.
- [12] H. Hochstadt, *Differential equations*, Dover Press, New York, NY, USA, 1975, Section 2.5.
- [13] J. Guckenheimer and P. Holmes, *Nonlinear oscillations, dynamical systems and bifurcations of vector fields*, Springer, New York, NY, USA, 1983, Section 1.1.

- [14] P.I. Rovira, R.A. Yarussi, R.W. Collins, R. Messier, V.C. Venugopal, A. Lakhtakia, K. Robbie and M.J. Brett, Transmission ellipsometry of a thin-film helicoidal bianisotropic medium, *Appl Phys Lett* **71** (1997), 1180–1182.
- [15] P.D. Sunal, A. Lakhtakia and R. Messier, Simple model for dielectric thin-film helicoidal bianisotropic media, *Opt Commun* **158** (1998), 119–126.
- [16] A. Lakhtakia, Energy flows in axially excited, locally biaxial, dielectric, helicoidal bianisotropic media (HBMs), *Opt Commun* **161** (1999), 275–286.
- [17] M.W. McCall and A. Lakhtakia, Development and assessment of coupled wave theory of axial propagation in thin-film helicoidal bianisotropic media. Part 1: reflectances and transmittances, *J Modern Opt* **47** (2000), 973–991.
- [18] M.W. McCall and A. Lakhtakia, Development and assessment of coupled wave theory of axial propagation in thin-film helicoidal bi-anisotropic media. Part 2: Dichroisms, ellipticity transformation and optical rotation, *J Modern Opt* **48** (2001), 143–158.
- [19] V. Belyakov, V.E. Dmitrienko and V.P. Orlov, Optics of cholesteric liquid crystals, *Sov Phys Usp* **22** (1979), 63–88.
- [20] I.J. Hodgkinson, Q.h. Wu, M. Arnold, M.W. McCall and A. Lakhtakia, Chiral mirror and optical resonator designs for circularly polarized light: suppression of cross-polarized reflectances and transmittances, *Opt Commun* **210** (2002), 201–211.
- [21] M.W. McCall, On the application of coupled mode theory for the modelling of fiber-Bragg gratings, *J Lightwave Technol* **18** (2000), 236–242.
- [22] V.C. Venugopal and A. Lakhtakia, Sculptured thin films: conception, optical properties and applications, *Electromagnetic fields in unconventional materials and structures* (O.N. Singh and A. Lakhtakia, eds), Wiley, New York, NY, USA, 2000.
- [23] I.J. Hodgkinson, Q.h. Wu, A. Lakhtakia and M.W. McCall, Spectral-hole filter fabricated using sculptured thin-film technology, *Opt Commun* **177** (2000), 79–84.
- [24] A. Lakhtakia, V.C. Venugopal and M.W. McCall, Spectral holes in Bragg reflection from chiral sculptured thin films: circular polarization filters, *Opt Commun* **177** (2000), 57–68.
- [25] K. Hill, Aperiodic distributed-parameter waveguides for integrated optics, *Appl Opt* **13** (1974), 1853–1856.
- [26] R. Kashyap, *Fiber Bragg gratings*, Academic Press, San Diego, CA, USA, 1999.
- [27] P. Baumeister, Bandpass filters for wavelength division multiplexing—modification of the spectral bandwidth, *Appl Opt* **37** (1999), 6609–6614.
- [28] M.W. McCall and A. Lakhtakia, Integrated optical polarization filtration via sculptured-thin-film technology, *J Modern Opt* **48** (2001), 2179–2184.



Martin W. McCall was born in London, UK, in 1962. He is currently a Reader in Photonics at Imperial College, London. Throughout his career he has worked at various times on nonlinear optics, nonlinear dynamics and electromagnetics of complex mediums. He is author or co-author of about 90 papers and an undergraduate text on classical mechanics.

Electromagnetics of Carbon Nanotubes

Sergey A. Maksimenko and Gregory Ya. Slepyan

Abstract

A review of theoretical results on electromagnetic properties of carbon nanotubes is presented, including the linear electrodynamics of nanotubes, non-linear transport, optical effects in nanotubes and foundations of their quantum electrodynamics. The method of effective boundary conditions is shown to be a universal tool for the study of electrodynamic problems of nanotubes. The concept of nanotube as a surface wave nanowaveguide in the infrared is developed, and scattering of electromagnetic waves at the nanotube edge is discussed. The high-efficiency generation of high-order harmonics in the nanotubes under the action of strong pumping fields is predicted. Experimental results on third-harmonic generation in a nanotube ensemble is compared with theoretical predictions. It is shown that, in a strong static field, the nanotubes exhibit negative differential conductivity. The formalism of electrodynamics of lossy dispersive mediums is applied to the problem of spontaneous radiation of an excited atom in the carbon nanotube.

1 Introduction

Rapid progress in the synthesis of a variety of different kinds of nanostructures with fascinating electronic and optical properties irreducible to properties of bulk mediums symbolizes a fundamental breakthrough in the physics and chemistry of condensed matter, significantly extending our knowledge of the nature of solids and our capabilities to control their properties. Solid-state nanostructures are nanoinhomogeneities of a different nature and geometry in semiconductor and dielectric mediums. Despite their different physical nature, these objects share the common property of having extremely small dimensions in one or several directions. These dimensions are approximately one or two orders of magnitude bigger than the characteristic interatomic distance. Under these conditions, the quantum nature of carrier motion is fully developed, and traditional constitutive parameters in electrodynamics defined through macroscopic averaging cannot be applied; at the very least, they require modification.

Milestones in the development of electrodynamics have always been related to practical problems arising from new ideas relating to the transmission and processing of electromagnetic signals. The advances of quantum electronics led to the development of the theory of open quasi-optical resonators [1]. The synthesis of high-quality optical fibers made fiber-optic communication feasible, which led to the development of the theory of open dielectric waveguides (including irregular and nonlinear waveguides) [2-3]. Progress in microwave microelectronics stimulated research in electrodynamics of microstrips and other planar structures [4]. Modern electromagnetic theory is characterized by the development of highly efficient numerical methods simulating diffraction from lossy bodies of an arbitrary spatial configuration [5]. Advances in physics and technology of solid-state carbon nanostructures led to the creation of a theoretical basis for their widespread application in nanoelectronics and other fields. Among them, research on the properties of carbon nanotubes (CNs)—quasi-one-dimensional carbon macromolecules—has continued to grow unabatedly [6]-[9]. Undoubtedly, electromagnetic simulation of nanostructures is one of the main directions for modern electrodynamics. In this regard, researchers face new problems, while the traditional electromagnetic methods gain new life in their application to new objects.

The first stage in modern quantum physics was completely associated with homogeneous mediums. This research resulted in the derivation of the dispersion equation for coupled states of the electromagnetic field and material particles. Solutions to this equation describe eigenwaves of the medium, quasi-particles, which differ from usual (free) particles by the complex behavior of their dispersion characteristic (energy versus quasi-momentum $\mathbf{p}$). Embedding nanoscale inhomogeneities into a homogeneous medium creates conditions for diffraction and scattering of quasi-particles and for their mutual transformation, just as this occurs with electromagnetic waves in usual irregular waveguides. An important role in this process is played by the resonance interaction between different modes and corresponding matching conditions. The first step in the incorporation of resonant interactions of quasi-particles was made in the theory of quantum semiconductor

superlattices (see, for example, Ref. 10), those high-frequency and optical properties turned out to be very unusual: negative differential conductivity, propagation of longitudinal (plasma) waves, and so on. Interaction of different modes in nanostructures appears to be even more complex due to the greater variety in the interacting modes and complex three-dimensional geometry of the inhomogeneities. Correspondingly, electronic and electromagnetic properties of nanomaterials also appear to be richer and more diverse. In particular, modern quantum theory of quasi-one-dimensional conductors predicts monomolecular electronic devices whose operation relies on quantum charge-transport processes [11-12]. CN-based transistors [13], tunneling diodes based on doped CN junctions [14], and Schottky diodes in CNs heterojunctions [15-16] are actively studied. One more important peculiarity of CNs is pronounced field-electron emission, a property that makes CNs attractive for cathodes in electronic devices [9].

Physical transport mechanisms—tunneling, avalanche and resistive instabilities, and interaction between charged quasi-particles and periodic potentials—are well known in solid-state physics. However, in nanotubes and in three-dimensional mediums, these mechanisms express themselves in a significantly different manner because of the quasi-one-dimensional conductivity of CNs. Therefore, the classical results for three-dimensional macroscopic bodies cannot be extended to nanotubes, which necessitates new fundamental investigations of electronic processes in quasi-one-dimensional macromolecules—and in CNs in particular. It should also be noted that many types of nanostructures exhibit chiral symmetry. Therefore, we can expect to observe many effects that have been uncovered in macroscopic linear and nonlinear chiral mediums.

In this chapter, we consider several issues of electrodynamics of and electron transport in CNs. The chapter is based on a series of works that cover problems of linear electrodynamics of carbon nanotubes [17]-[22], nonlinear transport in and nonlinear optics of CNs [22]-[27], and of the quantum electrodynamics (QED) of nanotubes [22], [28]. This is of interest in relation to the recent idea to use nano-objects in quantum networks that store and process quantum information being transmitted between nodes by entangled states of photons [29]. The approach developed here uses the analogy between electromagnetic fields in nanostructures and subjects of the macroscopic microwave electrodynamics [30], and includes well-known methods and techniques of macroscopic electrodynamics [5], [31]. The material is presented in the form of specific problems. It should, however, be noted that aside from the fundamental and applied interest in them, these particular problems are good illustrations of general trends in modern electrodynamics of nanostructures.

2 Electron transport in carbon nanotubes

2.1 Dispersion properties of π -electrons

Surface carbon structures, fullerenes and nanotubes, appear as a result of certain deformations of a planar monoatomic graphite layer (graphene), whose crystalline

structure is illustrated in Figure 1. In fullerenes, discovered in 1985 [32], the planar graphene surface is transformed into a closed sphere or spheroid containing regular hexagons (their number depends on the fullerene dimension) and 12 regular pentagons. In a CN, originally synthesized in 1991 [33], the graphite surface is transformed into an extended hollow cylindrical structure; see Figure 2. Thus, carbon atoms in CNs are situated regularly on a helical line with a certain wrapping angle (geometric chiral angle). Let $\mathbf{R}$ be the relative position vector in the graphene plane drawn between two sites on its honeycomb lattice, as shown in Figure 1. In terms of the lattice basic vectors $\mathbf{a}_1$ and $\mathbf{a}_2$, $\mathbf{R} = m\mathbf{a}_1 + n\mathbf{a}_2$, where m and n are integers. Thus, the geometric configuration of CNs can be classified by a dual index (m, n) —with $(m, 0)$ for *zigzag* CNs, (m, m) for *armchair* CNs, and $0 < n \neq m$ for *chiral* CNs. The cross-sectional radius of a CN and its chiral angle are given by [9]

$$R_{\text{cn}} = \frac{\sqrt{3}}{2\pi} b \sqrt{m^2 + mn + n^2}, \quad \theta_{\text{cn}} = \tan^{-1} \left(\frac{\sqrt{3}n}{2m + n} \right), \quad (1)$$

where $b = 0.142$ nm is the interatomic distance in graphene. Typically, CNs are $0.1\text{--}10\ \mu\text{m}$ in length, their cross-sectional radius varies within the range $1\text{--}10$ nm, while $0 \leq \theta_{\text{cn}} \leq 30^\circ$. Recently, synthesis of CNs of extremely small radius of

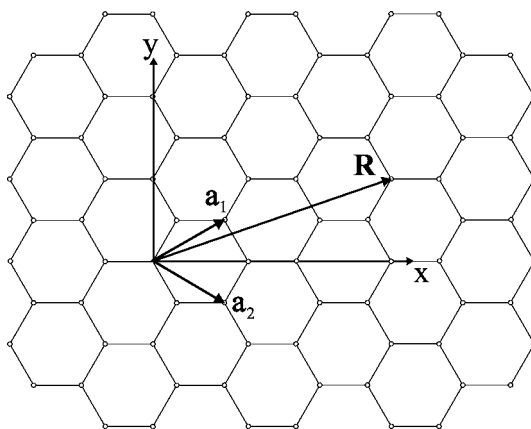


Figure 1 Configuration of the graphene crystalline lattice, $\mathbf{R} = m\mathbf{a}_1 + n\mathbf{a}_2$.

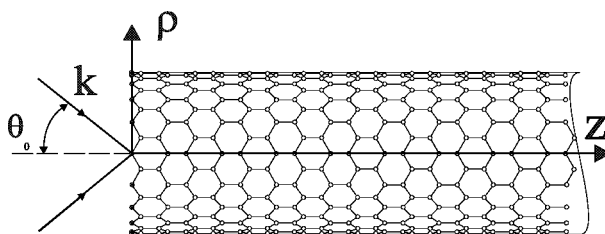


Figure 2 Model of an open-ended carbon nanotube. The wavevector $\mathbf{k}$ shows the direction of propagation of the cylindrical wave in the analysis of edge effects in nanotubes.

~ 0.4 nm has been reported [34]. In this chapter, we use a two-dimensional Cartesian coordinate system (x, y) for graphene and the circular cylindrical coordinate system (ρ, ϕ, z) for any CN, with the CN axis parallel to the z axis. The x axis is oriented along a hexagonal side; see Figure 1. The transition from graphene to a zigzag CN is established by the substitution $\{x \rightarrow z, y \rightarrow \phi\}$, while the transition from graphene to an armchair CN requires the substitution $\{y \rightarrow z, x \rightarrow \phi\}$. Gaussian units are used throughout, in conformity with CN literature.

Both single-wall and multiwall nanotubes have been synthesized [9]. The multilayer nanotubes have the form of N coaxial cylinders (the distance between the layers is 0.34 nm and $N \sim 10$ –12). Although the theory presented here has been developed for a single-wall CN, it should be noted that a multiwall CN can be treated as an ensemble of single-wall tubules, with a broad diameter distribution [19]. The weak intertube interactions in multiwall tubules only affect the low-energy electronic structure for $E < 0.1$ eV [35]. Along with carbon nanotubes, nanotubes doped with nitrogen and boron are also known [9].

Electromagnetic processes in any medium essentially depend on its electronic properties. Properties of electrons in nanotubes and electron transfer processes in them have been studied in detail both theoretically and experimentally. The theoretical analysis is usually confined to dynamics of π -electrons within the tight-binding approximation [36–37], which allows for interaction between only three adjacent atoms of the hexagonal structure. In the framework of this model, electron properties of graphene are described by the well-known dispersion law [38]

$$\mathcal{E}_{c,v}(\mathbf{p}) = \pm \gamma_0 \sqrt{1 + 4 \cos(ap_x) \cos\left(\frac{a}{\sqrt{3}}p_y\right) + 4 \cos^2\left(\frac{a}{\sqrt{3}}p_y\right)}, \quad (2)$$

where $\gamma_0 \approx 2.7$ eV is the overlapping integral, $a = 3b/2\hbar$, $\hbar$ is the Planck constant, and $p_{x,y}$ are the projections of the quasi-momentum. The upper and lower signs in (2) refer to the conduction and valence bands, marked by the indexes c and v , respectively. The range of definition of the quasi-momentum $\mathbf{p}$ (the first Brillouin zone) are the hexagons shown in Figure 3. The vertexes are the Fermi points where

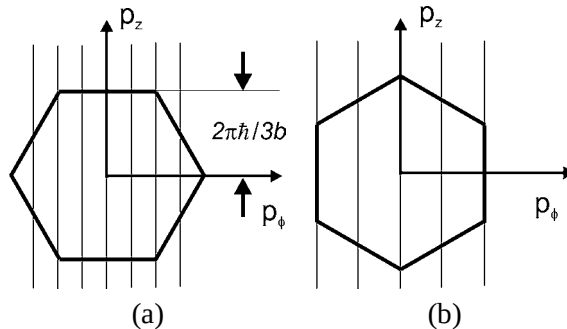


Figure 3 Configuration of the first Brillouin zone for (a) zigzag and (b) armchair CNs.

$\mathcal{E} = 0$, which is indicative of the absence of the forbidden zone for π -electrons in graphene. Recall that graphene is a semimetal: it lacks a bandgap, but the density of states at the Fermi level is zero.

The dispersion properties of nanotubes essentially differ from dispersion properties of graphene because of the difference in topology. In the cylindrical structure, electrons residing at the origin and at the point $\mathbf{R} = m\mathbf{a}_1 + n\mathbf{a}_2$ are identical, which quantizes the transverse quasi-momentum component:

$$p_\varphi = \frac{\hbar s}{R_{\text{cn}}}, \quad s = 1, 2, \dots, m. \quad (3)$$

The axial component of quasi-momentum p_z remains continuous. Relationships (2) and (3) and the substitution $\{p_x \rightarrow p_z, p_y \rightarrow p_\varphi\}$ yields the dispersion law for the zigzag nanotubes:

$$\mathcal{E}_{c,v}(p_z, s) = \pm \gamma_0 \sqrt{1 + 4 \cos(ap_z) \cos\left(\frac{\pi s}{m}\right) + 4 \cos^2\left(\frac{\pi s}{m}\right)}. \quad (4)$$

In order to evaluate the electron dispersion relation for armchair CNs from (2), the substitution $\{p_x \rightarrow p_\varphi, p_y \rightarrow p_z\}$ must be carried out; accordingly,

$$\mathcal{E}_{c,v}(p_z, s) = \pm \gamma_0 \sqrt{1 + 4 \cos\left(\frac{\pi s}{m}\right) \cos\left(\frac{a}{\sqrt{3}} p_z\right) + 4 \cos^2\left(\frac{a}{\sqrt{3}} p_z\right)}. \quad (5)$$

As follows from relationships (4) and (5), the first Brillouin zone in nanotubes is no longer a hexagon, but rather has the form of a set of one-dimensional zones: rectilinear segments inside the hexagon. Depending on whether these lines pass through the hexagon vertexes (Fermi points), the bandgap in the electron spectrum either disappears or appears. Unlike graphene, the density of states at the Fermi level in one-dimensional zones is nonzero. Accordingly, the nanotube is either metal or a semiconductor. As can be seen from Figure 3, the armchair nanotubes exhibit metallic conductivity at any m , whereas the zigzag nanotubes behave as a metal only for $m = 3q$, where q is an integer. For a metal nanotube of a small radius and for a nanotube of a very large radius ($m \rightarrow \infty$), Mintmire and White [39] have proposed the approximate dispersion law for π -electrons: $\mathcal{E}_{c,v}(\mathbf{p}) = \pm v_F |\mathbf{p} - \mathbf{p}_F|$. Here, $v_F = a\gamma_0$ is the velocity of π -electrons at the Fermi level and $\mathbf{p}_F$ is a constant vector defined as the quasi-momentum at the Fermi level. In both cases, the approximate dispersion law is applicable, because the regions near the Fermi points give the maximum contribution to the conductivity.

When placed in either an axial [40] or a transverse [41] magnetostatic field, the nanotube changes its type of conductivity. Due to this feature, the conductivity can be controlled over a wide range by varying the magnetization. An important property of the chiral nanotubes is that a voltage applied across its ends

produces an azimuthal current component [23], [42]-[44]. As a result, the trajectory of the current in the nanotube is helical, although the graphene conductivity is isotropic.

2.2 Bloch equation for π -electrons in carbon nanotubes

The theory of optical properties of CNs applied in the present chapter involves a direct solution of the quantum-mechanical equations of motion for π -electrons. Consider an infinitely long rectilinear single-wall CN oriented along the z axis and excited by the component of the electromagnetic field polarized along this axis: $\mathbf{E}(\mathbf{r}, t) = \mathbf{e}_z E_z(\mathbf{r}, t)$. Let the field be incident normally to the CN axis. In the tight-binding approximation, the motion of electrons in the CN crystalline lattice potential $W(\mathbf{r})$ is described by the Schrödinger equation

$$i\hbar \frac{\partial \Psi}{\partial t} = -\frac{\hbar^2}{2m_0} \Delta \Psi + [W(\mathbf{r}) - e(\mathbf{E} \cdot \mathbf{r})]\Psi, \quad (6)$$

where e and m_0 are the electron charge and mass. The solution can be represented by the Bloch wave expansion

$$\Psi = \sum_q C_q \Psi_q(\mathbf{p}, \mathbf{r}), \quad (7)$$

where the integral index q stands for the collection of quantum numbers characterizing states of π -electrons with a given quasi-momentum; in the framework of the two-band model, the index takes the values either v or c . $\Psi_q(\mathbf{p}, \mathbf{r}) = \hbar^{-1/2} \exp(i\mathbf{p}\mathbf{r}/\hbar) u_q(\mathbf{r})$ are the Bloch functions whose amplitudes $u_q(\mathbf{r})$ are periodic with respect to an arbitrary lattice vector $\mathbf{R}$ (see Figure 1): $u_q(\mathbf{r} + \mathbf{R}) = u_q(\mathbf{r})$.

Expansion (7) does not contain states of the continuous spectrum; consequently, we restrict consideration to the effects below the ionization threshold. Using expansion (7), it can easily be shown that coefficients C_q satisfy to the equation as follows (for details see, for example, Ref. 45):

$$i\hbar \frac{\partial C_q}{\partial t} = \mathcal{E}_q C_q - i\hbar e E_z \frac{\partial C_q}{\partial p_z} - e E_z \sum_{q'} C_{q'} R_{qq'}, \quad (8)$$

where

$$R_{qq'} = \frac{i\hbar}{2} \int_{\Omega} \left(u_q^* \frac{\partial u_{q'}}{\partial p_z} - \frac{\partial u_q^*}{\partial p_z} u_{q'} \right) d^3 \mathbf{r}. \quad (9)$$

Integration is carried out over the elementary volume Ω , with the honeycomb cell as the cross-section. Using standard representation of the density matrix elements

$\rho_{qq'} = C_q C_{q'}^*$, (8) can be transformed to the system as follows [27]:

$$\begin{aligned} \frac{\partial \rho_{vv}}{\partial t} + e E_z \frac{\partial \rho_{vv}}{\partial p_z} &= -\frac{i}{\hbar} e E_z (R_{vc}^* \rho_{vc} - R_{vc} \rho_{cv}), \\ \frac{\partial \rho_{vc}}{\partial t} + e E_z \frac{\partial \rho_{vc}}{\partial p_z} &= -\frac{i}{\hbar} e E_z [R_{vc}(2\rho_{vv} - 1) - \Delta R \rho_{vc}] - i \omega_{vc} \rho_{vc}, \\ \rho_{vv} + \rho_{cc} &= 1, \end{aligned} \quad (10)$$

where $\omega_{vc} = (\mathcal{E}_c - \mathcal{E}_v)/\hbar$ is the frequency of the transition. Here, the transition frequency as well as the matrix elements R_{vc} and $\Delta R = R_{vv} - R_{cc}$ are evaluated in the tight-binding approximation, taking into account transverse quantization of the charge carriers' motion and the hexagonal structure of the CN crystalline lattice. For zigzag CNs the matrix element R_{vc} takes the form

$$R_{vc}(p_z, s) = -\frac{b\gamma_0^2}{2\mathcal{E}_c^2(p_z, s)} \left[1 + \cos(ap_z) \cos\left(\frac{\pi s}{m}\right) - 2 \cos^2\left(\frac{\pi s}{m}\right) \right]. \quad (11)$$

The analogous expression for armchair (m, m) CNs is as follows:

$$R_{vc}(p_z, s) = -\frac{\sqrt{3}b\gamma_0^2}{2\mathcal{E}_c^2(p_z, s)} \sin\left(\frac{a}{\sqrt{3}}p_z\right) \sin\left(\frac{\pi s}{m}\right). \quad (12)$$

For CNs of both types, the relation $\Delta R/R_{vc} \ll 1$ can easily be derived; this allows us to reject the term ΔR in (10).

The system (10), supplemented with (4), (5), (11), and (12), constitutes a basic system for the analysis of the optical properties of CNs. Generally, its solution is nonlinear with respect to E_z . In the linear regime, the system (10) can be linearized. In the case of weak nonlinearity, optical properties of CNs are derived from (10) using the power expansion in E_z . Thus, in that case, nonlinear properties of CNs are characterized by nonlinear optical susceptibilities of different orders. In high-intensity external fields of subpicosecond optical pulses ($> 10^{10}$ W/cm²), the formalism of nonlinear susceptibilities becomes inefficient. In Ref. 27, the system (10) was solved numerically in the time domain. The method of characteristics [46] was used for integration of the system. Initial distribution of electrons in zones was specified by the Fermi equilibrium distribution function at room temperature and the periodicity on the Brillouin zone boundaries has been utilized.

Solution of (10) allowed us to evaluate the surface density of the induced axial current:

$$\begin{aligned} j_z(t) = \frac{2e\gamma_0}{(\pi\hbar)^2 R_{cn}} \sum_{s=1}^m \int \left\{ \frac{\partial \mathcal{E}_c(p_z, s)}{\partial p_z} \rho_{vv}(p_z, s) \right. \\ \left. + \mathcal{E}_c(p_z, s) R_{vc}(p_z, s) \text{Im}[\rho_{vv}(p_z, s)] \right\} dp_z. \end{aligned} \quad (13)$$

Generally, relaxation terms describing inelastic scattering of π -electrons propagating in CNs should be introduced into (10). This can be done phenomenologically [22] (e.g., in the framework of the relaxation-time approximation [10]) or on the basis of microscopic theory of electron-phonon interactions [41], [47]. An alternative approach is to solve (10) without relaxation terms and introduce corresponding corrections into the final results instead.

3 Linear electrodynamics of carbon nanotubes

3.1 Dynamic conductivity of carbon nanotubes

As was pointed out in the previous section, to reveal the linear optical response of a CN we must solve the linearized equations (10) and evaluate the surface current density. In the linear regime, the Bloch equations can be solved in the weak-field limit by the Fourier transform method. If we neglect therein the spatial dispersion, the optical response proves to be spatially local. In that case, for Fourier amplitudes of the axial current and field we obtain the relation [22]

$$j_z(\omega) = \sigma_{zz}(\omega) E_z(\omega), \quad (14)$$

where the axial conductivity of the CN is given by the following expression

$$\sigma_{zz}(\omega) = -\frac{ie^2\omega}{\pi^2\hbar R_{\text{cn}}} \left\{ \frac{1}{(\omega + i0)^2} \sum_{s=1}^m \int_{1st\text{ BZ}} \frac{\partial F_c(p_z, s)}{\partial p_z} \frac{\partial \mathcal{E}_c(p_z, s)}{\partial p_z} dp_z \right. \\ \left. - 2 \sum_{s=1}^m \int_{1st\text{ BZ}} \mathcal{E}_c(p_z, s) |R_{vc}(p_z, s)|^2 \right. \\ \left. \times \frac{F_c(p_z, s) - F_v(p_z, s)}{\hbar^2(\omega + i0)^2 - 4\mathcal{E}_c^2(p_z, s)} dp_z \right\}. \quad (15)$$

Here

$$F_{c,v}(p_z, s) = \frac{1}{1 + \exp \left[\frac{\mathcal{E}_{c,v}(p_z, s) - \mu}{k_B T} \right]} \quad (16)$$

is the equilibrium Fermi distribution function, T is the temperature, k_B is the Boltzmann constant, and μ is the chemical potential; in graphite and undoped CNs, $\mu = 0$. Integration in (15) is performed over the first Brillouin zone.

The CN dynamic conductivity law (14) is analogous to constitutive relations for three-dimensional conducting mediums in classical electrodynamics. However, there is a significant distinction: as in classical electrodynamics, the derivation of

(15) employed the macroscopic averaging, but a surface element was used instead of an infinitesimally small volume. Thus, $j_z(\omega)$ is the surface current density.

The relaxation effect is phenomenologically incorporated into (15) by substituting $(\omega + i0)^2 \rightarrow \omega(\omega + i\nu)$, where $\nu = 1/\tau$ is the relaxation frequency. The mean time of the electronic free pass for nanotubes we estimate as $\tau = 3 \times 10^{-13}$ s [48]. The first term in (15) describes intraband motion of π -electrons and corresponds to the first term in (13). The second term describes direct transitions between valence and conductivity bands and corresponds to the second term in (13). Note that the contribution of interband transitions is negligible in the frequency region determined by the condition

$$\omega < \omega_l = \begin{cases} 2\nu_F/R_{\text{cn}}, & \text{for metallic tubules,} \\ 2\nu_F/3R_{\text{cn}}, & \text{for semiconducting tubules.} \end{cases} \quad (17)$$

For typical nanotubes, ω_l falls into the infrared region. Figure 4 illustrates the behavior of the axial conductivity at frequencies of optical transitions.

Figure 5 shows $\sigma_{zz}(\omega)$ for zigzag CNs as a function of the radius (index m). For armchair CNs this function is monotonic, because these tubules always exhibit metallic conductivity. Irrespective of the nanotube type, its conductivity tends to the same limit equal to the graphene conductivity as $m \rightarrow \infty$ [38]. The results presented above demonstrate that a rigorous microscopic transport theory must be utilized for elaboration of the electrodynamics of CNs: constitutive relations for CNs cannot be properly introduced without such a theory. In particular, phenomenological models proposed in Refs. 49–51 prove to be unsatisfactory since they assume the CN conductivity to be identical to the graphene conductivity. Figure 4 shows that such an approximation is adequate only for large-radius CNs ($m > 100$),

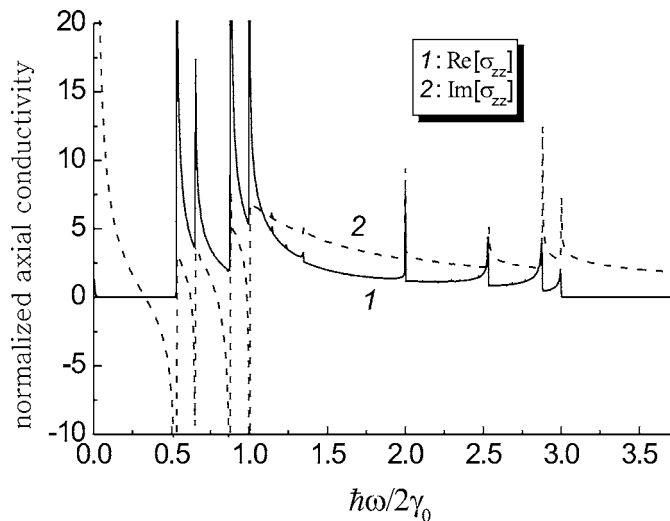


Figure 4 Frequency dependence of the total axial conductivity $\sigma_{zz}(\omega)$ of the (9, 0) metallic zigzag CN; $\tau = 3 \times 10^{-12}$ s and $T = 295$ K. The axial conductivity is normalized by $e^2/2\pi\hbar$.

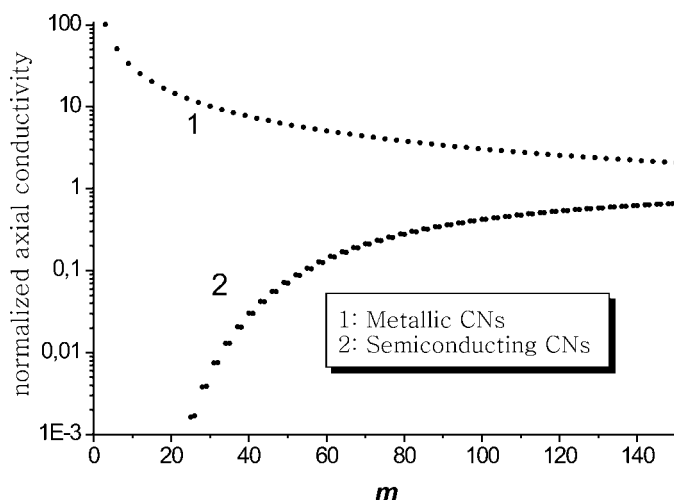


Figure 5 Normalized semiclassical conductivity $\sigma_{zz}(\omega)/\sigma_{\infty}$ for zigzag CNs as a function of m (and therefore of the cross-sectional radius R_{cn}); $\sigma_{\infty} = \lim_{m \rightarrow \infty} \sigma_{zz}$, $\tau = 3 \times 10^{-12}$ s, and $T = 264$ K.

where specific properties of CNs as low-dimensional structures do not manifest themselves since the role of transverse quantization becomes negligible.

3.2 Effective boundary conditions for CNs

Using the model of axial conductivity introduced in the previous section, we can now impose effective boundary conditions (EBCs) on the nanotube surface. This approach provides a general method for solving a wide range of problems of nanotube electrodynamics. The basic idea of the method is that the periodic structure is replaced with a smooth one-sided surface, on which appropriate EBCs for the electromagnetic field are imposed. These conditions are chosen in such a manner that the spatial structure of the electromagnetic field induced by the effective current that flows on the smooth homogeneous surface and the spatial structure of the electromagnetic field generated by the real current in the lattice are identical at a certain distance from the surface. The lattice parameters are included in the so-called EBC coefficients.

The EBCs are obtained as a result of the spatial averaging of the macroscopic fields over a physically infinitesimal element of the cylindrical surface. The condition that the tangential electric field component and the axial component of the magnetic field be continuous on the CN surface yields

$$\begin{aligned} E_{\varphi,z} \big|_{\rho=R_{\text{cn}}+0} - E_{\varphi,z} \big|_{\rho=R_{\text{cn}}-0} &= 0, \\ H_z \big|_{\rho=R_{\text{cn}}+0} - H_z \big|_{\rho=R_{\text{cn}}-0} &= 0. \end{aligned} \quad (18)$$

The next condition follows from the equation for the CN axial conductivity [19-20]. Its derivation utilizes the relation between the surface current density $j_z(\omega)$ and the discontinuity of the magnetic field component H_φ at the CN surface:

$$H_\varphi|_{\rho=R_{\text{cn}}+0} - H_\varphi|_{\rho=R_{\text{cn}}-0} = \frac{4\pi}{c} \sigma_{zz}(\omega) E_z|_{\rho=R_{\text{cn}}}. \quad (19)$$

In the regime of optical transitions, the electromagnetic response of a CN is significantly influenced by the spatial dispersion of π -electrons, provided, in particular, by the Coulomb screening effect. A theory of this effect in quasi-one-dimensional structures has been presented in Ref. 52. The spatial dispersion results in the CN conductivity σ_{zz} becoming a one-dimensional integral operator. As an example, consider the propagation in the CN along its axis of a traveling wave with a nonzero z component of the electric field: $E_z(\mathbf{r}, t) = \text{Re}\{E_z^0 \exp[i(hz - \omega t)]\}$. The planewave propagation considered in Section 2.2 corresponds to the particular case when $h = 0$. For such a traveling wave, the conductivity acquires a dependence on the wavenumber h , $\sigma_{zz} = \sigma_{zz}(h, \omega)$ and a concise expression for $\sigma_{zz}(h, \omega)$ is available [20]. Modification of the EBCs with the spatial dispersion accounted for has been presented in Refs. 18-20. The spatial dispersion is incorporated into EBCs by the change $\sigma_{zz}(\omega) \rightarrow \sigma_{zz}(\omega)[1 + \gamma(\omega)\partial^2/\partial z^2]^{-1}$, where $\gamma(\omega) = l_0/[k(1 + i\nu/\omega)]^2$, $k = \omega/c$ is the free-space wavenumber and the coefficient

$$l_0 = \frac{k^2}{2\sigma_{zz}(0, \omega)} \left. \frac{\partial^2 \sigma_{zz}(h, \omega)}{\partial h^2} \right|_{h=0} \left(1 + i\frac{\nu}{\omega}\right)^2 \quad (20)$$

characterizes the contribution of the spatial field inhomogeneity. After taking the foregoing into account, the EBC (19) changes to

$$\left[1 + \gamma(\omega)\frac{\partial^2}{\partial z^2}\right] \left(H_\varphi|_{\rho=R_{\text{cn}}+0} - H_\varphi|_{\rho=R_{\text{cn}}-0}\right) = \frac{4\pi}{c} \sigma_{zz}(\omega) E_z|_{\rho=R_{\text{cn}}}, \quad (21)$$

while (18) remain valid. The estimates $l_0 \approx (v_F/c)^2$ for metallic nanotubes of a small radius and $l_0 \approx 3(v_F/c)^2/4$ for any nanotube of a large radius were obtained in Refs. 19-20. For metallic CNs, calculations yield $l_0 \sim 10^{-5}$.

The conditions (18) and (19) or (18) and (21) constitute a complete system of EBCs for the electromagnetic field on the nanotube surface. They are analogous to the Weinstein–Sivov boundary condition [53-54] for grid structures and small-period grids in the macroscopic electrodynamics of microwaves.

3.3 Surface electromagnetic waves in nanotubes

To exemplify the EBC method, let us examine propagation of surface waves along an isolated infinite CN in free space and assume that the nanotube exhibits axial conductivity. The eigenwaves under study satisfy the homogeneous Maxwell

equations, boundary conditions (18) and (19), and the condition that there are no exterior current sources at infinity. The problem thus formulated is similar to the eigenwave problem for microwave slow-wave helical structures and can be solved by the field-matching technique.

The entire space is divided into two cylindrical partial domains—the domains inside and outside the tube. As in Ref. 53, we represent the electromagnetic field to be found in terms of the scalar Hertz potential Π_e and then, using the Maxwell equations and the radiation conditions ($\rho \rightarrow \infty$), we obtain

$$\Pi_e = A \exp(ihz + il\varphi) \begin{cases} I_l(\kappa\rho)K_l(\kappa R_{\text{cn}}), & \rho < R_{\text{cn}}, \\ I_l(\kappa R_{\text{cn}})K_l(\kappa\rho), & \rho > R_{\text{cn}}, \end{cases} \quad (22)$$

where A is an arbitrary constant; $\kappa = \sqrt{h^2 - k^2}$; while I_l and K_l are the modified Bessel functions of the first and second kind, respectively. The representation (22) directly satisfies the EBCs (18). Using an expression for the Wronskian of the modified Bessel functions, we obtain the dispersion relation for the surface wave in a CN:

$$\left(\frac{\kappa}{k}\right)^2 I_l(\kappa R_{\text{cn}})K_l(\kappa R_{\text{cn}}) = \frac{ic}{4\pi k R_{\text{cn}} \sigma_{zz}} \left[1 - \frac{1 + (\kappa/k)^2}{(1 + i\nu/\omega)^2} l_0 \right]. \quad (23)$$

Figure 6 shows the complex-valued slow-wave coefficient $\beta = k/h$ for the axially symmetric ($l = 0$) surface wave in the metallic (9, 0) CN obtained numerically from (23). Axially asymmetric modes are discussed in detail in Ref. 19. At low frequencies ($\omega < \nu$), when $kb < 10^{-7}$ (b is the C—C bond length; see Section 2.1) the nanotube demonstrates strong attenuation: $\text{Im}[\beta] \sim \text{Re}[\beta]$. We can thus conclude that the nanotubes are of no interest as surface waveguides at low frequencies. It is important that for nanotubes of typical lengths of $l_{\text{cn}} \sim 1 \mu\text{m}$, $l_{\text{cn}}\text{Re}[h] \ll 1$. This means that the CNs transmit low-frequency electric signals similar to electric

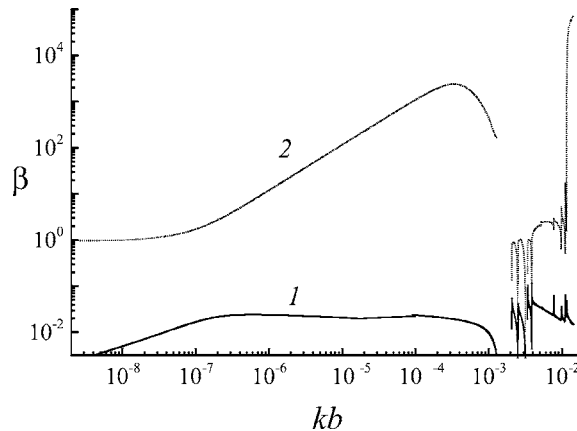


Figure 6 Frequency dependence of the complex-valued slow-wave coefficient β for a polar-symmetric surface wave in a (9, 0) metallic zigzag CN. Input parameters are the same as for Figure 4. 1: $\text{Re}[\beta]$, 2: $-\text{Re}[\beta]/\text{Im}[\beta]$.

circuits without wave effects. Unlike the low-frequency limit, in the infrared range ($10^{-5} < kb < 10^{-3}$ or $3 \times 10^{12} \text{ s}^{-1} < \omega/2\pi < 3 \times 10^{14} \text{ s}^{-1}$), the nanotubes propagate slowly decaying surface waves. The analysis has shown that, in the infrared range, the slow-wave coefficient $\text{Re}[\beta]$ of the semiconductor CNs is 1/10 that of the metallic ones, while the respective values of $\text{Im}[\beta]$ are comparable. Therefore, attenuation in semiconductor CNs is significantly higher than in the metallic ones. Moreover, since the semiconductor CNs are characterized by high slow-wave coefficients ($2 \times 10^{-3} < \text{Re}[\beta] < 2 \times 10^{-2}$), the electromagnetic field in such tubules is tightly localized near the surface.

We should also note that the slow-wave coefficient $\text{Re}[\beta]$ and phase velocity of the surface wave $v_{ph} = \text{Re}[\omega/h]$ are almost frequency-independent. Therefore, a wave packet will propagate in the nanotube without significant distortions, which is very important for possible application in nanoelectronics. This demonstrates that CNs can serve in the infrared regime as dispersionless surface-wave nanowaveguides, which may become high-efficiency nanoelectronic elements.

Such surface waves will also exist in the multiwall nanotube, which can be represented as a set of single-wall nanotubes. It has been shown that the electron interaction between the layers is negligible. This means that electrons move in the layer as in a single-wall nanotube of the same radius. In this situation, boundary conditions (18) and (21) remain valid for each layer, though the layer conductivities are different. Thus, a rigorous formulation of the surface wave problem for the N -layer tubule should consider $N + 1$ coaxial partial regions with boundary conditions (18) and (21) imposed at their boundaries. Characteristics of the surface wave can be efficiently controlled by choosing an appropriate number of layers.

Previously we have developed the concept of nanotubes as regular nanowaveguides. Practical application of CNs as waveguiding structures and antenna elements requires generation of different types of irregularities in CNs like those formed in ordinary macroscopic waveguides in the microwave range. There are several types of irregularities observed experimentally; e.g., junction of two CNs with different diameters [55], T -junctions formed by fusing two CNs of different diameters and chiralities perpendicular to each other [56], differently configured Y -junctions [56]-[58], crossed CN junctions [59], and a setup of two CNs connected to a gapped superconductor [60]. Note that embedding irregularities into a macroscopic waveguide does not change its electronic properties but manifests itself in the scattering of electromagnetic waves and mode transformation. In contrast, irregularities in a nanowaveguide may drastically change the conductivity's character. In our approach, this can be taken into account by means of corresponding transformation of EBCs (18) and (19) or (18) and (21) in the region adjoining the irregularity.

3.4 Edge effects in nanotubes

There are two alternative mechanisms of manifestation of edge effects in CNs: electronic and electromagnetic. The first mechanism is provided by the modification of the CN electronic structure entailed by the edge; in particular, new elec-

tronic states are localized in the vicinity of the edge [61]. The second mechanism is related to the electromagnetic wave diffraction at the CN edge. Here we concentrate on the electromagnetic diffraction edge effects, which are similar to those that occur in wire antennas [62]. Indeed, at optical frequencies, the length and radius of real nanotubes satisfy the conditions $kR_{\text{cn}} \ll 1$ and $kl_{\text{cn}} \sim 1$. These are the same conditions that characterize microwave wire antennas. This analogy not only indicates the importance of studying resonances associated with the finite length of nanotubes, but also indicates the analytical method: the problem of surface wave diffraction by an open end of the semi-infinite nanotube is solved by the Wiener–Hopf technique [62]. A finite nanotube can be analyzed with the help of the modified factorization method or by using the approximate solutions of integral equations for induced current as is done in antenna theory.

Consider the diffraction of an E -polarized cylindrical electromagnetic wave by an open end of a nanotube. Let the wave travel at the angle θ_0 to the tubule axis (see Figure 2). The scalar Hertz potential of this wave is well known:

$$\Pi_e^{(\text{inc})} = -\frac{i}{k \sin^2 \theta_0} H_l^{(\mu)}(k\rho \sin \theta_0) \exp(ikz \cos \theta_0 + il\varphi), \quad (24)$$

where $H_l^{(\mu)}$ are the cylindrical Hankel functions of the first and second kinds ($\mu = 1$ or 2). The scalar Hertz potential of the scattered field satisfies the Helmholtz equation and is related to the electromagnetic field in the standard manner [53], [62]. This allows us to use EBCs (18) and (21) to derive boundary conditions for the potential (for details, see Refs. 21 and 22). These conditions should be supplemented with the radiation conditions, which follow from the ultimate absorption principle, and with the edge conditions, which requires that no source is present on the sharp edge. The edge condition implies that the field energy in any finite spatial region containing the edge is finite.

The boundary-value problem formulated can be solved by the Wiener–Hopf technique. In the following, we proceed as in Ref. 21. For convenience, the space is assumed to be filled by a lossy medium, in which $k = k' + ik''$. The limit $k'' \rightarrow 0$ is carried out in the final expressions. Application of the Jones approach [63] leads us to a functional equation of the Wiener–Hopf type for two unknown functions $J_+(\alpha)$ and $Y_-(R_{\text{cn}}, \alpha)$ in the band enclosing the real axis in the complex α -plane:

$$J_+(\alpha)G(\alpha)R_{\text{cn}} = Y_-(R_{\text{cn}}, \alpha) - \frac{\tilde{\Phi}(\alpha)}{\xi \kappa^2}, \quad (25)$$

where $\kappa = \sqrt{\alpha^2 - k^2}$. Subscripts $\pm$ stand for the functions analytical in the upper and lower half-planes, respectively. In (25),

$$G(\alpha) = K_l(\kappa R_{\text{cn}})I_l(\kappa R_{\text{cn}}) - \frac{1 - l_0 \alpha^2 / k^2}{R_{\text{cn}} \xi(\omega) \kappa^2}, \quad (26)$$

$$\tilde{\Phi}(\alpha) = \frac{k H_l^{(\mu)}(k R_{\text{cn}} \sin \theta_0)}{\alpha + k \cos \theta_0};$$

$\tilde{\Phi}(\alpha)$ is the Fourier transform of the function $\Phi(\varphi, z) \exp(-il\varphi)$,

$$\Phi(\varphi, z) = \xi(\omega) k^2 \sin^2 \theta_0 \Pi_e^{(i)}(R_{\text{cn}}, \varphi, z), \quad \xi(\omega) = -4\pi i \sigma_{zz}(\omega)/ck. \quad (27)$$

The main idea of the technique used to solve (25) consists of factorization and decomposition of known functions that enter (25) to obtain two independent expressions for functions $J_+(\alpha)$ and $Y_-(R_{\text{cn}}, \alpha)$ [62]. As a result, the general solution for the diffracted field is expressed by quadratures. The function $G(\alpha)$ can be factorized as follows [63]:

$$\ln \left[G_{\pm}(\alpha) \sqrt{\alpha \pm k} \right] = \frac{1}{2\pi i} \int_{-\infty \pm \alpha_0}^{+\infty \pm \alpha_0} \ln \left[\sqrt{\alpha'^2 - k^2} G(\alpha') \right] \frac{d\alpha'}{\alpha' \pm \alpha}, \quad (28)$$

where α_0 is a real number such that $0 < \alpha_0 < \text{Im}[k]$. The factor $\sqrt{\alpha'^2 - k^2}$ provides the asymptotic behavior of the integrand necessary for the convergence of the integral. The integral in (28) cannot be analytically evaluated; therefore, the factorization should be performed numerically. By following the standard Wiener–Hopf procedure [63], we arrive at the following formula for the z component of the field outside the tubule:

$$E_z(\rho, \varphi, z) = \frac{\eta_l(\theta_0)}{2\pi i k} \exp(il\varphi) \int_C \exp(-i\alpha z) \frac{(\alpha - k) K_l(\kappa \rho) I_l(\kappa R_{\text{cn}})}{(\alpha + k \cos \theta_0) G_+(\alpha)} d\alpha, \quad (29)$$

where

$$\eta_l(\theta_0) = \frac{H_l^{(\mu)}(k R_{\text{cn}} \sin \theta_0)}{G_+(k \cos \theta_0)(1 + \cos \theta_0)}, \quad (30)$$

and the integration path C is shown in Figure 7 with the solid line. The field inside the tubule is obtained by interchanging ρ and R_{cn} in (29). The remaining components of fields $\mathbf{E}$ and $\mathbf{H}$ can easily be obtained in a similar manner.

Equation (29) is an exact analytical expression for the field scattered by a semi-infinite nanotube, which holds true in both the far zone and near the tubule. In the near zone, convergence of integrals in (29) becomes much slower; therefore, one should be careful when performing numerical integration in this case. In the far zone, the integrals in (29) can be estimated asymptotically by the saddle-point method. The standard procedure results in

$$\begin{aligned} E_z &\sim F(\theta, \theta_0) \sin \theta \frac{\exp\{ik\sqrt{\rho^2 + z^2}\}}{k\sqrt{\rho^2 + z^2}}, \\ H_\varphi, E_\theta &\sim F(\theta, \theta_0) \frac{\exp\{ik\sqrt{\rho^2 + z^2}\}}{k\sqrt{\rho^2 + z^2}}, \end{aligned} \quad (31)$$

where

$$F(\theta, \theta_0) = \eta_l(\theta_0) \frac{J_l(k R_{\text{cn}} \sin \theta)}{G_-(k \cos \theta)(\cos \theta + \cos \theta_0)} \cot\left(\frac{\theta}{2}\right) \exp\left(-i \frac{\pi}{4}\right); \quad (32)$$

$\theta = \pi/2 + \arctan(z/\rho)$. The function $F(\theta, \theta_0)$ is conventionally referred to as the edge scattering pattern. The total scattering pattern also contains components associated with surface polaritons [21].

To illustrate these results, we calculated the far-zone scattered power density $P_l(\theta) \sim |F_l(\theta, \theta_0)|^2$. In our calculations, we assumed $l = 0$ because this term dominates for realistic incident fields. Figure 8 shows the scattered power density versus

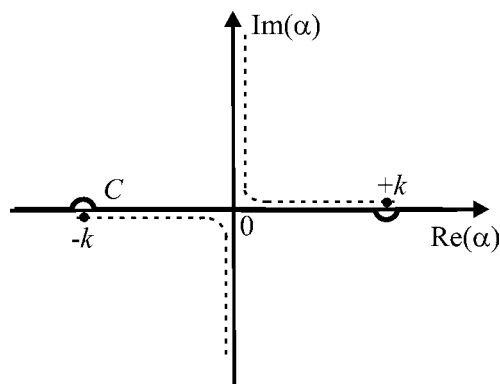


Figure 7 Contour C in the complex α -plane. The dashed lines show the branch cuts.

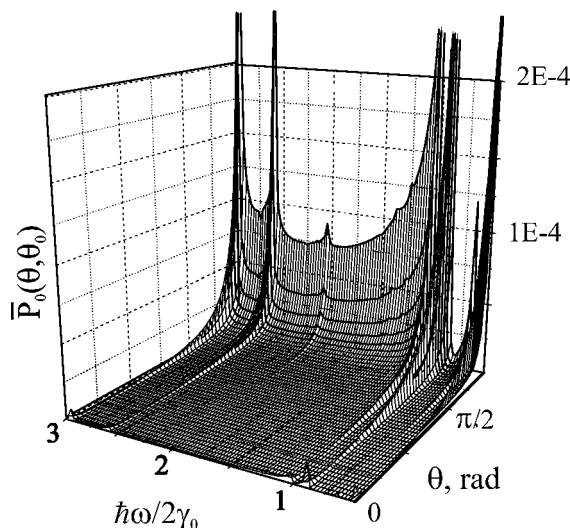


Figure 8 Density of the scattered power (normalized by the incident power density) $\bar{P}_0(\theta, \theta_0) = P_0(\theta)/[4\pi/c(1 + \cos \theta_0)^2]$, for the metallic (9, 0) nanotube at frequencies of inter-band transitions when $\theta_0 = \pi/4$.

frequency and angle for a (9, 0) metallic CN. As shown in the figure, the frequency dependence at a fixed angle θ exhibits strong oscillations: the scattered field significantly increases at frequencies that correspond to the optical transitions. It should be noted that according to our calculations, for the metallic tubules, the resonance scattering maximums are higher than those for the semiconductor tubules. The figure also shows that a relatively small detuning from the exact resonance frequencies significantly reduces intensity of the scattered field. Thus, we may conclude that, physically, the intense field scattering in CNs is related to the induction in the CN of a plasmon (which propagates from the CN edge along its axis) by the incident field. Therefore, the solution (31)–(32) shows that the edge resonances play a significant role in the scattering process. This solution is the basis for solving the problem of electromagnetic scattering by a finite CN.

4 Nonlinear processes in nanotubes

We have repeatedly stressed in the foregoing that nanostructures, and carbon nanotubes in particular, exhibit a strong spatial inhomogeneity and a great number of elementary resonances. A sufficiently strong dynamic nonlinearity is also typical of the nanostructures in a wide frequency range from microwave to ultraviolet frequencies. This nonlinearity can manifest itself in various electromagnetic processes: propagation of solitons, optical instability, dynamic chaos, or high-order harmonics generation. These effects are of interest from two points of view. First, they can be used for studying the physical properties of the nanostructures, i.e., their diagnostics. Second, these effects open new unique possibilities for controlling electromagnetic radiation, which is very promising for many optical and nanoelectronic applications.

Spatial inhomogeneity inherent to nanostructures essentially hampers description of nonlinear electromagnetic effects observed in them, because it involves nonlinear diffraction problems. Therefore, special simplifying assumptions are required in each particular case to reduce the original problem to a mathematical model that can be studied analytically or numerically. Two approaches should be mentioned that are most promising as applied to nanostructures. The first approach singles out the contribution of resonances by expanding the field in a set of specially chosen eigenmodes [31]. This approach was developed to solve nonlinear problems of macroscopic electromagnetics and was used, in particular, to analyze bistability in nonlinear diffraction lattices in the vicinity of the Wood anomalies [31]. An alternative approach—the electrodynamics of nonlinear composites—is based on the macroscopic averaging of the electromagnetic field in an ensemble of a large number of nonlinear scatterers whose size and spacings are much smaller than the wavelength [64–65].

As an example of a problem of the nonlinear electrodynamics of CNs, consider the generation within them of high-order harmonics of the incident field [24], [26–27]. Physically, the interest in generation of the high-order harmonics is caused primarily by searching for ways to create coherent far-ultraviolet and soft x-ray

sources. Gases [66]–[68] and solid surfaces [69–70] were studied as possible non-linear mediums. Generation of high-order odd harmonics in gases is caused by the tunneling of electrons from atomic orbitals to the continuous-spectrum states and back, under the effect of a strong oscillating pumping field. In solid surfaces, harmonics (both even and odd) are generated by transitions of electrons through the solid–free space interface at high (relativistic) velocities. For the pumping wave, subpicosecond pulses of a titanium-sapphire laser were used with power density $\sim 10^{14}$ W/cm². The structure of the spectrum envelope was found to be qualitatively the same in both cases: a region of relatively weak dependence of the line intensity on its number (plateau) abruptly terminates at a certain critical harmonic number. The pumping wave-to-high harmonics power conversion factor is between 10^{-6} and 10^{-7} in the plateau region. An alternative harmonic-generation mechanism was proposed in Refs. 24, 26 and 27.

4.1 Current density spectrum in an isolated nanotube

Let a CN interact with an intense laser pulse whose electric field is polarized along the CN axis. Assume that the pumping field frequency ω_1 satisfies inequalities (17), so that the contribution of interband transitions to the π -electrons motion can be neglected. This means that the motion of π -electrons is assumed to be quasi-classical. By expanding $\mathcal{E}_c(p_z, s)/\gamma_0$ (see (4) or (5) for zigzag and armchair CNs, respectively) and $F_c(p_z, s)$ (16) into a Fourier series in p_z with coefficients $\mathcal{E}_{sq}$ and F_{sq} , respectively, the surface current density can be represented by [24]

$$j_z(t) = \sum_{M=0}^{\infty} j_z^{(2M+1)}(\omega_1) \sin[(2M+1)\omega_1 t], \quad (33)$$

where

$$j_z^{(2M+1)}(\omega_1) = j_0 \sum_{s=1}^m \sum_{q=1}^{\infty} q \mathcal{E}_{sq} F_{sq} J_{2M+1}(\Lambda q), \quad (34)$$

$J_N(x)$ are the Bessel functions, and

$$F_{rs} = \frac{a}{2\pi} \int_0^{2\pi/a} F_c(p_z, s) \exp(-iar p_z) dp_z, \quad (35)$$

$$\mathcal{E}_{rs} = \frac{a}{2\pi \gamma_0} \int_0^{2\pi/a} \mathcal{E}_c(p_z, s) \exp(-iar p_z) dp_z.$$

In these expressions, $j_0 = 8e\gamma_0/\pi\hbar R_{\text{cn}}$ and $\Lambda = \Omega/\omega_1$, where Ω is the Stark frequency; $\Omega = aeE_z/\sqrt{3}$ and $\Omega = aeE_z$ for armchair and zigzag CNs, respectively.

Figure 9 plots typical spectrums of the current density for metallic and semiconductor nanotubes at various pumping field intensities. If a titanium-sapphire laser with $\lambda = 0.8 \mu\text{m}$ is used for pumping, $\Lambda = 1$ corresponds to the field $E_1 = 7 \times 10^9 \text{ V/m}$, or the intensity $I_1 = 1.3 \times 10^{13} \text{ W/cm}^2$. The most important physical feature of the spectrums shown in the figure is that these spectrums do not have a cutoff frequency typical of the spectrums studied by other authors [66]–[70]. Physically, this effect is due to the special dispersion law peculiar to π -electrons in conducting nanotubes. Figure 9 shows that a harmonic's spectrum falls much faster with the harmonic's number in semiconductor CNs than in metallic ones.

Figure 10 shows the calculation of the light generated in the spectral range 300–750 nm, around the third harmonic (TH) of the Cr:forsterite laser at 417 nm [27]. The picture demonstrates that these spectrums represent a continuous background superimposed on a narrow spectral line corresponding to the TH of the pump frequency. The TH generated by all samples of CNs in measurements carried out at Max Born Institute (Berlin, Germany) [27] is indeed emerging from a broad background, as illustrated in Figure 10 for a sample of aligned multiwall CNs. The relative intensity of this background is, however, much higher than in the theoretical prediction. The formation of a local plasma through the emission of free electrons from the CNs could be the reason for this discrepancy.

TH generation with nonresonant excitation in bulk crystals or gases can be described, in general, by the third-order polarization $P^{(3)}(3\omega_1) = \chi^{(3)}(3\omega_1)E^3(\omega_1)$, even for rather high intensities below the optical damage threshold. One would also expect a similar behavior for the TH yield of CNs. Figure 11(a) shows the theoretical dependence of the TH yield on the driving field intensity, for various types of CNs. Surprisingly, for pump intensities as low as 10^{10} W/cm^2 , the expected power

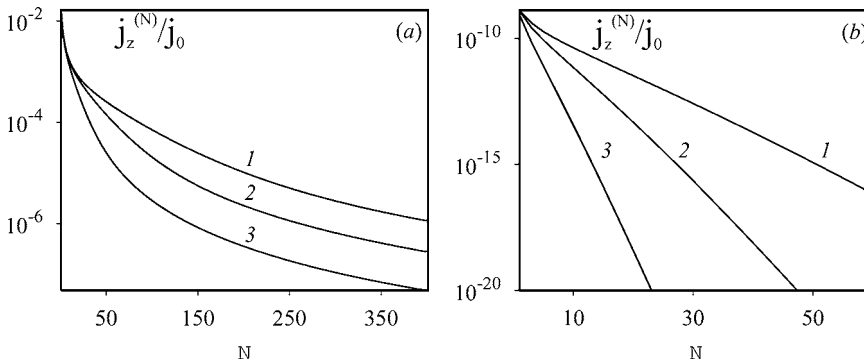


Figure 9 Envelope of the spectrum of high-order harmonics of the nonlinear current induced in (a) metallic (12, 0) and (b) semiconductor (11, 0) zigzag nanotubes by pumping pulses of different intensities: (1) $\Lambda = 1.0$, (2) 0.5, and (3) 0.2. The normalization factor j_0 for the metallic and semiconductor nanotubes is $2.6 \times 10^6 \text{ A/m}$ and $2.8 \times 10^6 \text{ A/m}$, respectively. $N = 2M + 1$ is the harmonic's number, $M = 0, 1, 2, \dots$

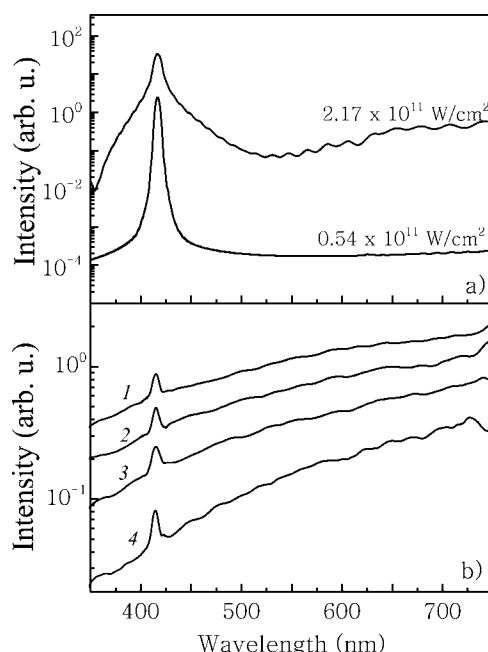


Figure 10 Broad background and TH signal generated by the interaction of intense laser radiation with aligned multiwall CNTs: (a) theory, (b) experiment; input intensities: 1: 2.3×10^{11} W/cm², 2: 1.7×10^{11} W/cm², 3: 1.3×10^{11} W/cm², and 4: 0.8×10^{11} W/cm². The experimental curves are corrected for the efficiencies of the monochromator, the photomultiplier and the transmission of the KG5 filters in the detection system.

law for the intensity dependence is broken:

$$j_z^{(3)}(\omega_1) \sim E_z^p(\omega_1), \quad (36)$$

with p between 2.04 and 2.58 for the considered types of CNTs. The theory also predicts that p depends not only on the type of CNTs and their diameter, but also on the pumping frequency. The experimental dependencies measured from samples of nonaligned multiwall CNTs in Figure 11(b) show good agreement with this theoretical prediction. Physically, this fact indicates that the interaction of CNTs with an intense laser pulse cannot be described by the perturbation approach, even for relatively low laser intensities. Note that the theoretical and experimental values of p agree remarkably well, except for the case of the orthogonally aligned array of CNTs. One interesting experimental result [see Figure 11(b)] is the observed decrease in slope of the TH intensity at a pump laser pulse intensity of $\sim 3 \times 10^{10}$ W/cm². Such saturation of the TH signal is also predicted by the theory, however at incident intensities about two orders of magnitude higher.

A similar situation, where the power expansion of polarization does not work, occurs for the fifth harmonic of the current j_z at intensities $\sim 10^{10}$ – 10^{11} W/cm². Our analysis shows that both theoretical and experimental values of the exponent p

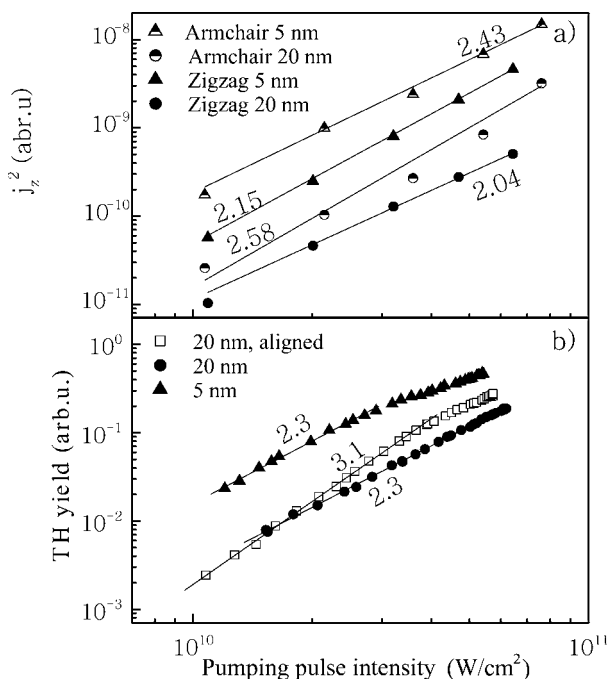


Figure 11 Dependence of the TH-generation efficiency on the intensity of the pumping pulse: (a) theory, (b) experiment.

for the fifth harmonic differ from 5 and, at the same time, are close to each other (4.0 and 4.26 for experiment and theory, respectively). In conclusion, the interaction of strong laser fields with samples of CNs cannot be described by a power expansion of the polarization. This results in the violation of the general expressions $j_z^{(3)}(\omega_1) \sim E_z^3(\omega_1)$ and $j_z^{(5)}(\omega_1) \sim E_z^5(\omega_1)$ for the dependences of the third and fifth harmonic yields on the input laser field, even for intensities as low as 10^{10} – 10^{11} W/cm². The results from a full-quantum theoretical model show good agreement with the experimental findings.

In this section, we have studied high-order harmonics of the current in a single nanotube. The next step in analyzing this problem is the study of high-order field harmonics in arrays of aligned CNs with allowance for dispersion. Such an array is effectively an anisotropic birefringent medium [18]. A consistent analysis of generation of high-order harmonics in nanotube arrays with allowance for the phase matching is given elsewhere [26].

4.2 Negative differential conductivity in an isolated nanotube

In the quasi-static regime, nonlinear properties of the charge carriers in CNs also exhibit themselves as portions with the negative differential conductivity (NDC), $dI/dV < 0$, in the current–voltage (I–V) characteristics [14], [25], [71]. In a CN interacting simultaneously with dc and ac fields in the vicinity of a given working

point in the I–V characteristics, instability evolves. This makes CNs attractive as potential nanoscale amplifying diodes similar to the macroscopic tunneling ones.

The I–V characteristics for tunneling electrons in individual single-wall CNs at low temperatures were measured in Refs. 48 and 72. At such temperatures $k_B T \ll \epsilon_c$, and $k_B T \ll \Delta\epsilon$ and conduction occurs through well-separated discrete electron states; here, ϵ_c is the charging energy, and $\Delta\epsilon = \pi \hbar v_F / l_{\text{cn}}$ is the energy level spacing. It is reported that $\Delta\epsilon \simeq 0.6$ meV for CNs with $l_{\text{cn}} \simeq 3 \mu\text{m}$ [48]; and an estimate $\epsilon_c \simeq 1.4e^2 \ln(l_{\text{cn}}/R_{\text{cn}})/l_{\text{cn}}$ is given in Ref. 73. For CNs described in Ref. 48, $R_{\text{cn}} \simeq 0.7$ nm and, thus, $\epsilon_c \simeq 2.5$ meV, which is in good agreement with experimental data [48]. Under these conditions, the current is produced by electrons tunneling through the CN in the presence of the Coulomb blockade induced by the long-range (unscreened) Coulomb interaction. Due to this mechanism, the I–V characteristics that were observed [48], [72] are analogous to that obtained via the scanning tunneling microscopy. As a result, the normalized differential conductivity $(V/I)(dI/dV)$ proves to be proportional to the local density of states. Therefore, the I–V characteristics described in Refs. 48 and 72 carry important information on the nanotube electron structure. On the other hand, tunneling in macromolecules (in nanotubes, in particular) can serve as a basis for monomolecular transistors [13].

In this section, we theoretically analyze I–V characteristics of the nanotubes at room temperatures when $k_B T > \epsilon_c$ and $k_B T \ll \Delta\epsilon$. Consider a single-wall zigzag nanotube exposed to a homogeneous axial dc field $E_z = V/l_{\text{cn}}$, where V is the voltage across the nanotube. We apply the semiclassical approximation, considering the motion of π -electrons as a classical motion of free quasi-particles in the field of the crystalline lattice with the dispersion law (4) extracted from quantum theory. The motion of quasi-particles in an external axial dc electric field is described by the Boltzmann kinetic equation, where $\partial/\partial t = \partial/\partial z = 0$ is assumed and the collision integral is taken in the relaxation-time approximation. Depending on the relaxation time, the relaxation term can describe electron–phonon scattering, electron–electron collisions, etc.

The surface current density is determined by (13). Expansions of $F_c(p_z, s)$ and $\mathcal{E}_c(p_z, s)/\gamma_0$ into Fourier series in p_z , carried out by analogy with the previous section, lead us to the equation [25]

$$j_z(E_z) = \frac{j_0}{2} \sum_{r=1}^{\infty} \frac{r^2 \Omega \tau}{1 + (r \Omega \tau)^2} \sum_{s=1}^m F_{rs} \mathcal{E}_{rs}, \quad (37)$$

with F_{rs} and $\mathcal{E}_{rs}$ defined by (35). Equation (37) states the basis for evaluation of the I–V characteristics of nanotubes.

Direct numerical integration in (35) for the coefficients F_{rs} and $\mathcal{E}_{rs}$ is technically difficult because the integrands are rapidly oscillating functions. Therefore, we resorted to the following technique. The change of variable $z = \exp(iap_z)$ was used to transform the original integrals into integrals over the closed path $|z| = 1$ in the complex plane. The integrands have two pairs of branch points in the z plane.

The integrand for F_{rs} also has an infinite number of first-order poles inside the unit circle. According to the Cauchy residue theorem, the integrals can be written in terms of integrals over banks of the branch cuts plus series of residues (for F_{rs}). The integrals over the cut banks do not contain oscillating functions and can easily be calculated numerically. The residue series converge rapidly and are also summed numerically.

Let us estimate constraints that follow from this theoretical model. As has been stated previously, the model describes motion of the quasi-particles by the classical Boltzmann kinetic equation. Thus, both interband transitions and quantum-mechanical corrections to the intraband motion are not accounted for in this model. The first of these approximations is valid when the inequality $\Omega \leq \omega_l$ holds true; the second assumption requires that $\Omega \leq \delta\mathcal{E}$. In these inequalities, ω_l is the low-frequency edge of the optical transition band given by (17) and $\delta\mathcal{E}$ is the width of the allowed band. The estimate $\delta\mathcal{E} \sim \gamma_0$ and the inequality (17) reduce both constraints imposed on the Stark frequency to the limitation on the intensity of the external electric field $|E_z| < \gamma_0/2eR_{\text{cn}}$. Our theoretical model also neglects the Coulomb interaction between the electrons. The role of this interaction in the carbon nanotubes has been addressed elsewhere [73]–[75]. It has been found that the short-range electron–electron interaction typical for CN arrays does not significantly contribute at high temperatures. Since the Coulomb interaction in an isolated CN is unscreened, it exhibits itself in a different manner, providing an observable effect over a wide temperature range. Therefore, the results obtained from our model are applicable primarily to CN arrays. For a single tubule, this model should be modified to allow for the long-range Coulomb interaction. As follows from Ref. 74, the change in the temperature dependence of the relaxation time τ is expected as the only result of the Coulomb interaction.

Figure 12 shows the I–V characteristic for undoped (with zero chemical potential) metallic zigzag nanotubes. As shown in the figure, the function j_z versus E_z is linear at small intensities of the imposed electric field, which corresponds to the ohmic conductivity. With the increase of the imposed electric field intensity, $\partial j_z / \partial E_z$ becomes smaller until the current density reaches its maximum value j_z^{max} at $E_z = E_z^{\text{max}}$. Increasing the intensity of the applied electric field will further decrease j_z . We thus predict the negative differential conductivity (NDC) $\partial j_z / \partial E_z < 0$.

The imposed field strength $E_z^{\text{max}} \approx 3.2 \times 10^3$ V/cm at which the NDC starts exhibiting itself was found to be unexpectedly weak. Indeed, nonlinearity in these structures is determined by the quantity aE_z . In quantum superlattices the spacing is about 10^{-6} cm [10], which is much greater than the characteristic interatomic distance b in the graphene lattice. Nevertheless, the NDC is observed in them almost at the same strength of the imposed field. Therefore, nonlinearity in nanotubes is much stronger.

To explain this phenomenon, let us compare the nonlinear conductivity mechanisms in nanotubes and in superlattices. In quantum superlattices, the dispersion law is $\mathcal{E}_{v,c}(p_z) = \pm \gamma'_0 [1 - \cos(2ap_z/3)]$, where γ'_0 is the overlapping inte-

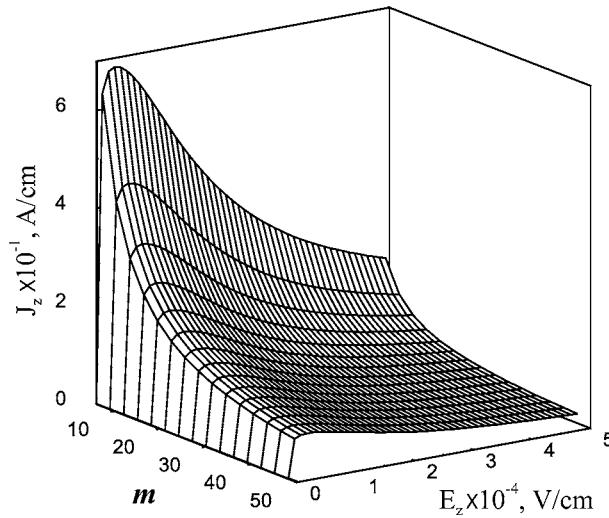


Figure 12 I–V characteristics of metallic zigzag nanotubes at $T = 287.5$ K and $\tau = 3 \times 10^{-12}$ s.

gral. Applying the method described previously to this dispersion law, we obtain the expression $j_z(E_z) = \sigma_{zz}E_z/(1 + i\tau r\Omega)$ instead of (37), where $\sigma_{zz} = \lim_{E_z \rightarrow 0}(\partial j_z/\partial E_z)$ is the linear conductivity. The comparison of these two expressions for the current density shows that a specific feature of nanotubes is that they produce high-order Stark harmonics. Our calculations show that the number of significant Stark harmonics is within 70–150 for the metallic nanotubes and within 200–300 for the semiconductor nanotubes. As a result, the high-order Stark components play a significant role in nanotubes, and the integral nonlinearity in nanotubes is much stronger than in superlattices. Impurities and defects in the lattice provide an additional mechanism of carrier scattering, which can be described quantitatively by the substitution $\tau \rightarrow \tau' = \tau\tau_1(\tau + \tau_1)^{-1}$, where the relaxation time τ_1 is determined by the impurities and defects. Since $\tau' < \tau$, doping increases $E_z^{\max}$ and decreases $\partial j_z/\partial E_z$ in the NDC region.

Thus, we have predicted the NDC effect in CNs, which is expected to be observable in sufficiently long CNs at room temperatures. As was emphasized before, the NDC causes the current instability. One can expect that simultaneously applied dc and ac fields will result in the dynamic electron localization (which is the nonlinear phase of the instability) and in the two-dimensional analog of the self-induced transparency. The effects mentioned are responsible for the absolute negative conductivity, which is thus predicted in nanotubes. Due to this phenomenon, regions must appear where nanotubes exhibit absolute negative conductivity and active properties, which creates the possibility of developing microwave and infrared oscillator nanodiodes in single nanotubes as well as in nanotube arrays.

The predicted NDC mechanism is not alone in creating the NDC effect. Another mechanism observed in inhomogeneous nanotubes is caused by π -electrons tunnelling through the potential barrier near the inhomogeneity [14], [71].

5 Quantum electrodynamics of carbon nanotubes

5.1 The Maxwell equations for electromagnetic field operators

In most cases, electromagnetic modeling of nanostructures assumes the number of photons involved in the process to be large enough to describe the electromagnetic field by classical equations. At the same time, peculiarities of traditional quantum electrodynamics (QED) effects such as spontaneous emission and electromagnetic fluctuations, as well as recently raised ideas to use nanostructures for storage and processing of quantum information, provide growing interest for QED of nanostructures and, in particular, nanotubes. For this aim, the quantum nature of the electromagnetic field in nanotubes should be taken into account. Since the nano-object (the CN) is an inhomogeneity much smaller than the photon wavelength, this issue appears to be significantly more complex than QED problems in homogeneous mediums. In this section, we concentrate on the problem of spontaneous emission of an atom located inside or in the vicinity of a CN.

Standard schemes of the electromagnetic field quantization are based on modal representations: in free space, these modes are plane waves; in cavities they are corresponding eigenmodes, for example. The quantum description of the electromagnetic field replaces coefficients of such modal representations by operators of creation and annihilation of photons associated with a particular mode.

Since the nanostructures are strongly inhomogeneous open systems, it is usually difficult to find an appropriate system of eigenmodes. Therefore, an alternative approach developed recently for lossy dispersive mediums [76]-[78] appears to be more convenient for the QED of nanostructures. This approach rejects the modal representation and allows for quantization in the Maxwell equations: the vectors $\mathbf{E}$ and $\mathbf{H}$ are replaced by corresponding operators that satisfy the appropriate commutation relations and define observable quantities as mean values of these operators. Following Refs. 76-78, let us introduce the electric field operator as $\hat{\mathbf{E}}(\mathbf{r}) = \hat{\mathbf{E}}^{(+)}(\mathbf{r}) + \hat{\mathbf{E}}^{(-)}(\mathbf{r})$, where

$$\hat{\mathbf{E}}^{(+)}(\mathbf{r}) = \int_0^\infty \hat{\underline{\mathbf{E}}}(\mathbf{r}, \omega) d\omega, \quad \hat{\mathbf{E}}^{(-)}(\mathbf{r}) = [\hat{\mathbf{E}}^{(+)}(\mathbf{r})]^* \quad (38)$$

and $*$ indicates the Hermitian conjugate. The magnetic field operator $\hat{\mathbf{H}}(\mathbf{r})$ is defined in the same manner. Operators $\hat{\underline{\mathbf{E}}}$ and $\hat{\underline{\mathbf{H}}}$ are subject to radiation conditions at infinity. They satisfy the Maxwell equations

$$\nabla \times \hat{\underline{\mathbf{E}}} = ik \hat{\underline{\mathbf{H}}}, \quad \nabla \times \hat{\underline{\mathbf{H}}} = -ik \hat{\underline{\mathbf{E}}} + \frac{4\pi}{c} \hat{\mathbf{j}}^{ext}, \quad (39)$$

where $\hat{\mathbf{J}}^{ext}$ is the exterior current operator. The effective boundary conditions (18) and (19) for (39) are rewritten as

$$\begin{aligned} \mathbf{n} \times \left(\hat{\underline{\mathbf{E}}} \Big|_{\rho=R_{cn}+0} - \hat{\underline{\mathbf{E}}} \Big|_{\rho=R_{cn}-0} \right) &= \mathbf{0}, \\ \mathbf{n} \times \left(\hat{\underline{\mathbf{H}}} \Big|_{\rho=R_{cn}+0} - \hat{\underline{\mathbf{H}}} \Big|_{\rho=R_{cn}-0} \right) &+ \frac{4\pi}{c} \hat{J}_z^{ns} \mathbf{e}_z = \frac{4\pi}{c} \sigma_{zz}(\omega) \hat{E}_z \mathbf{e}_z, \end{aligned} \quad (40)$$

where $\mathbf{n}$ is the unit vector along the external normal to the CN surface, $\sigma_{zz}(\omega)$ is the axial dynamical conductivity of CN, see (15), and $\hat{J}_z^{ns}$ is the operator of an axial noise current. The latter is expressed as $\hat{J}_z^{ns} = \{\hbar\omega \text{Re}[\sigma_{zz}(\omega)]/\pi\}^{-1} \hat{f}(\mathbf{R}, \omega)$ in terms of two-dimensional scalar field operator $\hat{f}(\mathbf{R}, \omega)$ (where $\mathbf{R}$ lies on the CN surface) satisfying standard bosonic commutation relations:

$$\begin{aligned} [\hat{f}(\mathbf{R}, \omega), \hat{f}^*(\mathbf{R}', \omega')] &= \delta(\mathbf{R} - \mathbf{R}') \delta(\omega - \omega'), \\ [\hat{f}(\mathbf{R}, \omega), \hat{f}(\mathbf{R}', \omega')] &= [\hat{f}^*(\mathbf{R}, \omega), \hat{f}^*(\mathbf{R}', \omega')] = 0, \end{aligned} \quad (41)$$

where $\delta(x)$ is the Dirac delta function and $[\hat{f}_1, \hat{f}_2] = \hat{f}_1 \hat{f}_2 - \hat{f}_2 \hat{f}_1$. The axial noise current is responsible for the correct commutation relations of the operators $\hat{\underline{\mathbf{E}}}$ and $\hat{\underline{\mathbf{H}}}$ [79]. The homogeneous Maxwell equations (39) along with boundary conditions (40) describe the quantum electrodynamics of CNs.

One of the most important applications of this quantization scheme is the problem of dynamics of an excited two-level atom located inside (or near) a dielectric body with relative permittivity $\epsilon(\mathbf{r}, \omega)$ [80]–[82]. This problem considers an electric dipole transition, characterized by the dipole moment $\boldsymbol{\mu}$ and frequency ω_A , in an electrically neutral atom located at position $\mathbf{r} = \mathbf{r}_A$. The general expression for the spontaneous radiation time τ_{sp} for this system is as follows:

$$\Gamma = \frac{1}{\tau_{sp}} = \frac{8\pi}{\hbar} k_A^2 \mu_\alpha \mu_\beta \text{Im}[G_{\alpha\beta}(\mathbf{r}_A, \mathbf{r}_A, \omega_A)], \quad (42)$$

where $k_A = \omega_A/c$, and $G_{\alpha\beta}$ are the components of the classical dyadic Green function that accounts for the dielectric body. Summation is implied with respect to indexes α and β .

Equation (42) can be interpreted physically as follows. Spontaneous emission is the process of interaction between an excited atom and vacuum states of the electromagnetic field. The vacuum states are diffracted by the dielectric body similar to the diffraction of conventional electromagnetic fields. This effect is taken into account by the difference between the dyadic $G_{\alpha\beta}$ and the free-space Green dyadic $G_{\alpha\beta}^{(0)}$. Since $\text{Im}[G_{\alpha\beta}^{(0)}(\mathbf{r}_A, \mathbf{r}_A, \omega_A)] = \omega_A \delta_{\alpha\beta} / 6\pi c$, where $\delta_{\alpha\beta}$ is the Kroneker delta, it can easily be seen that as $\epsilon(\mathbf{r}, \omega) \rightarrow 1$, and (42) simplifies to the well-known formula for the spontaneous decay time in free space [83]: $\tau_{sp}^{(0)} = 3\hbar/4k_A^3 \mu^2$. We

should also note that the Lamb shift in the transition frequency due to the presence of the dielectric body, which differs from that in free space, has been reported [80-81].

Equation (42) was used to analyze various physical situations; for example, to calculate the spontaneous decay time of an excited atom in a spherical microcavity [80-81] and to study the influence of the local field effects on the spontaneous emission in optically dense gases and solid dielectrics [82]. Next, we show how the foregoing quantization scheme can be used to study the spontaneous emission process in carbon nanotubes.

5.2 Spontaneous decay of an excited atom in the carbon nanotube

Consider the spontaneous decay of an excited atom located inside a CN at a distance ρ_0 from its axis [28]. The dipole moment of the atom is assumed to be aligned with the z axis. We have demonstrated previously that the EBC method is an appropriate tool for many problems of nanotube electrodynamics. Application of this method to the problem of the spontaneous decay of an atom inside a nanotube yields (42) [80-81].

Since the dipole moment of the atom is parallel with the CN axis, only the longitudinal component of the dyadic Green function is of physical interest. Let us represent this component in terms of the scalar Green function $\bar{G}$ of the atom in the CN:

$$G_{zz} = \frac{1}{k^2} \left(\frac{\partial^2 \bar{G}}{\partial z^2} + k^2 \bar{G} \right). \quad (43)$$

In turn, this function can be represented as

$$\bar{G} = \begin{cases} \tilde{G}^+, & \rho > R_{\text{cn}} \\ G_0 + \tilde{G}^-, & \rho < R_{\text{cn}} \end{cases}, \quad (44)$$

where $G_0 = \exp(ik\rho)/4\pi\rho$ is the free-space Green function and $\tilde{G}^\pm$ are unknown functions that satisfy the homogeneous Helmholtz equation and boundary conditions on the CN surface, which follow from EBCs (18) and (19) as

$$\begin{aligned} \tilde{G}^+ \Big|_{\rho=R_{\text{cn}}} &= (G_0 + \tilde{G}^-) \Big|_{\rho=R_{\text{cn}}}, \\ \left[\Delta \frac{\partial}{\partial \rho} (\tilde{G}^+ - \tilde{G}^-) - \xi(\omega) \left(\frac{\partial^2 \bar{G}}{\partial z^2} + k^2 \bar{G} \right) \tilde{G}^+ \right] \Big|_{\rho=R_{\text{cn}}} &= \Delta \frac{\partial G_0}{\partial \rho} \Big|_{\rho=R_{\text{cn}}}. \end{aligned} \quad (45)$$

In these expressions, $\Delta = 1 + \gamma(\omega)\partial^2/\partial z^2$, and the function $\sigma_{zz}(\omega)$ is given by (15).

Let us seek functions $\tilde{G}^\pm$ as expansions in terms of cylindrical functions. Using (42), we arrive at the following expression for the spontaneous decay rate of an atom in a CN: $\Gamma = \zeta_0 \Gamma_0$, where $\Gamma_0 = 1/\tau_{sp}^{(0)}$,

$$\zeta_0 = 1 + \frac{3\pi R_{cn}}{16k_A^3} \sum_{p=-\infty}^{\infty} \text{Im} \int_C \frac{\beta_A \kappa_A^4 I_p^2(\kappa_A \rho_0) K_p^2(\kappa_A R_{cn})}{1 - \beta_A R_{cn} \kappa_A^2 I_p(\kappa_A \rho_0) K_p(\kappa_A R_{cn})} dh, \quad (46)$$

$\kappa_A = \sqrt{h^2 - k_A^2}$, and $\beta_A = -\xi(\omega_A)/[1 - \gamma(\omega_A)h^2]$. The integration path C in the complex plane is shown in Figure 7. The coefficient ζ_0 directly characterizes the effect of diffraction of the vacuum states on the spontaneous decay rate of an atom in the nanotube. It should be emphasized that the integral in (46) cannot be reduced to an integral with finite limits as was done elsewhere [84] for a perfectly conducting cylinder. This circumstance is due to the contribution to the atom's spontaneous decay of surface waves propagating in the CN. By analogy with the classical diffraction theory, we can expect that this contribution to be significant.¹

For the inner region ($r_A < R_{cn}$), (46) is modified by the simple interchange of r_A and R_{cn} in the numerator of the integrand. Note the divergence of the integral in (46) at $r_A = R_{cn}$, i.e., when the atom is located directly on the CN surface. This divergence originates from the averaging procedure over a physically infinitely small volume when describing CN optical properties. Such an averaging does not assume any additional atoms on the CN surface; to take them into consideration the procedure must be modified. Thus, the domain of applicability of our model is restricted by the condition $|r_A - R_{cn}| > b$.

The decay of the excited atom interacting with a medium may proceed both via real photon emission (radiative decay) and via virtual photon emission with subsequent excitation in the medium of quasi-particles (nonradiative decay). Both of these decay channels are present in the atomic spontaneous decay rate Γ described by (46).

The partition of the total Γ into radiative and nonradiative contributions is not a trivial problem. For an atom near a microsphere, the radiative contribution Γ_r has been estimated by using the Poynting vector [80-81]. The radiative contribution has also been estimated for an atom inside an optical fiber [85]. Following this approach, let us estimate the distribution of the spontaneous emission intensity $I(\mathbf{r}, t)$ at large distances $|\mathbf{r}| \rightarrow \infty$. In the spherical coordinates, $(|\mathbf{r}|, \phi, \theta)$ associated with the atom, we obtain

$$I(\mathbf{r}, t) \simeq \frac{1}{|\mathbf{r}|^2} k_A^4 |\mu|^2 \sin^2 \theta \left| \sum_{p=-\infty}^{\infty} \Xi_p(-ik_A \sin \theta) e^{ip\phi} \right|^2 \exp(-\Gamma t), \quad (47)$$

¹We refer the reader to Ref. 20, where the model of the perfectly conducting cylinder is shown to fail in describing correctly the CN electromagnetic properties.

with

$$\Xi_p(x) = \begin{cases} \frac{I_p(xr_A)}{1 + R_{cn}\beta_A x^2 I_p(xR_{cn})K_p(xR_{cn})}, & r_A < R_{cn} \\ I_p(xr_A) - \frac{R_{cn}\beta_A x^2 I_p^2(xR_{cn})K_p(xr_A)}{1 + R_{cn}\beta_A x^2 I_p(xR_{cn})K_p(xR_{cn})}, & r_A > R_{cn} \end{cases}. \quad (48)$$

Then the contribution of the radiative channel is given by

$$\begin{aligned} \frac{\Gamma_r}{\Gamma} &= \frac{c}{2\pi\hbar\omega_A} \lim_{|\mathbf{r}| \rightarrow \infty} \int_0^\infty dt \int_0^{2\pi} d\phi \int_0^\pi |\mathbf{r}|^2 I(\mathbf{r}, t) \sin\theta d\theta \\ &= \frac{3}{4\zeta(\omega_A)} \sum_{p=-\infty}^\infty \int_0^\pi |\Xi_p(-ik_A \sin\theta)|^2 \sin^3\theta d\theta. \end{aligned} \quad (49)$$

Figure 13 shows the results of the calculations of $\zeta(\omega_A)$ according to (46) for metallic and semiconducting zigzag CNs. The atom is located on the CN axis. The frequency range $0.305 < \hbar\omega_A/2\gamma_0 < 0.574$ corresponds to visible light. Lower frequencies $\hbar\omega_A/2\gamma_0 < 0.305$ correspond to infrared waves emitted by highly excited Rydberg atomic states. At these frequencies, a large difference (three to four orders of magnitude) is seen in the value of $\zeta(\omega_A)$ for metallic and semiconducting CNs. The difference is caused by the Drude-type conductivity (intraband electronic transitions) dominating in this region, whose relative contribution to the total CN conductivity is larger in metallic CNs than in semiconducting ones [19-20], [86].

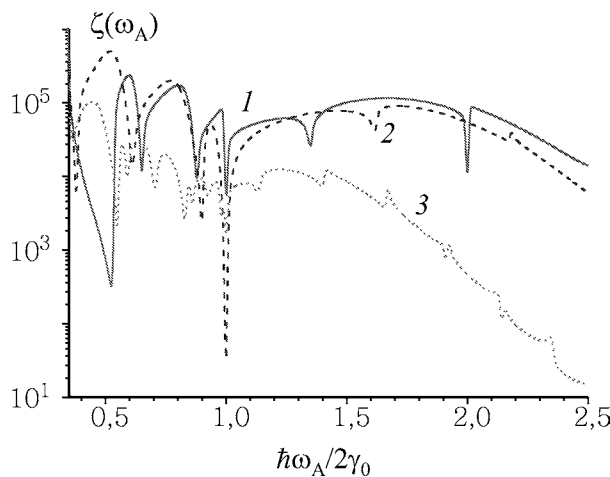


Figure 13 Factor $\zeta(\omega_A)$ calculated from (46) for an atom in the center of different zigzag CNs: 1: (9, 0), 2: (10, 0), 3: (23, 0). Surface axial conductivity σ_{zz} appearing in (46) was calculated in the τ -approximation with $\tau = 3 \times 10^{-12}$ s [19].

As the frequency increases, interband transitions start manifesting themselves and $\zeta(\omega_A)$ becomes irregular. At high frequencies, there is no significant difference between metallic and semiconducting CNs of approximately equal radii. The function $\zeta(\omega_A)$ has dips when ω_A equals the interband transition frequencies; in particular, there is a dip at $\hbar\omega_A = 2\gamma_0$ for all CNs considered. It is essential that $\zeta(\omega_A) \gg 1$ throughout the entire frequency range considered. This lets us formulate the central result of the present analysis: *the spontaneous decay probability of an atom in the vicinity of a CN is larger by a few orders of magnitude than that of the same atom in free space*. In other words, the Purcell effect [87] is extraordinarily strong in CNs. This is physically explained by the photon vacuum renormalization: the density of photonic states (and, as a consequence, the atomic decay rate) near a CN effectively increases as per

$$\rho^{\text{eff}}(\omega) = \zeta(\omega) \frac{\omega^2}{\pi c^3}, \quad (50)$$

since, along with ordinary free photons, photonic states coupled with CN electronic quasi-particle excitations appear. The presence of a CN is seen to drastically accelerate the spontaneous decay process of an excited atomic state.

The possible existence of slow surface electromagnetic waves in CNs has been demonstrated [19-20]. Such waves are responsible for the strong Purcell effect for an atom in the spherical microcavity [80]. The results of the present analysis agree qualitatively with those obtained in Refs. 80 and 81. However, there is the risk of going beyond the applicability limits of the two-level model and Markovian approximation. Indeed, considering the spontaneous radiation of the atom in the near-surface regime, one gets for the tangential atomic dipole orientation [80-81]

$$\frac{\Gamma}{\Gamma_0} \approx \frac{3\epsilon''(\omega_A)}{8|\epsilon(\omega_A) + 1|^2} \frac{1}{(k_A \Delta r)^3} + O(\Delta r^{-1}), \quad (51)$$

where Δr is the atom-surface distance, $\epsilon(\omega)$ the permittivity of the under-surface medium, and $\epsilon''(\omega) = \text{Im}[\epsilon(\omega)]$. Seemingly, approaching the surface, one obtains arbitrary large Γ . However, in doing so one has to remain within the applicability domain of the macroscopic approximation.

The relation (51) has been derived under the condition that Δr is much smaller than all other parameters, or, more physically, when the atom is placed so close to the surface that it sees the quasi-plane, and the surface curvature is irrelevant. For CNs of small enough radii ($m \approx 10-30$), this condition contradicts the inequality $|r_A - R_{\text{cn}}| > b$ determining the applicability limits of the macroscopic approximation for this particular task. As a consequence, the CN surface curvature turns out to be essential and (46) cannot be, in principle, reduced to any equation similar to (51). Thus, the large Purcell effect in CNs has nothing to do with the near-surface regime.

Figure 14 shows $\zeta(\omega_A)$ for an atom located outside a CN at different distances from the CN surface. The qualitative behavior of $\zeta(\omega_A)$ is similar to that represented in Figure 13 for an atom inside a CN. It is seen that $\zeta(\omega_A)$ rapidly decreases

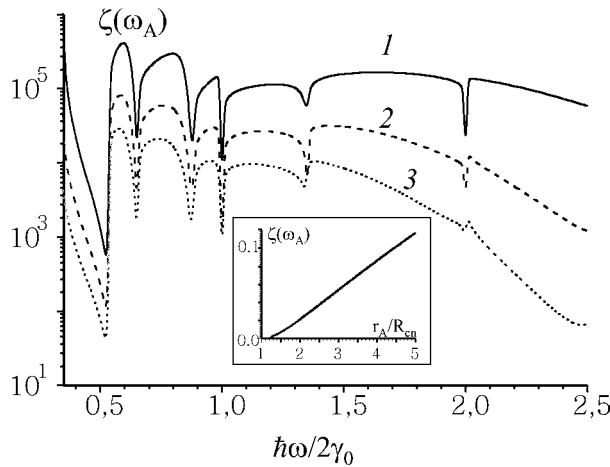


Figure 14 The function $\zeta(\omega_A)$ for an atom at different distances outside a zigzag (9, 0) CN. 1: $r_A = 1.5 R_{\text{CN}}$, 2: $2.0 R_{\text{CN}}$, 3: $2.5 R_{\text{CN}}$. Inset: $\xi(\omega_A)$ at $\omega_A = 3\gamma_0/\hbar$ as a function of r_A/R_{CN} for an atom near a (9, 0) CN in the model of the ideally conducting cylinder.

with increasing distance—as it should be—in view of the evident fact that photonic states coupled with CN electronic excitations (those increasing effective density $\rho^{\text{eff}}(\omega)$) are spatially localized on the CN surface, and their coupling strength with the excited atom decreases with increasing distance of the atom from CN. Figure 15 shows the ratio Γ_r/Γ calculated according to (49) for an atom in the center of different CNs.

Note that $\Gamma_r/\Gamma = W_s(\omega_A)/\hbar\omega_A$, with $W_s(\omega_A)$ being the total power of the atomic spontaneous radiation far from CN. The ratio is very small, indicating that the nonradiative decay dominates. However, the radiative decay is seen to essentially contribute in the vicinity of the interband transition frequencies. Therefore, the frequency dependence of $W_s(\omega_A)$ (which, in principle, can be measured experimentally) reproduces CN electronic structure peculiarities. The main conclusion one can draw from Figure 15 is that the Purcell effect in CNs, along with the increase of the atomic spontaneous decay rate, manifests itself by decreasing the power of the spontaneous radiation.

Our model of the atomic spontaneous decay in the presence of a CN includes, as a limiting case, the model of the ideally conducting cylinder considered elsewhere [84]. In particular, (46) reduces for the outward atomic position to (15) and (18) of Ref. 84 as $\sigma_{zz} \rightarrow \infty$. The inset in Figure 14 shows $\zeta(\omega_A)$ at $\omega_A = 3\gamma_0/\hbar$ ($k_A R_{\text{CN}} \simeq 0.01$) as a function of r_A/R_{CN} for this case. The dependence is similar to that reported in Ref. 84 for a z -oriented dipole at $k_A R_{\text{CN}} = 1$. For the atom inside CN, (46) yields $\zeta(\omega_A) \rightarrow 0$ as $\sigma_{zz} \rightarrow \infty$. The result is natural since, in this case, only one electromagnetic eigenmode can propagate in the CN; this mode is essentially transverse and, consequently, is not coupled with the axially oriented atomic dipole moment. However, the actual $\zeta(\omega_A)$ behavior discussed above is quite different from that predicted by the ideally conducting cylinder model, since the latter does not account for CN electronic quasi-particle excitations respon-

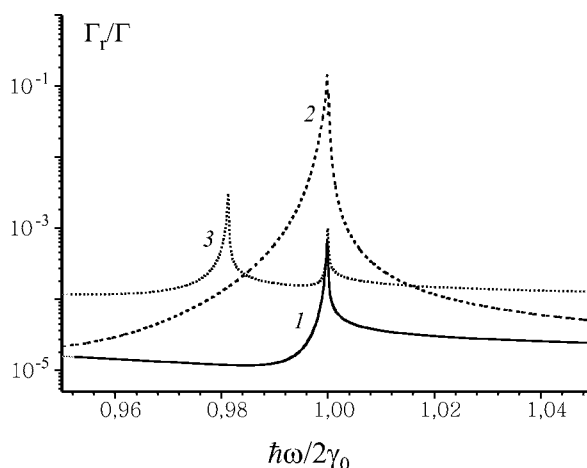


Figure 15 Ratio Γ_r/Γ calculated from (49) for an atom in the center of different zigzag CNs; 1: (9, 0), 2: (10, 0), and 3: (23, 0).

sible for the nonradiative atomic decay dominating the total spontaneous decay process.

Our theory may be generalized to cover the transverse atomic electric dipole orientation, electric quadrupole and magnetic dipole atomic transitions, the properties of organic molecules inside/outside CNs [88]. The mechanism that was revealed of the photon vacuum renormalization is likely to manifest itself in other phenomena in CNs such as Casimir forces or electromagnetic fluctuations.

The presented results may be tested by methods of atomic fluorescent spectroscopy and may have various physical consequences. In particular, the effect of the drastic increase of the atomic spontaneous decay rate may turn out to be of practical importance in problems of the laser control of atomic motion [89], increasing the ponderomotive force acting on an atom moving in the vicinity of a CN in a laser field. One might expect the Purcell effect peculiarities that we predict for CNs to manifest themselves in macroscopic anisotropically conducting waveguides with strong wave deceleration (for example, in microwave spiral or collar waveguides with highly excited Rydberg atoms inside).

6 Conclusion

In this chapter, we have analyzed a wide range of linear and nonlinear electromagnetic problems and associated issues of electron transport through carbon nanotubes. QED as applied to nanotubes has also been formulated and used for consideration of the problem of atomic spontaneous emission near a CN. The choice of problems and methodology presented here were dictated by the following reasoning: First, we aimed to stress the close connection between traditional problems of classical electrodynamics of microwaves and new problems arising from technological progress in synthesis and application of nanostructures. Such a con-

nection allows us to extend to nanostructures the rich experience and mathematical approaches accumulated in classical electrodynamics. On the other hand, in our chapter we have demonstrated the peculiarities of electromagnetic problems in nanostructures irreducible to problems in classical electrodynamics due to the complex conductivity law and pronounced field inhomogeneity. The nanotube is a nice example to demonstrate the correctness of both ideas, but it is not the only one. Nanostructures as a whole are the field of applicability of such a methodology. Certainly, in this chapter we have touched upon only a restricted set of problems, one that is far from being complete. It shows, however, that the range of problems is very wide and that the methods and techniques of traditional electrodynamics can be successfully adapted to nanostructures.

Acknowledgments

We are grateful to Drs. O.M. Yevtushenko and I. Herrmann for long-term collaboration, and to Dr. I. Bondarev for fruitful cooperation in the investigation of spontaneous emission process in carbon nanotubes. The research was partially supported by the BMBF under Project No. BEL-001-01, and from the Belarus Foundation for Fundamental Research under Project Nos. F02-176 and F02 R-047.

References

- [1] L.A. Weinstein, *Open resonators and open waveguides*, Golem, New York, NY, USA, 1969.
- [2] M.J. Adams, *An introduction to optical waveguides*, Wiley, New York, NY, USA, 1981.
- [3] V.V. Shevchenko, *Tapers in open waveguides*, Golem, Boulder, CO, USA, 1971.
- [4] K.C. Gupta, R. Garg and R. Chadha, *Computer aided design of microwave circuits*, Artech House, Boston, MA, USA, 1981.
- [5] A.S. Ilyinsky, G.Ya. Slepyan and A.Ya. Slepyan, *Propagation, scattering and dissipation of electromagnetic waves*, Peter Peregrinus, London, UK, 1993.
- [6] M.S. Dresselhaus, G. Dresselhaus and P.C. Eklund, *Science of fullerenes and carbon nanotubes*, Academic Press, New York, NY, USA, 1996.
- [7] T.W. Ebbesen (ed), *Carbon nanotubes: Preparation and properties*, CRC Press, Boca Raton, FL, USA, 1997.
- [8] R. Saito, G. Dresselhaus and M.S. Dresselhaus, *Physical properties of carbon nanotubes*, Imperial College Press, London, UK, 1998.
- [9] M.S. Dresselhaus, G. Dresselhaus and Ph. Avouris, *Carbon nanotubes*, Springer, Berlin, Germany, 2001.
- [10] F.G. Bass and A.A. Bulgakov, *Kinetic and electrodynamic phenomena in classical and quantum semiconductor superlattices*, Nova Science, New York, NY, USA, 1997.

- [11] Ph. Avouris, T. Hertel, R. Martel, T. Schmidt, H.R. Shea and R.E. Walkup, Carbon nanotubes: Nanomechanics, manipulation, and electronic devices, *Appl Surf Sci* **141** (1999), 201–209.
- [12] P.G. Collins and Ph. Avouris, Nanotubes for electronics, *Sci Am* **283** (6) (2000), 62–69 (December issue).
- [13] S.J. Tans, A.R.M. Verschueren and C. Dekker, Room-temperature transistor based on a single carbon nanotube, *Nature* **393** (1998), 49–52.
- [14] A.A. Farajian, K. Estarjani and Y. Kawazoe, Nonlinear coherent transport through doped nanotube junctions, *Phys Rev Lett* **82** (1999), 5084–5087.
- [15] F. Leonard and J. Tersoff, Role of Fermi-level pinning in nanotube Schottky diodes, *Phys Rev Lett* **84** (2000), 4693–4696.
- [16] A.A. Odintsov, Schottky barriers in carbon nanotube heterojunctions, *Phys Rev Lett* **85** (2000), 150–153.
- [17] G.Ya. Slepyan, S.A. Maksimenko, A. Lakhtakia, O.M. Yevtushenko and A.V. Gusakov, Electronic and electromagnetic properties of nanotubes, *Phys Rev B* **57** (1998), 9485–9497.
- [18] A. Lakhtakia, G.Ya. Slepyan, S.A. Maksimenko, O.M. Yevtushenko and A.V. Gusakov, Effective medium theory of the microwave and the infrared properties of composites with carbon nanotube inclusions, *Carbon* **36** (1998), 1833–1838.
- [19] G.Ya. Slepyan, S.A. Maksimenko, A. Lakhtakia, O.M. Yevtushenko and A.V. Gusakov, Electrodynamics of carbon nanotubes: Dynamic conductivity, impedance boundary conditions and surface wave propagation, *Phys Rev B* **60** (1999), 17136–17149.
- [20] S.A. Maksimenko and G.Ya. Slepyan, Electrodynamical properties of carbon nanotubes, *Electromagnetic fields in unconventional materials and structures* (O.N. Singh and A. Lakhtakia, eds), Wiley, New York, NY, USA, 2000, 217–255.
- [21] G.Ya. Slepyan, N.A. Krapivin, S.A. Maksimenko, A. Lakhtakia, O.M. Yevtushenko, Scattering of electromagnetic waves by a semi-infinite carbon nanotube, *Arch Elektron Übertrag* **55** (2001), 273–280.
- [22] S.A. Maksimenko and G.Ya. Slepyan, Electrodynamics of carbon nanotubes, *J Commun Technol Electron* **47** (2002), 235–252.
- [23] O.M. Yevtushenko, G.Ya. Slepyan, S.A. Maksimenko, A. Lakhtakia and D.A. Romanov, Nonlinear electron transport effects in a chiral carbon nanotube, *Phys Rev Lett* **79** (1997), 1102–1105.
- [24] G.Ya. Slepyan, S.A. Maksimenko, V.P. Kalosha, J. Herrmann, E.E.B. Campbell and I.V. Hertel, Highly efficient high harmonic generation by metallic carbon nanotubes, *Phys Rev A* **61** (1999), R777–R780.
- [25] A.S. Maksimenko and G.Ya. Slepyan, Negative differential conductivity in carbon nanotubes, *Phys Rev Lett* **84** (2000), 362–365.
- [26] G.Ya. Slepyan, S.A. Maksimenko, V.P. Kalosha, A.V. Gusakov and J. Herrmann, High-order harmonic generation by conduction electrons in carbon nanotube rope, *Phys Rev A* **63** (2001), 053808.

- [27] C. Stanciu, R. Ehlich, V. Petrov, O. Steinkellner, J. Herrmann, I.V. Hertel, G.Ya. Slepyan, A.A. Khrutchinski, S.A. Maksimenko, F. Rotermund, E.E.B. Campbell and F. Rohmund, Experimental and theoretical study of third-order harmonic generation in carbon nanotubes, *Appl Phys Lett* **81** (2002), 4064–4066.
- [28] I.V. Bondarev, G.Ya. Slepyan and S.A. Maksimenko, Spontaneous decay of excited atomic states near a carbon nanotube, *Phys Rev Lett* **89** (2002), 115504.
- [29] S.Ya. Kilin, Quantum information, *Phys Usp* **42** (1999), 435–452.
- [30] G.Ya. Slepyan and S.A. Maksimenko, Effective boundary conditions in electrodynamics of nanostructures, *Proceedings of Bianisotropic'2000: 8th international conference on electromagnetics of complex media* (A.M. Barbosa and A.L. Topa, eds), Lisbon, Portugal (2000), 175–180.
- [31] F.G. Bass, A.Ya. Slepyan and G.Ya. Slepyan, Resonant oscillations of diffraction structures with weak nonlinearity, *Microwave Opt Technol Lett* **19** (1998), 203–208.
- [32] H.W. Kroto, J.R. Heath, S.C. O'Brien, R.F. Curl and R.E. Smalley, C⁶⁰: Buckminsterfullerene, *Nature* **318** (1985), 162–165.
- [33] S. Iijima, Helical microtubules of graphitic carbon, *Nature* **354** (1991), 56–58.
- [34] L.-Ch. Qin, X. Zhao, K. Hirahara, Y. Miyamoto, Y. Ando and S. Iijima, The smallest carbon nanotube, *Nature* **408** (2000), 50.
- [35] Y.K. Kwon, S. Saito and D. Tomanek, Effect of intertube coupling on the electronic structure of carbon nanotube ropes, *Phys Rev B* **58** (1998), R13314–R13317.
- [36] R. Saito, M. Fujita, G. Dresselhaus and M.S. Dresselhaus, Electronic structure of graphene tubules based on C₆₀, *Phys Rev B* **46** (1992), 1804–1811.
- [37] M.F. Lin and K.W.-K. Shung, Plasmons and optical properties of carbon nanotubes, *Phys Rev B* **50** (1994), 17744–17747.
- [38] P.R. Wallace, The band theory of graphite, *Phys Rev* **71** (1947), 622–634.
- [39] J.W. Mintmire and C.T. White, Universal density of states for carbon nanotubes, *Phys Rev Lett* **81** (1998), 2506–2509.
- [40] W. Tian and S. Datta, Aharonov–Bohm-type effect in graphene tubules: A Landauer approach, *Phys Rev B* **49** (1994), 5097–5100.
- [41] H. Suzuura and T. Ando, Phonons and electron-phonon scattering in carbon nanotubes, *Phys Rev B* **65** (2002), 235412.
- [42] Y. Miyamoto, S.G. Louie and M.L. Cohen, Chiral conductivities of nanotubes, *Phys Rev Lett* **76** (1996), 2121–2124.
- [43] P. Kral, E.J. Mele and D. Tomanek, Photogalvanic effects in heteropolar nanotubes, *Phys Rev Lett* **85** (2000), 1512–1515.
- [44] O.V. Kibis, Electronic phenomena in chiral carbon nanotubes in the presence of a magnetic field, *Physica E* **12** (2002), 741–744.
- [45] Yu.A. Il'inskii and L.V. Keldysh, *Electromagnetic response of material media*, Plenum Press, New York, NY, USA, 1994.

- [46] R. Courant and D. Hilbert, *Methods of mathematical physics, Vol. 1*, Interscience, New York, NY, USA, 1962.
- [47] L.M. Woods and G.D. Mahan, Electron-phonon effects in graphene and arm-chair (10,10) single-wall carbon nanotubes, *Phys Rev B* **61** (2000), 10651–10663.
- [48] S.J. Tans, M.H. Devoret, H. Dai, A. Thess, R.E. Smalley, L.J. Geerligs and C. Dekker, Individual single-wall carbon nanotubes as quantum wires, *Nature* **386** (1997), 474–477.
- [49] F.J. García-Vidal, J.M. Pitarke and J.B. Pendry, Effective medium theory of the optical properties of aligned carbon nanotubes, *Phys Rev Lett* **78** (1997), 4289–4292.
- [50] W. Lü, J. Dong and Z.-Ya. Li, Optical properties of aligned carbon nanotube systems studied by the effective-medium approximation method, *Phys Rev B* **63** (2000), 033401.
- [51] J.M. Pitarke and F.J. García-Vidal, Electronic response of aligned multishell carbon nanotubes, *Phys Rev B* **64** (2001), 073404.
- [52] M.F. Lin and D.S. Chuu, π -plasmons in carbon nanotube bundles, *Phys Rev B* **57** (1998), 10183–10187.
- [53] L.A. Weinstein, *Electromagnetic waves*, Soviet Radio, Moscow, USSR, 1988 (in Russian).
- [54] D.J. Hoppe and Y. Rahmat-Samii, *Impedance boundary conditions in electromagnetics*, Taylor & Francis, Washington, DC, USA, 1995.
- [55] R. Tamura and M. Tsukada, Analysis of quantum conductance of carbon nanotube junctions by the effective-mass approximation, *Phys Rev B* **58** (1998), 8120–8124.
- [56] M. Menon, D. Srivastava and S. Saini, Fullerene-derived molecular electronic devices, *Semiconductor Sci Technol* **13** (1998), A51–A54.
- [57] A.N. Andriotis, M. Menon, D. Srivastava and L. Chernozatouskii, Ballistic switching and rectification in single wall carbon nanotube Y junctions, *Appl Phys Lett* **79** (2001), 266–268.
- [58] A.N. Andriotis, M. Menon, D. Srivastava and L. Chernozatonskii, Rectification properties of carbon nanotube “Y-junctions,” *Phys Rev Lett* **87** (2001), 066802.
- [59] Y.-G. Yoon, M.S.C. Mazzoni, H.J. Choi, J. Ihm and S.G. Louie, Structural deformation and intertube conductance of crossed carbon nanotube junctions, *Phys Rev Lett* **86** (2001), 688–691.
- [60] C. Bena, S. Vishveshwara, L. Balents and M.P.A. Fisher, Quantum entanglement in carbon nanotubes, *Phys Rev Lett* **89** (2002), 037901.
- [61] A. Rochefort, D.R. Salahub and Ph. Avouris, Effects of finite length on the electronic structure of carbon nanotubes, *J Phys Chem B* **103** (1999), 641–646.
- [62] L.A. Weinstein, *The theory of diffraction and the factorization method*, Golem, New York, NY, USA, 1969.

- [63] R. Mittra and S.W. Lee, *Analytical techniques in the theory of guided waves*, Macmillan, New York, NY, USA, 1971.
- [64] G.Ya. Slepyan, S.A. Maksimenko, F.G. Bass and A. Lakhtakia, Nonlinear electromagnetics in chiral media: self-action of waves, *Phys Rev E* **52** (1995), 1049–1058.
- [65] V.M. Shalaev, *Nonlinear optics of random media: Fractal composites and metal-dielectric films*, Springer, Berlin, Germany, 2000.
- [66] A. L'Huillier, K. Schafer and K. Kulander, Higher-order harmonic generation in xenon at 1064 nm: The role of phase matching, *Phys Rev Lett* **66** (1991), 2200–2203.
- [67] J. Krause, K. Schafer and K. Kulander, High-order harmonic generation from atoms and ions in the high intensity regime, *Phys Rev Lett* **68** (1992), 3535–3538.
- [68] J. Macklin, J. Kmetec and C. Gordon III, High-order harmonic generation using intense femtosecond pulses, *Phys Rev Lett* **70** (1993), 766–769.
- [69] D. von der Linde, T. Engers, G. Jenke, P. Agostini, G. Grillon, E. Nibbering, A. Mysyrowicz and A. Antonetti, Generation of high-order harmonics from solid surfaces by intense femtosecond laser pulses, *Phys Rev A* **52** (1995), R25–R27.
- [70] P.A. Norreys, M. Zepf, S. Moustazis, A.P. Fews, J. Zhang, P. Lee, M. Bakarezos, C.N. Danson, A. Dyson, P. Gibbon, P. Loukakos, D. Neely, F.N. Walsh, J.S. Wark and A.E. Dangor, Efficient extreme uv harmonics generated from picosecond laser pulse interactions with solid targets, *Phys Rev Lett* **76** (1996), 1832–1835.
- [71] F. Leonard and J. Tersoff, Negative differential resistance in nanotube devices, *Phys Rev Lett* **85** (2000), 4767–4770.
- [72] A. Bezryadin, A.R.M. Verschueren, S.J. Tans and C. Dekker, Multiprobe transport experiments on individual single-wall carbon nanotubes, *Phys Rev Lett* **80** (1998), 4036–4039.
- [73] R. Eigger and A.O. Gogolin, Effective low-energy theory for correlated carbon nanotubes, *Phys Rev Lett* **79** (1997), 5082–5085.
- [74] C.L. Kane, L. Balents and M.P.A. Fisher, Coulomb interactions and mesoscopic effects in carbon nanotubes, *Phys Rev Lett* **79** (1997), 5086–5089.
- [75] Yu.A. Krotov, D.-H. Lee and S.G. Louie, Low energy properties of (n,n) carbon nanotubes, *Phys Rev Lett* **78** (1997), 4245–4248.
- [76] R.J. Glauber and M. Lewenstein, Quantum optics of dielectric media, *Phys Rev A* **43** (1991) 467–491.
- [77] R. Matloob, R. Loudon, S.M. Barnett and J. Jetters, Electromagnetic field quantization in absorbing dielectrics, *Phys Rev A* **52** (1995), 4823–4838.
- [78] T. Gruner and D.-G. Welsch, Green-function approach to the radiation-field quantization for homogeneous and inhomogeneous Kramers–Kronig dielectrics, *Phys Rev A* **53** (1996), 1818–1829.
- [79] W. Vogel, D.-G. Welsch and S. Wallentowitz, *Quantum optics: An introduction*, Wiley, New York, NY, USA, 2001.

- [80] H.T. Dung, L. Knöll and D.-G. Welsch, Spontaneous decay in the presence of dispersing and absorbing bodies: General theory and application to a spherical cavity, *Phys Rev A* **62** (2000), 053804.
- [81] H.T. Dung, L. Knöll and D.-G. Welsch, Decay of an excited atom near an absorbing microsphere, *Phys Rev A* **64** (2001), 013804.
- [82] M. Fleischhauer, Spontaneous emission and level shifts in absorbing disordered dielectrics and dense atomic gases: A Green's-function approach, *Phys Rev A* **60** (1999), 2534–2539.
- [83] V.B. Berestetskii, E.M. Lifshitz and L.P. Pitaevskii, *Quantum electrodynamics*, Pergamon Press, Oxford, UK, 1982.
- [84] V.V. Klimov and M. Ducloy, Allowed and forbidden transitions in an atom placed near an ideally conducting cylinder, *Phys Rev A* **62** (2000), 043818.
- [85] T. Sondergaard and B. Tromborg, General theory for spontaneous emission in active dielectric microstructures: Example of a fiber amplifier, *Phys Rev A* **64** (2001), 033812.
- [86] S. Tasaki, K. Maekawa and T. Yamabe, π -band contribution to the optical properties of carbon nanotubes: Effects of chirality, *Phys Rev B* **57** (1998), 9301–9318.
- [87] E.M. Purcell, Spontaneous emission probabilities at radio frequencies, *Phys Rev* **69** (1946), 681.
- [88] E.P. Petrov, V.N. Bogomolov, I.I. Kalosha and S.V. Gaponenko, Spontaneous emission of organic molecules embedded in a photonic crystal, *Phys Rev Lett* **81** (1998), 77–80.
- [89] M.O. Scully and M.S. Zubairy, *Quantum optics*, Cambridge University Press, Cambridge, UK, 2001.



Sergey A. Maksimenko was born in Belarus in 1954. He received a M.S. degree in physics of heat and mass transfer from Belarus State University, Minsk, in 1976; a Ph.D. degree in theoretical physics from a Belarus State University in 1988; and a Sc.D. degree in theoretical physics from Institute of Physics, Minsk, in 1996. He is currently Deputy Vice-Rector of Belarus State University and a Head of Laboratory at the Institute for Nuclear Problems at Belarus State University. He has authored or coauthored more than 80 conference and journal papers.

Dr. Maksimenko was a member of the Scientific Advisory Committee of *Bianisotropics*'98. His current research interests are electromagnetic wave theory, diffraction by periodic mediums and structures, and electromagnetic processes in quasi-one- and zero-dimensional nanostructures in condensed matter.



Gregory Ya. Slepyan was born in Minsk, Belarus, in 1952. He received a M.S. degree in radioengineering from the Minsk Radioengineering Institute in 1974; a Ph.D. degree in physics from Belarus State University, Minsk, in 1979; and a Sc.D. degree in physics from Kharkov State University, Kharkov, Ukraine, in 1988. He is currently a Principal Researcher at the Institute for Nuclear Problems at Belarus State University. He has authored or coauthored more than 100 conference and journal theoretical papers, as well as two books. Dr. Slepyan is a member of the editorial board of *Electromagnetics*, and he was a member of the Scientific Advisory Committee of the conferences *Bianisotropics'97* and *Bianisotropics'2000*. His current research interests are diffraction theory, microwave and millimeter-wave circuits, nonlinear waves and oscillations, and nanostructures in condensed matter with applications to nanoelectronics.

Randomness in Complex Materials

H. John Caulfield, Donald O. Henderson, and
Mikhail A. Noginov

Abstract

Randomness serves many purposes where materials sciences and optical sciences meet. It can provide the raw material for self-organization. It can uniformize optical properties, make manufacturing easier, and assure a great deal of noise immunity. Although all of those features are commonly exhibited together, we illustrate them with examples in which one tends to dominate the other. Specifically, we discuss the examples of powder lasers, self-organized pseudorandom surfaces for Raman spectroscopy, and designer diffusers.

1 Introduction

Randomness is normally viewed as something undesirable. It suggests ignorance, noise, sloppiness, and other unpleasant things. Our goal here is to explore the uses of randomness in complex material mediums.

By way of background, it is worth pointing out that randomness is a respected and much-used tool in mathematics and engineering. Random signals are used to probe systems. Random noise (under well-defined conditions) can be used to enhance desirable components of weak signals or images. This area is called *stochastic resonance* [1]. No agricultural or medical trial can go on without a great deal of attention to the randomization of certain key variables. So, it seems reasonable to explore randomization in materials sciences.

We argue in this chapter that randomness has four specific and important virtues in material sciences: (i) it provides the needed *raw material* for self-organization; (ii) it can achieve a degree of uniformization of optical properties; (iii) it can be much simpler and cheaper than highly organized materials such as crystals (intermediate states such as glass are explicitly excluded from this discussion); (iv) it can offer great immunity or insensitivity to noise or damage. Let us look at each of these claims in a general context before we present some specific examples.

Self-organization is one of the wonderful emergent properties of what physicists like to call *nonlinear dynamics* and materials scientists (after Prigogine [2]) like to call *behavior very far from equilibrium*. There are many other names as well. The most important cases for our purposes are those in which a new kind of ordered behavior arises quite suddenly in response to a small change in some control variable. Many properties of such systems are general enough to allow universal rules to be applied. What is important for this discussion is that these ordered states arise from random states far from equilibrium. Random pre-organization is essential for the organized state to emerge.

Uniformity is a goal often sought in optics. Randomness in materials can help uniformize any optical parameter: color, spatial pattern, temporal variations, polarization, etc. This randomization uses the mechanism of multiple scattering.

One of the primary activities of materials scientists is to grow crystals. Crystals are highly organized, macroscopic structures that may exhibit simple, predictable behavior. Anisotropies, chirality and the like (see the chapter of Weiglhofer of this volume) can remove all simplicity from such systems, though. Unfortunately, from the view of device makers, growing crystals is still very difficult, and large crystals of some useful materials either do not exist or are much too expensive for widespread use. Random materials are often mesoscopic crystals: bigger than quantum dots, so they have predictable bulk properties, yet far easier to make. Readers interested in more detail on this issue are referred to the chapter of Mackay of this volume. As the characteristic frequencies of inhomogeneities are comparable to the wavelength, simpler methods such as the Maxwell Garnett formula do not apply.

In 1955, the Rand Corporation published a book titled *A Million Random Digits with 100,000 Normal Deviates* [3]. Those digits had many practical uses. When

he was an undergraduate student at Rice University, the lead author used to post “Today’s Random Digits” on the Physics Department bulletin board, until people started calling him a random deviate. The question to contemplate here is: Was it important to proofread to make certain the digits posted were the ones in the book? Our answer is: no. Random is random. If someone makes a random error in copying a random number, its randomness does not decrease. By the same reasoning, random materials are likely to be quite immune to mistakes, noise, and the like.

The reason these considerations had to be discussed in the introduction is that one can treat them as distinct subjects only in theory. In practice, most or all of these advantages obtain in every example we have chosen for this chapter.

2 Raw material for self-organization

Arguably, self-organization is the most important useful product of randomness. In the absence of *a priori* information on what a self-organizing system will be asked to do or *choose* to do, the best preparation is to present many randomized options. This is a trick used over and over in mathematics. Moore’s law has held for about five decades. Earlier, computing was slow and expensive. Monte Carlo sampling allows one to approximate the unknown with predictable expected error if we provide truly random sampling [4-5]. Likewise, the validity of every agricultural or medical testing program depends on and assumes true randomness. For the same reasons, random initial states are critical for truly flexible self-organization. Let us give some examples.

One of the most exciting and unexpected things in nonlinear optics is mutual self-conjugation, in which two independent beams phase-conjugate each other perfectly, even though they are mutually incoherent and may not even coexist in the material at the same time [6]. How can that possibly happen? What couples the uncoupled beams? The answer, surprisingly, is noise [7]. Random scatterers scatter each beam forward and backward. In the appropriate photorefractive material and in the appropriate geometry, charge movements are impacted in such a way that the forward scatter of one beam couples with the backscatter of the other, and conversely. *Without a little randomness, this effect could not occur. Nonscattering crystals will not exhibit the effect.*

Mutual phase conjugation, in turn, allows all sorts of otherwise impossible self-organized useful behaviors to occur. We ourselves have achieved the following ones: (i) we can couple light efficiently from a handheld laser into a single-mode fiber [7]; (ii) we can allow optical connections to adaptively follow changing alignments [6]; (iii) we can perform conventional interferometry on the output of a self-conjugated beam from a multimode fiber; and (iv) we can send a strong beam to retrace the path a weaker beam took to get to us [7]. These by no means exhaust the complexity effects observable in photorefractive materials. For a more general discussion of one-beam, two-beam, and three-beam complexity in nonlinear optics, see Ref. 8.

Another common case of self-organized coupling of optics with randomness is enhanced backscatter. Figure 1 shows scattering from a regular crystal (Bragg scat-

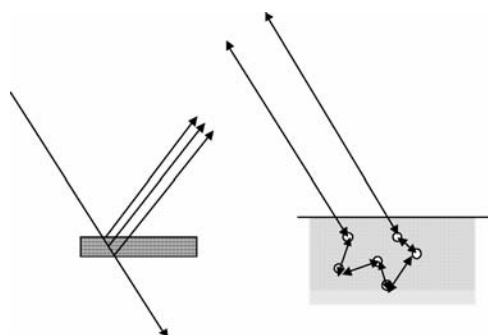


Figure 1 Left: Bragg scattering is resonant enhancement of scattering in certain directions as a result of the regular structure of a crystal. Right: Enhanced backscatter is the random equivalent. It also yields enhanced scatter when a resonance occurs in self-organized paths between incoming and outgoing parts of the same beam of light.

tering) and its random counterpart called *enhanced backscatter*. When a path that can reverse the beam direction occurs, enough coherent interaction among back- and forward-scattered beams occurs to offer a mild enhancement of the backscatter. If this occurs in a highly nonlinear random medium, the enhancement itself can be enhanced. We can put these effects to purposes both physically interesting and potentially useful, by converting random powders into self-organized lasers.

3 Random lasing in scattering solid-state materials

3.1 History and the state of the art

In a laser with a Fabry–Pérot cavity and resonant feedback [9-10], stimulated emission is spatially coherent and its frequencies are primarily determined by the eigenmodes of the resonator. In 1966, Ambartsumyan *et al.* proposed a new type of laser wherein nonresonant feedback occurs through reflection off a highly scattering medium used in place of the back mirror [11-12]. In the laser with nonresonant (incoherent) feedback, the emission frequency was determined by the resonant frequency of the gain medium (instead of the eigenmodes of the resonator); the laser had low coherence, and its photon statistics and the dynamics of the line narrowing were different from those in regular lasers. Such types of laser could be useful when high spatial uniformity of illumination and high stability of the emission wavelength are desired.

In 1967–1968, Letokhov took a step further and theoretically predicted the possibility of generating laserlike light by scattering particles with negative absorption in the case when the motion of photons is diffuse [13-14]. In the proposed system, the same scattering material would play the roles of an active laser medium and of a resonator providing feedback. Solving the diffusion equation for the propagation of emitted photons in the amplifier medium together with the rate equation for gain, Letokhov calculated the laser threshold as well as the linewidth of the stimulated

emission; and he also predicted damped oscillations in the emission dynamics—the so-called *relaxation oscillations* [15]–[17]. This was probably the first report of what we now call a *random laser*.

In 1986, Markushev *et al.* found that above a pumping energy threshold, the duration of the emission pulse in $\text{Na}_5\text{La}_{1-x}\text{Nd}_x(\text{MoO}_4)_4$ powder shortens by approximately four orders of magnitude [18]. About the same strong enhancement was found in the intensity of the strongest spectral component of the $^4\text{F}_{3/2}$ – $^4\text{I}_{11/2}$ transition (at a free-space wavelength of $\sim 1.06\ \mu\text{m}$), with the linewidth narrowed significantly. Only one narrow emission line emerged above the threshold. The plot of the intensity of this emission line *versus* pumping energy resembled the input/output relationships in regular lasers. Since the observed experimental features were characteristic of conventional lasers, Markushev *et al.* concluded that the observed emission was stimulated radiation from $\text{Na}_5\text{La}_{1-x}\text{Nd}_x(\text{MoO}_4)_4$ powder.

Many experimental studies of solid-state random lasers following the pioneering work [18] were focused on *pulsed* stimulated emission in neodymium-activated scattering materials [19]–[26]. Experimentally, only spontaneous emission is observed in random laser materials at weak pumping. In this regime, the spatial uniformity of luminescence from the surface of the sample resembles that of pumping. At more intense pumping, amplified spontaneous emission (ASE) strengthens, thereby narrowing the emission spectrum. When pumping power reaches the threshold, the spectral and temporal properties of emission change significantly. The well-known phenomena observed in neodymium-activated random (powder) lasers include the following: (i) a dramatic shortening of the emission pulse and narrowing of the emission spectral line above the threshold [18]–[20], [23]–[26] (see Figures 2 and 3); (ii) input-output dependence, which is characteristic of conventional lasers [18], [23], [27] (see Figure 4); (iii) multiple-pulsed emission resembling relaxation oscillations in regular lasers, observed experimentally [19], [22]–[24], [28] and examined theoretically [17], [19], [23–24], [29]–[31];

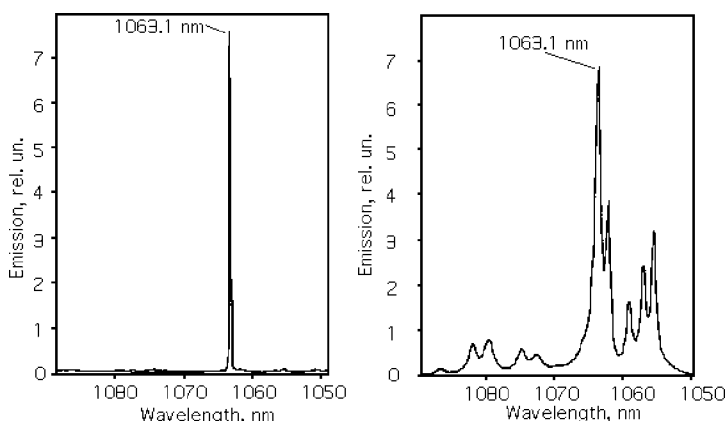


Figure 2 Emission spectrum of $\text{NdAl}_3(\text{BO}_3)_4$ powder, when the pumping wavelength is 532 nm, and the pumping pulse duration is 10 ns. Left: Pumping intensity of $\sim 30\ \text{mJ}/\text{cm}^2$ is below the threshold. Right: Pumping intensity of $\sim 240\ \text{mJ}/\text{cm}^2$ is above the threshold [23].

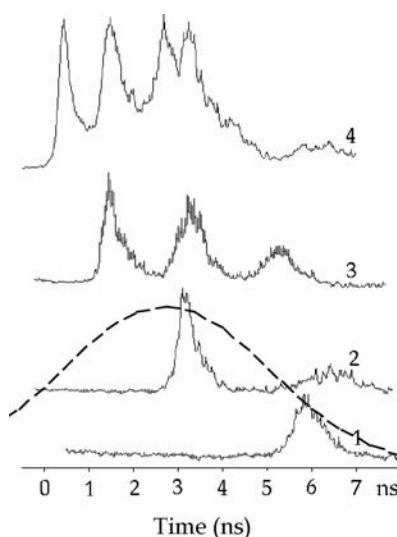


Figure 3 Pulses of stimulated emission in $\text{NdAl}_3(\text{BO}_3)_4$ powder (1) near the threshold energy (200 mJ/cm^2), (2) at 320 mJ/cm^2 energy, (3) at 380 mJ/cm^2 energy, and (4) at 780 mJ/cm^2 energy. The bell-shaped trace in the bottom of the figure is the approximate position and shape of the pumping pulse (at 532 nm wavelength). The mean powder particle size in this particular experiment was equal to $3.6 \mu\text{m}$ [23].

and (iv) drift of the stimulated emission frequency as well as hopping of the emission line from one discrete frequency to another [20], [23], [25]. In contrast with the predictions of Ambartsumyan *et al.* [11-12], the stability of the emission wavelength in random lasers is not very high.

Both low [22], [32]-[35] and reasonably high [33]-[35] degrees of coherence have been experimentally realized in Nd^{3+} random lasers in different experiments. In some experiments, localized bright spots are observed in the near-field emission above the threshold [33]-[35].

The early models describing the behavior of random lasers, which contradicted each other to some extent, were based on ASE with incoherent feedback provided by scatterers [13-14], [36], internal resonances in single particles [21], and coherent resonance modes in the ensembles of coupled particles [19], [24].

During the 1990s, many new random laser materials operating in different regimes were explored. These include Pr^{3+} powder lasers [25], Ti:sapphire powder lasers [28], and random lasers based on pulverized LiF with color centers [37]. Stimulated emission in laser powders pumped by electron beams has been reported too [34], [38]. Auzel and Goldner [26] observed the narrowing of the emission spectral line and the shortening of the emission pulses, similar to those in random lasers, in Er:YLiF_4 powder. They explained the phenomenon in terms of superfluorescence in the so-called Dicke-Bonifacio sense [39]. Not to a small degree, a renaissance in the area of random lasers was inspired by Lawandy *et al.* [40], who in 1994 observed a similar phenomenon in liquid dye with scatterers. (Liquid random lasers are not discussed in this chapter.)

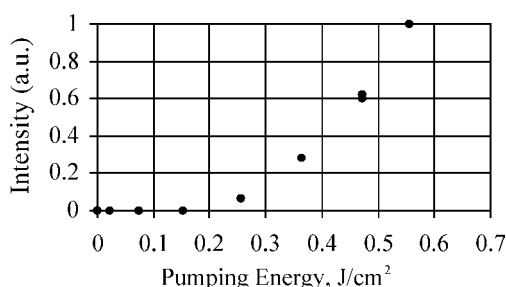


Figure 4 Dependence of the 1063.1-nm emission intensity in $\text{NdAl}_3(\text{BO}_3)_4$ powder on the pumping pulse energy (at 532-nm wavelength) [23].

The two random laser materials studied in great detail in recent years are ZnO nanopowder [41]–[48] and scattering polymers with gain [49]–[51]. (Note that laser emission in ZnO powder was first reported in 1981 [52].) Both materials have broadband gain spectrums and exhibit somewhat similar stimulated emission properties that are discussed next.

At weak pumping, only spontaneous emission is observed; and the spatial uniformity of luminescence from the surface of the sample resembles that of pumping. At more intense pumping, ASE strengthens and thus narrows the emission spectrum. When pumping power reaches the threshold, the first narrow emission peak appears in the spectrum close to the top of the ASE band. The spectral width of this peak can be several orders of magnitude smaller than that of the ASE band. The intensity of the narrow spectral peak (which depends on the experimental conditions) is typically comparable with the emission intensity below the threshold. Simultaneously, first bright spots appear in the in the near-field emission pattern [44–45]. With further increase of the pumping power, both the number of sharp peaks in the spectrum and the number of bright spots in the near-field emission pattern increase.

The transformation of the emission spectrum in ZnO powder with the increase of the pumping intensity is shown in Figure 5. The spectral positions of the narrow emission peaks are specific to the particular pumped spot. When the pumping beam is moved along the surface of the sample, the wavelengths of the emission lines change randomly. Qualitatively, this behavior is similar to that of a conventional laser with an optically clear gain medium and a highly scattering (“speckle”) back mirror [53]. It was concluded that sharp spectral spikes manifest high coherence modes of stimulated emission, and analysis of the photon count statistics supported that conclusion [46].

To explain the sharp lines observed in the stimulated emission spectrum of ZnO powder, Cao *et al.* suggested that, because of very strong scattering, the closed-loop paths (which serve as ring laser cavities) can be formed in a random medium through multiple optical scattering [41–42] (Figure 5, inset). Accordingly, narrow lines observed in the random laser emission spectrums represent the eigenmodes of the ring cavities. This simple model qualitatively explains some of the experimentally observed effects. However, to provide for a reasonably high quality factor

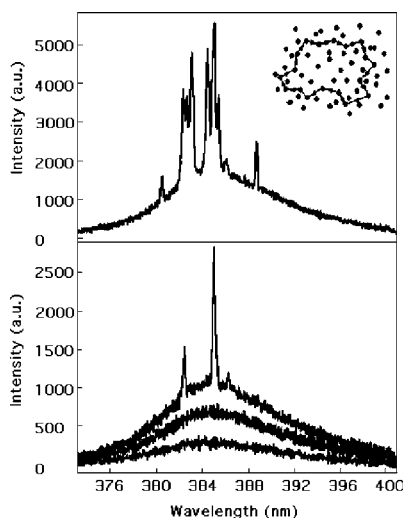


Figure 5 Transformation of the emission spectrum in ZnO powder with increase of the pumping intensity (from bottom to top) 400, 562, 763, and 875 kW/cm². Inset: a schematic diagram of the closed-loop path [42].

of the cavity, the loop model should assume deterministic reflections off scatterers, similar to specular reflections. Such reflections are not very probable when the average size ($\sim 0.1 \mu\text{m}$) of ZnO particles is much smaller than the emission wavelength ($\sim 0.4 \mu\text{m}$). Note that the incorporation of deterministic scattering of light is, in principle, possible in the model that treats the scattering material as an ensemble of dipole oscillators associated with individual particles [54].

The modes—which are quasi-stationary distributions of the electromagnetic field—are obtained as eigensolutions of the Maxwell equations in a medium with modulated permittivity ε and appropriate boundary conditions. Theoretical examination of random-laser modes and their properties has been carried out in large number of works [46], [55]–[59]. In particular, calculations predict (i) threshold behavior of the stimulated emission in a random medium [46], [55]–[58]; (ii) *hot spots* in the near-field emission pattern [45], [58]; (iii) relaxation oscillations [57]; (iv) an increase in the number of the modes with increasing pumping intensity and system length; and (v) the repulsion of laser modes related to localization effects [55]–[58]. The last result seems to be in disagreement with that of Ref. 59 predicting that the localized modes of the passive medium are not modified by the resonance gain.

3.2 Formation of a coherent mode

Numerical simulations do not help much to understand *how* the coherent mode is formed. Let us propose here a qualitative model explaining self-organization of a coherent stimulated emission state in random lasers from chaotic distribution of spontaneously emitted photons with individual random frequencies and phases. Let

us first consider a hypothetical collection of randomly positioned point sources of monochromatic light, all of which have random phases, as shown in Figure 6. A high degree of monochromaticity is possible in random lasers, provided the size of the lasing medium is much smaller than the longitudinal coherence length (typically, 1 cm [32]). In principle, one can insert several time-delay devices (slabs, prisms, reflectors, etc.) in the medium and, with the use of these devices, synchronize phases in certain (larger or smaller) ensembles of light sources, as also depicted in Figure 6. This is equivalent to the formation in a medium of a three-dimensional interference pattern, such that the light sources belonging to selected ensembles are predominantly positioned in the hot spots of the field distribution.

Certainly, any given configuration of time-delay devices—to provide phase synchronization within an ensemble—is wavelength-specific. If the emission wavelength is changed slightly, all time delays have to be adjusted to maintain synchronization. However, it is also possible that the same configuration of time-delay devices can support many discrete frequencies, like regular laser resonators do.

Many solid-state random laser materials are either dielectric/semiconductor powders or ceramics. They consist of a gain medium inside microcrystallites, air gaps between particles, and random scatterers (microcrystallites and their surfaces). With some modification, the following discussion can also be applied to scattering polymers.

Under weak optical pumping, an active medium emits spontaneous radiation. If the spectral band of the luminescence transition is broad, then the spectral density of spontaneous emission is low. The spontaneously emitted photons propagate in different (random) directions, and the corresponding waves have different (random) phases. At some instants, certain fortuitous ensembles of emitting atoms are realized, when randomly moving photons go through just the right combination of scatterers, such that point light sources illuminate each other with synchronized ra-

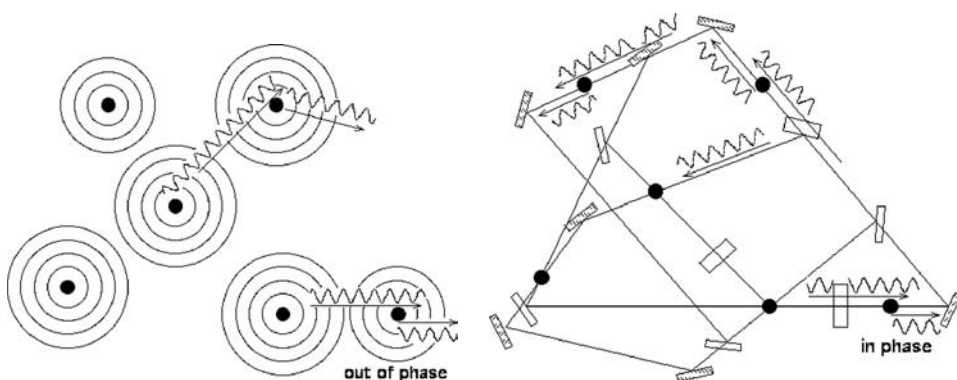


Figure 6 Left: Point sources of monochromatic light emitting with random phases. Right: An ensemble of the same point sources along with time-delay devices (shown as rectangular blocks) introduced to synchronize radiation phases, providing for the formation of a complex coherent electromagnetic mode.

diation; see Figure 6. The probability of such events is low. Besides, because of low gain, spontaneously emitted photons cannot induce a sufficient number of stimulated transitions to support the mode. Thus, instantaneous coherent configurations, randomly created from chaotic field distribution, vanish momentarily.

Stronger pumping leads to more efficient ASE and a corresponding narrowing of the spectral emission line. Thus, the spectral density of the emitted photons increases to make the realization of a synchronized mode more probable. What is even more important, when the coherent state (i.e., the synchronized-phase mode) actualizes, it has a high probability of being amplified *via* stimulated emission and becoming self-sustained. It appears likely that during some initial period of time the mode self-adjusts its shape and phase, thereby increasing the contrast of interference pattern and the efficiency of stimulated emission. Obviously, the higher the contrast of the interference pattern and the better the spatial overlap of its hot spots and the emitting centers, the more efficient is the stimulated emission.

Although both *photons* and *waves* are used in the foregoing discussion, the formation of a spatially distributed coherent mode is easier to understand in terms of waves and interfering electromagnetic fields. Similarly to the case of conventional lasers, the formation of a coherent mode in random lasers is more difficult to explain in terms of photons (particles). In the spectral domain, the appearance of a synchronized mode should obviously lead to the emergence of a narrow peak. The frequency of this peak is to be specified by an effective three-dimensional resonator constituted by an ensemble of excited atoms, which are coherently coupled by radiation. Thus, the model explains both sharp peaks in the spectrum and bright spots in the near-field emission pattern observed in many experimental works.

Auzel and Goldner [26] attributed enhanced emission in Er:LiYF₄ powder at cryogenic temperatures to superfluorescence (in the Dicke-Bonifacio sense [39]). Empirically, self-organization of a coherent state, as shown in Figure 6, has a certain analogy with the self-organization of cooperative spontaneous emission in the case of superfluorescence. In both cases, the phases of emitting centers in an ensemble are getting synchronized before the narrow-line intense coherent light radiation is emitted. In the case of superfluorescence, the emitting centers adjust their phases of oscillation through interaction *via* the electromagnetic field. On the contrary, in the synchronization scheme described here, nothing is changed within emitting centers. (This is a principal difference between the two cases.) Only photon walk paths and the phases of corresponding waves change randomly until a favorable combination realizes and gets locked in. One can say that in the proposed model, in contrast with superfluorescence, synchronization occurs regardless of the phase relaxation time; and thus it can be expected (and is observed experimentally) at room temperature.

The proposed model provides a plausible scenario for the formation of a *single coherent mode* via randomness. In principle, the same philosophy can be extended to the description of random lasers with incoherent feedback, wherein *stochastic resonators* formed by the scattering medium support large number of modes coupled through scattering [14].

4 Ease of manufacturing

Raman spectroscopy is undoubtedly the best way to discriminate among molecular structures. And it works at all wavelengths. The trouble is that for most materials, it is extremely insensitive. Recently, researchers have been able to produce gigantic enhancements of the Raman sensitivity by placing the molecules on specially prepared substrates—a technique called surface-enhanced Raman spectroscopy (SERS) [60]. It turns out that the best way to make these surfaces involves randomness.

That self-assembly arises from an initial condition of randomness is well-represented in the fabrication of SERS substrates. These substrates typically consist of an ensemble of mesoscopic metal particles, which may be fabricated by many approaches, including arrested precipitation from solution, chemistry in reverse micelles, inert-gas-induced condensation of metal vapor, and pulsed laser deposition, to mention a few. The metallic particles are then deposited onto a microscope slide.

Over the last few years, the sensitivity of SERS has advanced to the stage of single molecule detection corresponding to an enhancement factor of $\sim 10^{15}$, which is defined as

$$\rho(\mathbf{r}_m, \omega) = \left| \frac{\mathbf{E}(\mathbf{r}_m, \omega)}{E_{inc}(\omega)} \right|^4, \quad (1)$$

where $\mathbf{E}(\mathbf{r}_m, \omega)$ is the total electric field vector at position $\mathbf{r}_m$ of a molecule, and $E_{inc}(\omega)$ is the electric field amplitude associated with the incident plane wave of angular frequency ω . It is evident from (1) that the enhancement in the SERS signal depends not only on the total electric field at the position of the molecule on a metallic surface, but also on the morphology of the metallic surface. In addition, the field enhancement has been shown to be very sensitive to the topology of the metallic particles whose assemblage defines the surface. In particular, the largest experimental enhancements reported are for surfaces composed of metallic nanocrystals at the threshold of percolation. Model calculations that include coupling between *touching* metallic objects qualitatively predict the trend in the experimentally reported enhancements for surfaces with increasing fractional surface coverage, but severely underestimate the magnitude of the enhancements by as much as 10^6 . Much of the disagreement probably arises from two sources: (i) the details of the surface morphology and the degree of connectivity between the metallic particles on the surface, and (ii) the position of the molecule on surface. In addition, there has been some concern expressed about the chemical inertness of the metal surface used as a SERS substrate, particularly in the case of silver substrates, which pose more potential for chemical reactivity than those made from gold.

Based on the aforementioned considerations, control of the surface morphology and topology of a SERS substrate should play a central role in designing materials to render maximum enhancement. The control of morphology and topology is essentially a problem that is addressed by defining the parameter space for fabricating SERS substrates and is discussed next.

There are many methods for fabricating SERS substrates, which include mechanical roughening of a metal surface, thermal evaporation of metal onto a substrate, finely focusing ion beams for patterning a surface, and lithography. While these approaches for making SERS substrates have been successful to varying degrees, problems remain in reproducible fabrication of substrates with the same morphologies and topologies, particularly for substrates prepared by mechanical roughening and thermal evaporation. On the other hand, the finely focused ion-beam and lithographic techniques do provide excellent topological and morphological control, but do not have the spatial resolution required for patterning surfaces which support interparticle coupling effects associated with large enhancements.

An approach that we have developed for fabricating SERS substrates is pulsed laser deposition. This is a clean and simple technique as it involves a target, a source of coherent pulsed radiation and a substrate for collecting the particles generated in laser ablation plume. The parameter space for fabricating the SERS substrates is determined by the laser parameters (power density, wavelength, pulse-width, and number shots delivered to the target) as well as the target-to-substrate distance and orientation. Ancillary to the laser parameters are post-processing strategies that require a programmed temperature routine for thermal annealing in a controlled atmosphere.

Atomic force microscope (AFM) images of a gold nanocrystal SERS substrate fabricated by pulsed laser deposition are shown in Figure 7. The upper portion of the figure consists of gold nanocrystals with a mean height of ~ 3 nm. Subsequent annealing of the as-deposited substrate renders gold nanocrystals of mean height ~ 15 nm. Although it is not evident from the AFM images in the figure, the annealing conditions used actually transform the as-deposited sample from a nonpercolating structure into a percolating one. The transition from a nonpercolating topology to one that does percolate is confirmed by van der Pauw measurements that showed a resistivity of $10^5 \Omega \text{ cm}$ for the as-deposited sample and $2.2 \times 10^{-5} \Omega \text{ cm}$ after annealing for 1 hour at 400°C .

These results clearly indicate that pulsed laser ablation, when combined with a post-ablation annealing routine, does provide substrates with properties desired for very large SERS effects. This conclusion follows directly from the AFM images and resistivity measurements, which show that the surface is composed of gold nanocrystals that are in contact with each other to the extent that the ensemble of nanocrystals supports percolation.

The available results show that we can fabricate—by pulsed laser ablation and temperature-programmed annealing routines—gold nanocrystal SERS substrates that possess desirable attributes. However, there are questions that should be addressed to gain a better understanding of the details of the kinds of surface morphologies for optimized SERS responses. The issues that we address here are as follows: (i) the dependence of the SERS signal on the nanocrystal size; (ii) dependence of the SERS signal on substrates composed of nanocrystals with a bimodal size distribution; and (iii) a comparison of the SERS signal for gold and silver nanocrystal substrates with the constraint that both substrates have the same average nanocrystal size and are at the percolation threshold. It is expected that a

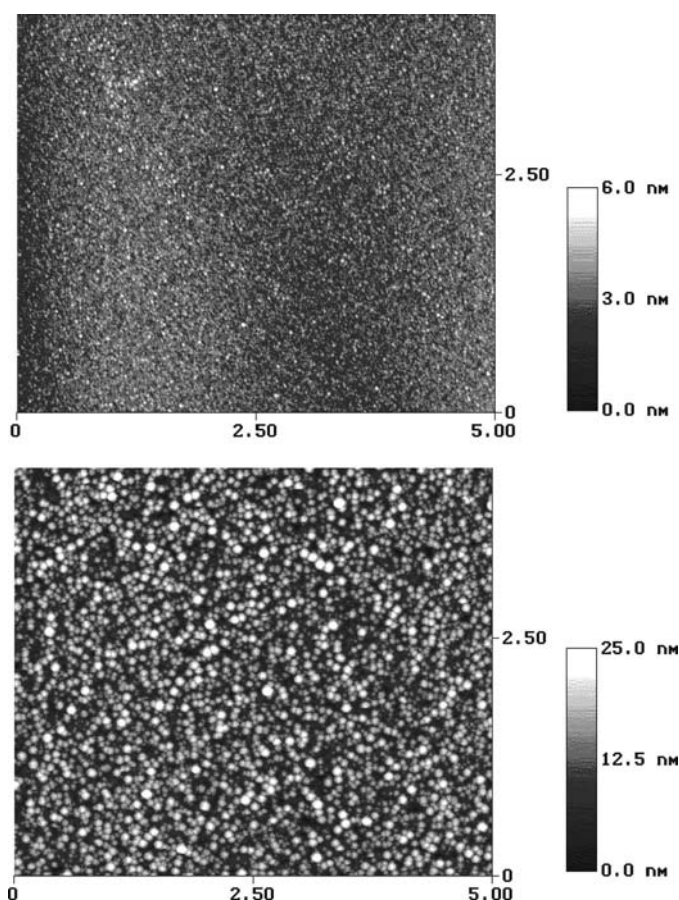


Figure 7 AFM images of gold nanocrystals fabricated by pulsed laser deposition, before annealing (upper) and after annealing at 400°C for 1 hour (lower).

resolution of these concerns will provide a deeper insight into optimization of the SERS signal, and thereby set the stage for establishing an improved protocol for fabricating SERS substrates.

Control of the nanocrystal size is achieved by varying the laser parameters, the annealing program, and the introduction of an inert gas (collisional cooling) for driving nucleation and nanocrystal growth during ablation.

The energy density threshold for ablation of gold as well as the energy density that leads to the so-called *splashing effects* have been established. As splashing is generally undesirable for synthesizing nanocrystals of uniform size, we have investigated energy density dependence on nanocrystal size above the ablation threshold and below the splashing threshold. A bimodal particle size distribution was achieved by pulsed laser deposition a layer of nanocrystals of a size determined by the energy density followed by a subsequent layer of another size, also determined by the energy density.

We have shown very good control of nanocrystal size by performing ablation in the presence of an inert gas for synthesizing silicon quantum dots using pressures in the range of 0.01 to 0.1 torr. At these pressures, we are able to control the nanocrystal size from 2 nm (0.01 torr) to 8 nm (0.1 torr). This approach can be used for fabricating gold nanocrystals and should produce the same degree of control for nanocrystals size, but the appropriate gas pressures will have to be identified.

We have also taken an approach for fabricating SERS substrates that exercises control over the aggregate size (which is defined as the number of nanocrystals in an aggregate). The motivation behind these investigations is to ultimately determine how the SERS effect depends on the aggregate size as well as to establish the aggregate size, which renders the largest enhancement. To control the aggregate size, we polish an MgO(100) oriented crystal 1° off-axis. This produces a highly stepped surface that when annealed at 1200°C causes the formation of islands ranging in size 0.3 to $1.5\ \mu\text{m}$. Subsequent electron-beam evaporation of gold onto the island-decorated MgO surface followed by annealing at 900°C leads to the selective growth of nanocrystals on the island surfaces, as shown in Figure 8. The AFM images in this figure reveal that gold nanocrystals show little variation in size and have an average diameter of 80 nm; furthermore, there is no evidence indicating the presence of gold nanocrystals in the domains between the islands. Consequently, these substrates will only exhibit the SERS effect when probe molecules are situated on the surface of the gold nanocrystal-decorated islands. Moreover, the inter-island distances are in many cases sufficiently large that a Raman microscope could interrogate individual islands for the SERS effect. These studies would then allow evaluation of how aggregate size affects the enhancement of the

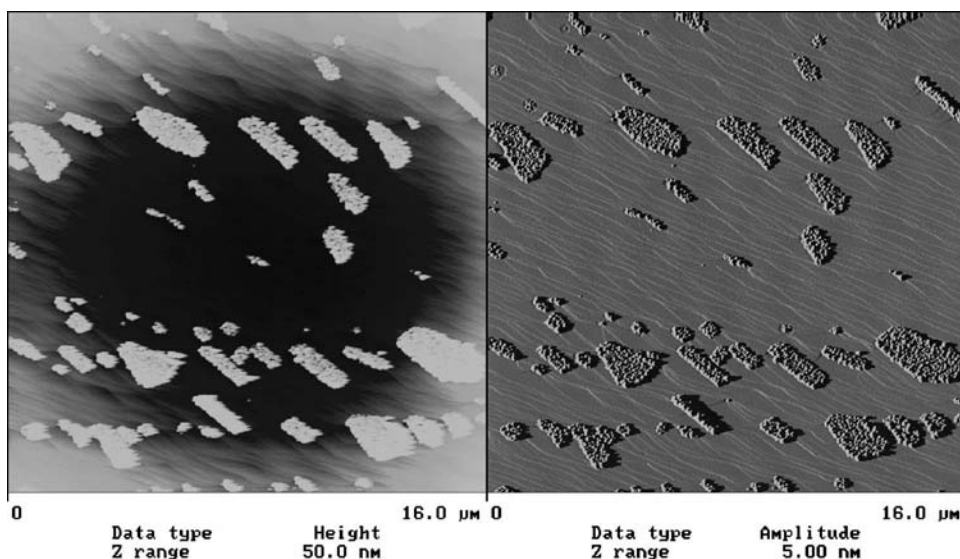


Figure 8 AFM images of gold nanocrystals on an MgO (100) surface. Left: Height-mode image. Right: Amplitude-mode image. A cross-section analysis indicates that the nanocrystals are typically 80 nm in height.

Raman signal. Clearly, the results of such measurements would then serve as a basis for designing SERS substrates, which are optimized with respect to aggregate size.

5 Uniformization of optical roperties

Sometimes we want our beams of light to be highly monochromatic, highly polarized, and collimated extremely well. With lasers, such beams are easy to produce. They are the opposites of random beams. At other times, we want our light beams to be white (or uniformly polychromatic), unpolarized (really randomly polarized), and diffuse. These are quite random beams. We need diffusers to produce and control randomness.

Diffusers can be classified in many ways. They can be optically thin or thick, depending on whether the average ray undergoes one or many interactions with the diffuser *en route* to the observation region. They can be conventional or holographic. They can be natural or manufactured. The manufactured ones can be crude (of ground glass or opal glass, for instance) or optimized for a specific purpose.

Thin diffusers have properties—such as not depolarizing very much—that are occasionally useful. Almost any interaction of light with anything will change the polarization a little, with the following notable exception. Light normally incident on clear, homogeneous glass is not changed in polarization. But all other attributes of light are impacted by the Fresnel equations and many other factors. Multiple scattering depolarizes by changing the ray's direction many times, taking the associated electromagnetic field vectors along for the ride. Sometimes, depolarization is the desired effect.

Diffusers can mix wavelengths. If you grind up bright red glass and put it in a clear glass, it will no longer look red. It will be a pale pink. The diffusion mixes in the white room light or sunlight. If you add water to the glass to reduce the scatter, it will look pink again. A very powerful tunable spectral filter—called a Christiansen filter—can be built to exploit this effect. The basic idea is 115 years old [61]. The diffusion depends strongly on both temperature and wavelength, and is thus quite tunable.

Diffusers always redirect light. Thin diffusers redirect light about the specular rays. The ideal thick diffuser does not have memory. That is, the scatter pattern is the same regardless of the direction of the input light, a phenomenon called Lambertian scatter. Everything in between is possible as well.

We have been most concerned with designer diffusers. It turns out that the best way to make them is through randomness. We record speckle patterns, which are actually the effects of random scattering. By controlling the expected size and shape of the speckles, we control the size and shape of the scattering pattern. By controlling the phase modulation these essentially binary patterns produce, we control the degree of polarization from 0 to 100%. This was proved many years ago [62]. With modern manufacturing methods, these custom diffusers have become cheap and reliable.

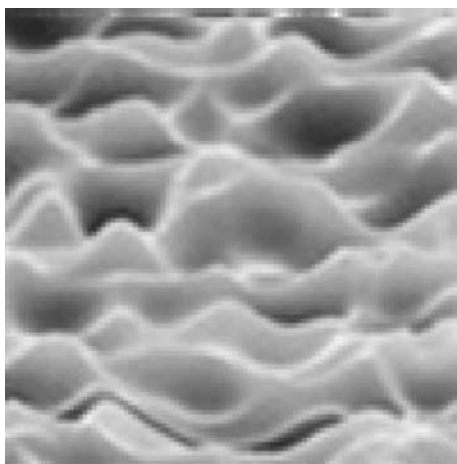


Figure 9 SEM image of a 60° full-angle transmissive holographic shaping diffuser with random but controlled morphology, at 1500 × magnification. [Courtesy of Physical Optics Corporation.]

Figure 9 shows a scanning electron microscope (SEM) image of a diffuser made by Physical Optics Corporation based on this concept. Of more interest perhaps are pictures of what such diffusers can accomplish, as illustrated in Figure 10.

6 Conclusion

Albert Einstein was wrong when he stated, “I cannot believe that God would choose to play dice with the universe.” Dice are key to many things, from how your brain functions to materials. God may not use them, but we can and should.

Randomness can be a friend to the makers and users of complex mediums. But, it needs to be controlled and directed to our needs by either self-organization or clever manufacture. The diverse examples we showed in this chapter are merely some of our favorites. You may well have others, and we would enjoy hearing from you about them.

The goal of this chapter, however, is not to gather random examples of random effects on optical materials. Rather, it is to suggest that there is a pattern here that, once understood, can be employed in many ways we have not yet thought of. We want to commend randomness as a tool for complex mediums.

Acknowledgments

The authors thank Alexander L. Burin and Natalia Noginova for useful discussions. M.A.N.’s work was supported by NASA grant #NCC-1-01049 and NSF CREST project #HRD-9805059.

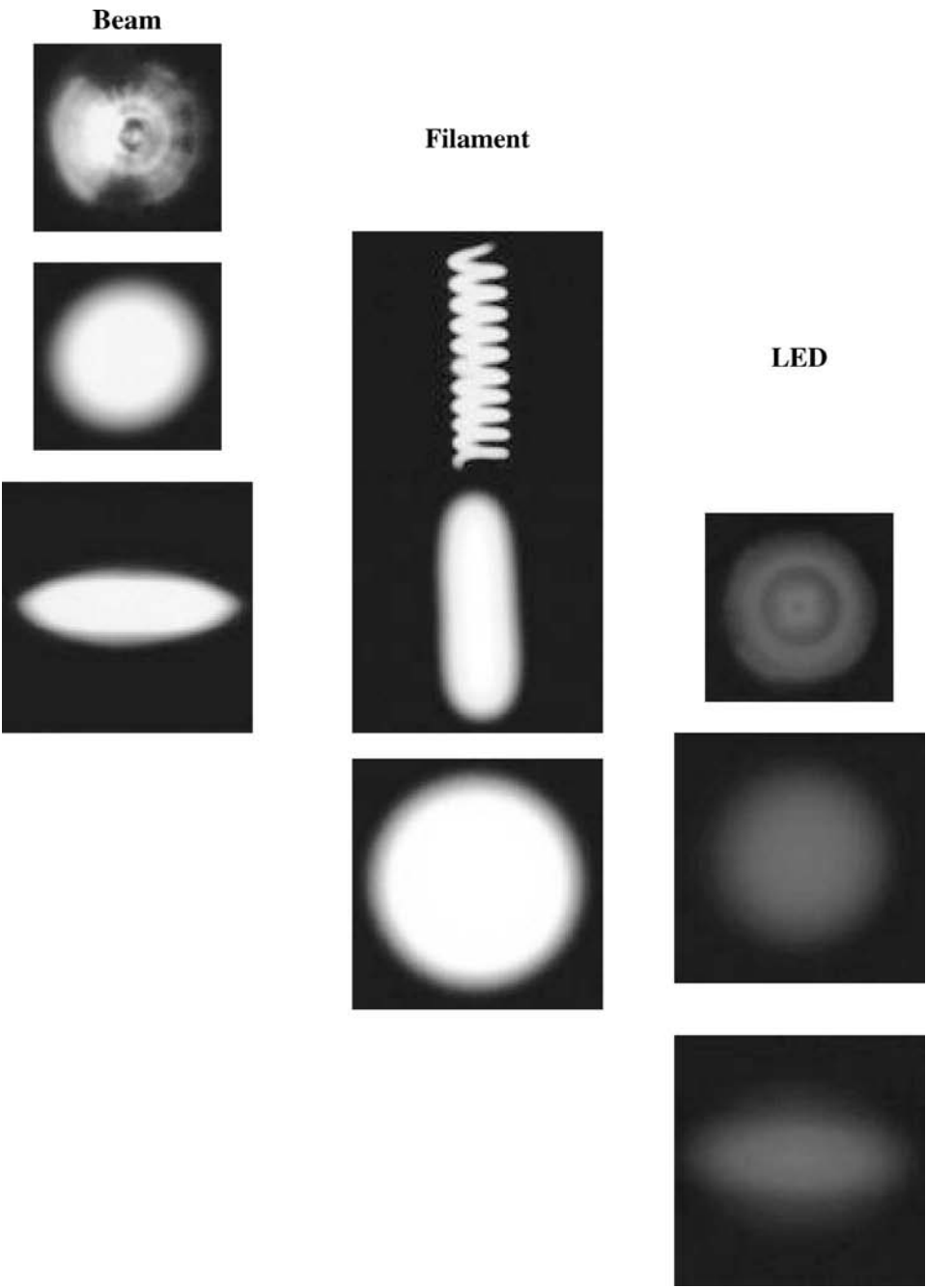


Figure 10 Each of these columns shows the prediffuser beam on top and the effects of particular light-shaping diffusers below—one circularly scattering and the other elliptically scattering. [Courtesy of Physical Optics Corporation.]

References

- [1] B. Ando and S. Graziani (eds), *Stochastic resonance: theory and applications*, Kluwer, Dordrecht, The Netherlands, 2000.
- [2] I. Prigogine and I. Stengers, *Order out of chaos: man's new dialogue with nature*, Bantam, New York, NY, USA, 1989.
- [3] Rand Corporation, *A million random digits with 100,000 normal deviates*, Free Press, Glencoe, IL, USA, 1955.
- [4] C.P. Robert and G. Casella, *Monte Carlo statistical methods*, Springer, New York, NY, USA, 1999.
- [5] I. Bilinskis, A. Mikelsons and A.K. Mikelson, *Randomized signal processing*, Prentice-Hall, New York, NY, USA, 1992.
- [6] J. Shamir, H.J. Caulfield and B.M. Hendrickson, Wavefront conjugation and amplification for optical communication through distorting media, *Appl Opt* **27** (1998), 2912–2914.
- [7] H.J. Caulfield, Q.B. He, M.P. Schamschula and J. Shamir, Reduced coherence, stability, and temporal overlap requirements for holography, *Appl Opt* **30** (1991), 4170–4171.
- [8] H.J. Caulfield, N. Kukhtarev, T. Kukhtareva, M.P. Schamschula and P. Banarjee, One-, two-, and three-beam optical chaos and self-organization effects in photorefractive materials, *Mater Res Innovat* **2** (1999), 194–199.
- [9] A.M. Prokhorov, Molecular amplifier and generator for submillimeter waves, *Sov Phys JETP* **7** (1958), 1140–1141.
- [10] A.L. Schawlow and C.H. Townes, Infrared and optical masers, *Phys Rev* **112** (1958), 1940–1949.
- [11] R.V. Ambartsumyan, N.G. Basov, P.G. Kryukov and V.S. Letokhov, A laser with a nonresonant feedback, *IEEE J Quant Electron* **2** (1966), 442–446.
- [12] R.V. Ambartsumyan, P.G. Kryukov and V.S. Letokhov, Dynamics of emission line narrowing for a laser with nonresonant feedback, *Sov Phys JETP* **24** (1967), 1129–1134.
- [13] V.S. Letokhov, Stimulated emission of an ensemble of scattering particles with negative absorption, *JETP Lett* **5** (1967), 212–215.
- [14] V.S. Letokhov, Generation of light by a scattering medium with negative resonance absorption, *Sov Phys JETP* **26** (1968), 835–840.
- [15] O. Svelto, *Principles of lasers*, 4th ed, Plenum Press, New York, NY, USA, 1998.
- [16] A. Yariv, *Quantum electronics*, 3rd ed, Wiley, New York, NY, USA, 1989.
- [17] W. Koechner, *Solid-state laser engineering*, 5th ed, Springer, Berlin, Germany, 1999.
- [18] V.M. Markushev, V.F. Zolin and Ch.M. Briskina, Luminescence and stimulated emission of neodymium in sodium lanthanum molybdate powders, *Sov J Quant Electron* **16** (1986), 281–283.
- [19] V.M. Markushev, N.É. Ter-Gabriélyan, Ch.M. Briskina, V.R. Belan and V.F. Zolin, Stimulated emission kinetics of neodymium powder lasers, *Sov J Quant Electron* **20** (1990), 772–777.

- [20] N.È. Ter-Gabriélyan, V.M. Markushev, V.R. Belan, Ch.M. Briskina and V.F. Zolin, Stimulated emission spectra of powders of double sodium and lanthanum tetramolybdate, *Sov J Quant Electron* **21** (1991), 32–33.
- [21] N.È. Ter-Gabriélyan, V.M. Markushev, V.R. Belan, Ch.M. Briskina, O.V. Dimitrova, V.F. Zolin and A.V. Lavrov, Stimulated radiation emitted by lithium neodymium tertaphosphate $\text{LiNd}(\text{PO}_3)_4$ and neodymium pentaphosphate $\text{NdP}_5\text{O}_{14}$ powders, *Sov J Quant Electron* **21** (1991), 840–841.
- [22] C. Gouedard, D. Husson, C. Sauteret, F. Auzel and A. Migus, Generation of spatially incoherent short pulses in laser-pumped neodymium stoichiometric crystals and powders, *J Opt Soc Am B* **10** (1993), 2358–2363.
- [23] M.A. Noginov, N.E. Noginova, H.J. Caulfield, P. Venkateswarlu, T. Thompson, M. Mahdi and V. Ostroumov, Short-pulsed stimulated emission in the powders of $\text{NdAl}_3(\text{BO}_3)_4$, $\text{NdSc}_3(\text{BO}_3)_4$, and $\text{Nd}:\text{Sr}_5(\text{PO}_4)_3\text{F}$ laser crystals, *J Opt Soc Am B* **13** (1996), 2024–2033.
- [24] Ch.M. Briskina, V.M. Markushev and N.È. Ter-Gabriélyan, Use of a model of coupled microcavities in the interpretation of experiments on powder lasers, *Quant Electron* **26** (1996), 923–927.
- [25] V.F. Zolin, The nature of plaser-powdered laser, *J Alloys Compounds* **300-301** (2000), 214–217.
- [26] F. Auzel and P. Goldner, Coherent light sources with powder: stimulated amplification versus super-radiance, *J Alloys Compounds* **300-301** (2000), 11–17.
- [27] M. Bahoura, K.J. Morris and M.A. Noginov, Threshold and slope efficiency of $\text{Nd}_{0.5}\text{La}_{0.5}\text{Al}_3(\text{BO}_3)_4$ ceramic random laser: effect of the pumped spot size, *Opt Commun* **201** (2002), 405–412.
- [28] M.A. Noginov, N. Noginova, S.U. Egarievwe, H.J. Caulfield, C. Cochrane, J.C. Wang, M.R. Kokta and J. Paitz, Study of the pumping regimes in Tisapphire and $\text{Nd}_{0.5}\text{La}_{0.5}\text{Al}_3(\text{BO}_3)_4$ powders, *Opt Mater* **10** (1998), 297–303.
- [29] D.S. Wiersma and A. Lagendijk, Light diffusion with gain and random lasers, *Phys Rev E* **54** (1996), 4256–4265.
- [30] M.A. Noginov, S.U. Egarievwe, N. Noginova, J.C. Wang and H.J. Caulfield, Demonstration of a second harmonic powder laser, *J Opt Soc Am B* **15** (1998), 2854–2860.
- [31] G. van Soest, F.J. Poelwijk, R. Sprik and A. Lagendijk, Dynamics of a random laser above threshold, *Phys Rev Lett* **86** (2001), 1522–1525.
- [32] M.A. Noginov, S.U. Egarievwe, N. Noginova, H.J. Caulfield and J.C. Wang, Interferometric studies in a powder laser, *Opt Mater* **12** (1999), 127–134.
- [33] A.A. Lichmanov, Ch.M. Briskina, V.M. Markushev, V.N. Lichmanova and N.P. Soshchin, Degree of coherence and dimensions of the generation region of powder lasers, *J Appl Spectrosc* **65** (1998), 818–825.
- [34] A.A. Lichmanov, Ch.M. Briskina, N.P. Soshchin and V. F. Zolin, Lasing in powders and its use for data processing, *Bull Russ Acad Sci Phys* **63** (1999), 922–926.

- [35] A.A. Lichmanov, Ch.M. Brickina, V.N. Lichmanova, N.P. Soshchin and V.F. Zolin, Experimental studies of the lanthanide doped lasing powders (plasers), *Proc LASERS'98*, Tucson, AZ, USA, 1999, 725–731.
- [36] S. John and G. Pang, Theory of lasing in a multiple-scattering medium, *Phys Rev A* **54** (1996), 3642–3652.
- [37] M.A. Noginov, N.E. Noginova, S.U. Egarievwe, H.J. Caulfield, P. Venkateswarlu, A. Williams and S.B. Mirov, Color-center powder laser: the effect of pulverization on color-center characteristics, *J Opt Soc Am B* **14** (1997), 2153–2160.
- [38] G. Williams, B. Bayram, S.C. Rand, T. Hinklin and R.M. Laine, Laser action in strongly scattering rare-earth-doped dielectric nanophosphors, *Phys Rev A* **65** (2001), 013807.
- [39] R. Bonifacio and L.A. Lugiato, Cooperative radiation processes in two-level systems: superfluorescence, *Phys Rev A* **11** (1975), 1507–1521.
- [40] N.M. Lawandy, R.M. Balachandran, A.S.L. Gomes and E. Sauvain, Laser action in strongly scattering medium, *Nature* **368** (1994), 436–438.
- [41] H. Cao, Y.G. Zhao, H.C. Ong, S.T. Ho, J.Y. Dai, J.Y. Wu and R.P.H. Chang, Ultraviolet lasing in resonators formed by scattering in semiconductor polycrystalline films, *Appl Phys Lett* **73** (1998), 3656–3658.
- [42] H. Cao, Y.G. Zhao, S.T. Ho, E.W. Seelig, Q.H. Wang and R.P.H. Chang, Random laser action in semiconductor powder, *Phys Rev Lett* **82** (1999), 2278–2281.
- [43] H. Cao, Y.G. Zhao, X. Liu, E.W. Seelig and R.P.H. Chang, Effect of external feedback on lasing in random media, *Appl Phys Lett* **75** (1999), 1213–1215.
- [44] H. Cao, J.Y. Xu, E.W. Seeling and R.P.H. Chang, Microlaser made of disordered media, *Appl Phys Lett* **76** (2000), 2997–2999.
- [45] H. Cao, J.Y. Xu, D.Z. Zhang, S.-H. Chan, S.T. Ho, E.W. Seelig, X. Liu and R.P.H. Chang, Spatial confinement of laser light in active random media, *Phys Rev Lett* **84** (2000), 5584–5587.
- [46] H. Cao, Y. Ling and C.Q. Cao, Photon statistics of random lasers with resonant feedback, *Phys Rev Lett* **86** (2001), 4524–4527.
- [47] H. Cao, J.Y. Xu, Y. Ling, S.-H. Chang, S.T. Ho, E.W. Seelig, X. Liu and R.P.H. Chang, Random lasers with coherent feedback, *Photonic crystals and light localization in the 21st century* (C.M. Soukoulis, ed), Kluwer, Dordrecht, The Netherlands, 2001.
- [48] H. Cao, Random lasers with coherent feedback, *Optical properties of nanostructured random media* (V.M. Shalaev, ed), Springer, New York, NY, USA, 2002.
- [49] R.C. Polson, J.D. Huang and Z.V. Vardeny, Random lasers in π -conjugated polymer films, *Syn Metals* **119** (2001), 7–12.
- [50] R.C. Polson, A. Chipouline and Z.V. Vardeny, Random lasing in p-conjugated films and infiltrated opals, *Adv Mater* **13** (2001), 760–764.
- [51] R.C. Polson, J.D. Huang and Z.V. Vardeny, Analysis of random lasers in thin films of π -conjugated polymers, *Photonic crystals and light localization in*

- the 21st century* (C.M. Soukoulis, ed), Kluwer, Dordrecht, The Netherlands, 2001.
- [52] V.A. Nikitenko, A.I. Tereschenko, I.P. Kuz'mina and A.N. Lobachev, Stimulated emission of ZnO at high level of single photon excitation (in Russian), *Optika i Spektroskopiya* **50** (1981), 605–607.
 - [53] P.C. de Oliveira, J.A. McGreevy and N.M. Lawandy, Speckle-mirror laser, *Opt Lett* **22** (1997), 700–702.
 - [54] A.L. Burin, M.A. Ratner, H. Cao and R.P.H. Chang, Model for a random laser, *Phys Rev Lett* **87** (2001), 215503.
 - [55] X. Jiang and C.M. Soukoulis, Time-dependent theory for random lasers, *Phys Rev Lett* **85** (2000), 70–73.
 - [56] Q. Li, K.M. Ho and C.M. Soukoulis, Mode distribution in coherently amplifying laser medium, *Physica B* **296** (2001), 78–84.
 - [57] C.M. Soukoulis, X. Jiang, J.Y. Xu and H. Cao, Dynamic response and relaxation oscillations in random lasers, *Phys Rev B* **65** (2002), 041103.
 - [58] X. Jiang and C.M. Soukoulis, Theory and simulations of random lasers, *Photonic crystals and light localization in the 21st century* (C.M. Soukoulis, ed), Kluwer, Dordrecht, The Netherlands, 2001.
 - [59] C. Vanneste and P. Sebbah, Selective excitation of localized modes in active random media, *Phys Rev Lett* **87** (2001), 183903.
 - [60] M. Kerker (ed), *Selected papers on surface enhanced Raman scattering*, SPIE Press, Bellingham, WA, USA, 1999.
 - [61] M. Cloupeau and S. Klarsfeld, Visualization of thermal fields in saturated porous media by the Christiansen effect, *Appl Opt* **12** (1973), 198–199.
 - [62] H.J. Caulfield, Optically generated kinoforms, *Opt Commun* **4** (1971), 201–202.



H. John Caulfield is the Distinguished Research Professor at Fisk University and holds several other academic and industrial positions. He has published 10 books, 32 book chapters, 6 conference proceedings, more than 200 refereed journal papers, 30 patents, and a number of popular articles. He was Editor of *Optical Engineering* for six years and has in the past or does now serve on the editorial boards of a dozen other journals. He has been mentioned and sometimes featured in every newsstand magazine on business, science, and news as well as two airline magazines. He is a Fellow

of SPIE (The International Society for Optical Engineering) and the Optical Society of America. He has won more prizes and medals from SPIE than anyone ever (Fellow, President's Award, Governor's Award, Dennis Gabor Award).

Donald O. Henderson received his B.S. in Chemistry from Eastern Oregon State University, La Grande, OR, USA, in 1979; and his Ph.D. in Physical Chemistry

from Texas Technological University, Lubbock, TX, USA, in 1987. He was a post-doctoral associate in the Chemistry Department at Vanderbilt University, Nashville, TN, USA from 1985 to 1987. He joined the Physics Department at Fisk University, Nashville, TN, USA in 1987 as a Research Associate and is currently a Research Professor of Physics and an Associate Professor of Chemistry. He works chiefly on interfacial phenomena, and the application of ion beams for materials modification and synthesis. He has more than 170 publications in refereed journals and presentations and has given several invited lectures at National Laboratories, Universities and International meetings. He currently serves on the Editorial Advisory Board of *Materials Science Foundations*. He is also an Adjunct Professor of Physics at Vanderbilt University.



Mikhail A. Noginov obtained his M.S. degree in Electronics Engineering from the Moscow Institute for Physics and Technology, and his Ph.D. degree in physics from the General Physics Institute of the Russian (USSR) Academy of Sciences. He then continued his training as a Research Associate at Massachusetts Institute of Technology, Cambridge, MA, USA. Dr. Noginov is currently an Associate Professor in the Department of Physics at Norfolk State University, Norfolk, VA, USA. He has coauthored one book chapter, nearly 70 journal publications, more than 30 proceedings publications and nearly 100 conference presentations. Dr. Noginov is a reviewer and panel member of several journals and funding agencies.

Nonlinear Spatial Structures

William J. Firth and John M. McSloy

Abstract

Spontaneous spatial patterns occur in nonlinear systems with spatial coupling, e.g., through diffraction or diffusion. Strong enough nonlinearity can induce breaking of spatial symmetry, such that a pattern becomes more stable than the original featureless state. Instances discussed are in nonlinear optics, but the phenomena have a universal character and are the basis of spatial differentiation in nature, from crystals to clouds, from giraffe coats to galaxies. The basic theory and phenomena of pattern formation are reviewed, with examples from experiments and simulations (mainly from optics). Patterns usually consist of repeated units, and such units may exist in isolation as localized structures. Such structures are akin to spatial solitons and are potentially useful in image and/or information processing. The nature and properties of such structures are discussed and illustrated.

1 General introduction

Patterns pervade both the natural world around us and our synthetic human environment; see Figure 1. This printed page is an example, being an array of dark objects (letters) distributed quasi-regularly on a white page. It is an interesting example because it illustrates a number of points that will be important later in this chapter. First, irregularity is essential to the conveying of information by the pattern of words and letters on the printed page: a perfectly regular sequence such as aaaaaaaaaaaaaaaaaaaaaaaaaa conveys little or no information. Thus, while perfectly regular patterns are of great fundamental importance and interest, storage and transmission of information requires patterns with controlled irregularity. For example, simply allowing any “a” in the foregoing sequence to be absent gives a class of binary sequences capable (if long enough) of encoding the entire works of Shakespeare through, e.g., ASCII (or Morse) code. Second, the pattern on this page is imposed, not spontaneous. It is formed by applying ink to the paper in a prescribed arrangement. In contrast, the pattern of coloration on an animal’s coat is not prescribed—a cat’s DNA may require it to be black-and-white, but does not control the precise layout of the black and white patches (as cloning experiments illustrate [1]). Spontaneous patterns arise when the dynamics of a system allows two or more stable states for the same external conditions (e.g., black or white). This requires that the system’s dynamics be nonlinear [2]. Some kind of spatial coupling is also needed, so that these states can compete for territory, just as ice and water compete for the Arctic Ocean, or vegetation and bare soil in arid regions [3].

These ideas are developed here, primarily in an optics context. The phenomena have a universal character, however, and so have the mathematical methods used to analyze and model them. An outline of the most basic models is presented next, both to establish notation and concepts and as a convenient summary for readers not familiar with dynamical systems theory or pattern formation [4].

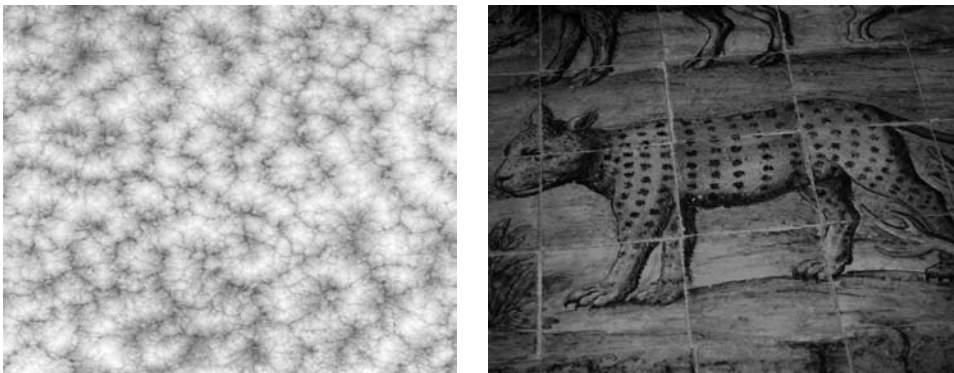


Figure 1 Left: Closed small-cell marine clouds, where the darker regions are indicative of sinking air motions. Right: An artist's impression of a leopard's coat pattern (Capri, Ch. San. Michele).

1.1 Basic models

To quantify the foregoing, and to illustrate the necessity for nonlinearity and spatial coupling, let us consider some very simple mathematical models. Suppose we have some system variable u which has some dynamics. About the simplest case we could envisage is where u obeys

$$\frac{\partial u}{\partial t} = \alpha u. \quad (1)$$

This dynamics is trivial: u grows exponentially if $\alpha > 0$, and decays exponentially to zero if $\alpha < 0$. The state $u = 0$ is its only fixed point, i.e., steady state.

Let us extend it by adding a driving term:

$$\frac{\partial u}{\partial t} = \alpha u + D. \quad (2)$$

Now there is an obvious nontrivial fixed point $u = -D/\alpha$. If D varies in space (e.g., represents a pattern), then u develops exactly the same dependence. In other words, this is a model for imposing a pattern on a system. Fine, but there are problems. What about stability? We easily find that any perturbation to the steady state grows exponentially if $\alpha > 0$, and so the steady state is unstable. If $\alpha < 0$, however, perturbations decay, and the steady state is stable. Good, but as soon as we cease to present the pattern D , the system reverts to the model (1) and the pattern disappears; the system has no memory.

Another ingredient is needed for that: nonlinearity [2]. Consider

$$\frac{\partial u}{\partial t} = \alpha u - \gamma u^3. \quad (3)$$

This system is just like the linear equation (1) if u is small, but for larger amplitudes the cubic term becomes important. It is a simple version of the Ginzburg–Landau equation [4]. As well as $u = 0$, (3) has the additional fixed points $u = \pm\sqrt{\alpha/\gamma}$ (provided α and γ have the same sign). What about their stability? From the foregoing, it is clear that $u = 0$ is stable if $\alpha < 0$, and unstable if $\alpha > 0$. To test the stability of the other fixed points, we linearize (3) close to them, writing $u = \pm\sqrt{\alpha/\gamma} + \delta$, substituting into (3), and neglecting all terms of quadratic or higher order in δ . Hence, this reduces to

$$\frac{\partial \delta}{\partial t} = -2\alpha\delta, \quad (4)$$

from which it is evident that both nontrivial fixed points are stable if $\alpha > 0$, and unstable if $\alpha < 0$.

Let us in the meantime assume that γ is positive; then there is a single stable fixed point at $u = 0$ if α is negative, and three if positive, of which $u = 0$ is unstable

and the other two are stable. Furthermore, it is easy to show that u can never change sign under the dynamics (3), and that any initial positive value of u is attracted to $u = \sqrt{\alpha/\gamma}$, and any initial negative value to $-\sqrt{\alpha/\gamma}$. Figure 2(a) shows the fixed points and their stability in a standard graph for such nonlinear systems.

When the parameter $\alpha = 0$, this is called a *bifurcation* point, because the character of the dynamics changes there through the appearance of new solutions. Physically, a bifurcation point is usually associated with some kind of threshold. This particular bifurcation is called a pitchfork, because of its shape; see Figure 2(a). Given this dynamics, we can try to impose a pattern on this system by initializing (3) with some desired pattern $u(r) = P(r)$. The result will be $u = \sqrt{\alpha/\gamma}$, wherever $P > 0$, and $u = -\sqrt{\alpha/\gamma}$, wherever $P < 0$. The system (3) is, in effect, a one-bit digitizing memory—as will be obvious to any reader familiar with the dynamics of bistable systems, of which (3) is an archetype.

The case of γ negative needs a mention, because it is important later on. In that case the pitchfork is reversed [Figure 2(b)], because the non-zero fixed points exist only for $\alpha < 0$. Their stability is still governed by (4), and so they are unstable. The related dynamics is also indicated in Figure 4(b): whenever u lies “outside” the parabola, it diverges with time. Model (3) thus cannot describe the global dynamics of the system in this case: one must introduce additional nonlinear terms—quintic or higher—to allow additional, stable, fixed points with larger $|u|$. A bifurcation such as this [Figure 3(b)], where the new solutions coexist with the existing one(s) below the bifurcation point, is said to be subcritical; while one like Figure 3(a), where they exist only above the bifurcation, is supercritical. It is evident from Figure 2 that the new solution grows smoothly out of the old at a supercritical

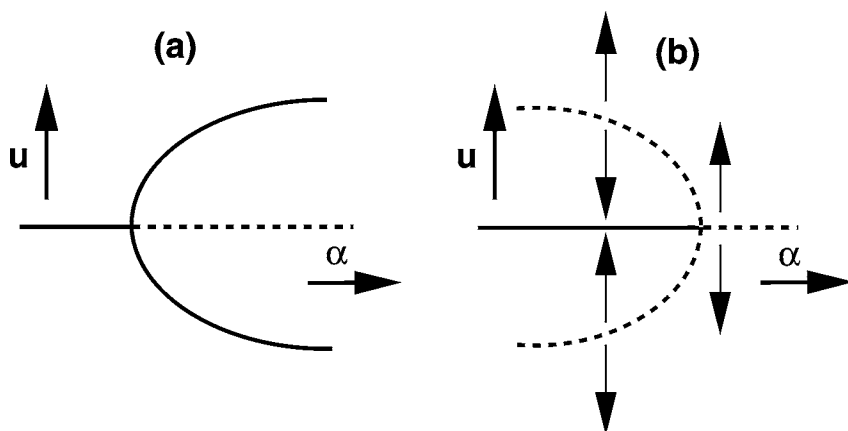


Figure 2 Bifurcation diagrams for fixed points of the Ginzburg–Landau equation (3): (a) $\gamma > 0$; the zero solution becomes unstable at $\alpha = 0$, and two new stable solutions appear; (b) $\gamma < 0$; the zero solution again becomes unstable at $\alpha = 0$, but the new solution curves bend backwards, so that there is no stable solution for $\alpha > 0$. (Conventionally, stable states are graphed with a solid line, unstable ones with a broken one.) Also shown in (b) are arrows indicating the evolution of u : if u lies “inside” the parabola, it evolves to the stable zero solution, but if outside, it diverges.

bifurcation, but that a subcritical bifurcation involves a discontinuous switching of the output state. In phase transition parlance, they are second-order and first-order transitions, respectively [4].

1.2 Basics of pattern formation

Spontaneous pattern formation requires spatial coupling as well as nonlinearity, and so we now develop our simple models appropriately. In imposing a pattern on the system (3), positive and negative initial values are attracted to the respective positive and negative fixed points, which are quite distinct. Thus, a sine wave $u(t=0) = a \sin Kx$ should evolve into a square wave, with discontinuities at the locations where $\sin Kx = 0$. This is physically unreasonable, and one might expect that in a real system the singularity would be smoothed by diffusion (or perhaps diffraction). Inclusion of diffusion in nonlinear equations like (3) does indeed remove such singularities, but usually also results in a dynamical competition between the different solutions. In general, one solution will be more stable than the other, and will invade and annihilate it. For example, above the freezing point water will invade ice, and *vice versa* below. The underlying reason is that diffusion is wave-vector selective, favoring small K over large, and $K = 0$ (no pattern) above all.

Pattern formation should also be wavevector-selective, but should favor finite K , corresponding to the pattern period. Consider the following modified version of (3):

$$\frac{\partial u}{\partial t} = \alpha u - \gamma u^3 - (\nabla^2 + k_c^2)u. \quad (5)$$

Before commenting on the structure, let us try a solution of the form $u = a(t) \cos Kx$, and assume that $|a|$ remains small enough that the nonlinear term can be neglected. Substitution yields

$$\frac{\partial a}{\partial t} = \alpha a - (-K^2 + k_c^2)a. \quad (6)$$

This equation has the same form as (1), and we can immediately conclude that $a = 0$ is the only stable solution for $\alpha < 0$, but that as α becomes positive, wave patterns close enough to $|K| = k_c$ become excited. This is exactly what we would expect to happen in a system showing spontaneous pattern formation. Below a certain threshold there is no pattern but, on increasing a suitable parameter, structure appears, concentrated around a particular spatial period.

Equation (2) is a form of Swift–Hohenberg equation [4] and is a generic model for spontaneous pattern formation, in the sense that many different physical scenarios (such as convection) can be converted into Swift–Hohenberg form when close to the onset of pattern formation [4]. How does the pattern actually evolve? Is there a stable steady-state pattern, and, if so, what is its amplitude? These are interesting

questions, but they cannot be exactly answered because (6), unlike (7), is nonlinear and in general cannot be solved analytically. The cubic term will give rise to spatial harmonics, leading to an infinite set of coupled equations for the spatial Fourier transform of u [8]. That said, (7) shows that all wavevectors far from $|K| = k_c$ are damped, and so the evolution will be dominated by K 's close to k_c (which is called the critical wavevector). The bifurcation diagram for the amplitude of the pattern will therefore look very much like those in Figure 2.

In particular, pattern formation can be supercritical or subcritical. Again, the latter case requires higher-order nonlinearity to enable a stable patterned solution. The spatial-coupling term in (6) was plucked “out of the air” but is pretty much forced by typical symmetry requirements. The Laplacian operator is translation- and rotation-symmetric, which means that the resulting patterns have no preferred orientation or phase (i.e., peaks and troughs can be located arbitrarily). The overall square on the operator ensures that both larger and smaller wavevectors than k_c are damped. The particular value of k_c is determined by the underlying physics, but it can be seen that the Swift–Hohenberg operator is about the simplest with this high symmetry and wavevector-selectivity. The rotational invariance of (6) means that only the magnitude, not the direction, of the wavevector of the evolving pattern is determined by the system. It follows that, in two or more dimensions, there can be competition between different orientations of the evolving patterns. The winner of such a competition can be determined by a slight environmental bias, e.g., the water-depth effect that ensures that ocean waves end up parallel to the beach. There can also be competition between different types of patterns, one of the commonest being between stripes (rolls) and hexagons (spots). Again, the competition is usually close, which is why there are both striped and spotted cats—and indeed why a spotted cat like the cheetah actually has a striped tail.

To summarize: the formation and evolution of regular spatial patterns is an interesting feature of many branches of science, including biology, chemistry, hydrodynamics and optics. Spontaneous formation of spatial patterns implies a spontaneous breaking of the symmetry of the system, which in turn implies the coexistence of several—or many—stable output states for a given input. Thus, complex patterns are indicative of potential for image recognition and data processing capacity, an idea that is discussed later in this chapter.

2 Pattern formation in nonlinear optics

Studies of optical systems increasingly involve simulations of their nonlinear dynamics in the space-time domain, a development that is essential to describe real, as opposed to idealized, systems. The significance is much wider, however, because the space-time dynamics of lasers and associated nonlinear optical systems is also important for applications. Technological demand for high-power laser beams of high spatial coherence and quality is raising questions as to how broad-area lasers can be controlled more subtly than by strangling their output with apertures. Miniaturized and integrated nonlinear optical systems involving all-solid-state lasers, es-

pecially semiconductor lasers, are making great progress. The low cost and great robustness of these systems are transforming and greatly expanding the range of laser applications [5]-[7]. But, this increasing complexity of nonlinear optoelectronic devices and systems raises serious questions of stability and functionality. Because monolithic devices have few of the manual adjustments on which tabletop lasers depend, these questions must be addressed at the design stage; and an understanding of the spatial structure and characteristics of the output, and of potential problems, assumes key importance.

Pattern formation can obviously be regarded as a problem in such systems—therefore needing to be understood in order to be eliminated. On the other hand, systems capable of pattern formation are usually also capable of supporting the modified patterns and localized structures with which information can be stored or processed. For these reasons, many research groups now focus their attention on systems where the spatial profile of the electromagnetic field assumes importance. Several relevant review articles (each with extensive bibliography) have been published over the last decade or so. Some of these are cited later in this chapter, and a subset of their topics are briefly reviewed. Images and diagrams are particularly useful in this field, and are liberally used to illustrate the presentation. Images produced by one's own group and collaborators are the most accessible, and so the models and illustrations presented are selected accordingly. Nevertheless, the universal character of pattern and structure formation means that the work presented here is representative of a much wider selection of research in the field.

Nonlinear optical systems with nontrivial spatial degrees of freedom (which excludes most fiber and waveguide systems) can be roughly divided into two classes: dissipative and Hamiltonian (conservative). The latter tend to involve propagation in bulk materials, and often exhibit self-organized spatial structures that may be loosely described as spatial solitons. They are not discussed here, but several edited books and review articles are available to the interested reader [9]-[12]. Some of these also include reviews dealing with dissipative systems, which are the main topic here.

As the name implies, dissipative systems are lossy. A nonlinear optical medium enclosed between mirrors is such a system, lossy due to output-coupling. A laser is the classic example, one whose threshold dynamics are actually very similar to those of the simple models described in Section 1. Laser patterns are not discussed here, however, and the interested reader is referred to the literature [13]-[15].

A very simple dissipative system, namely a single slice of nonlinear medium coupled to a single feedback mirror, has been shown theoretically [16-17] and experimentally (most comprehensively using sodium vapor [18]-[21]) to give rise to pattern formation when illuminated by a broad, smooth input beam. Figure 3 illustrates some recent results, in which control of experimental parameters (including the polarization of the optical field) allows a variety of regular patterns [18]-[20], and also quasi-patterns [21], to be observed. As is inevitable in real systems, the input beam has finite width and its intensity drops away from its central maximum. The patterns formed are, however, still quite well described by simpler

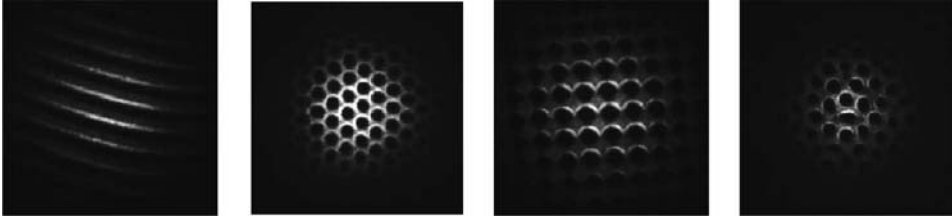


Figure 3 Examples of spontaneous optical patterns observed in sodium vapor with single feedback mirror: (left to right) rolls, hexagons [18]–[20], squares, and a quasi-pattern [21].

“planewave” models in which both the medium and the input field are assumed infinitely extended in the transverse plane.

Alternatively, one can consider a driven optical cavity containing a nonlinear medium, where the intracavity field is sustained against losses by an input driving field. (In a laser the loss is balanced, on average, by the gain in the laser medium.) Such systems are a main topic of this section. As an experimental example, patterns have been observed in reflection from a semiconductor microcavity composed of a GaAs multiple-quantum-well nonlinear medium sandwiched between two Bragg mirrors [22–23]. The small size (μm) and comparatively large bandwidth (>1 GHz) of semiconductor microresonators are of potential interest for applications, investigations of which were the goal of ESPRIT project PIANOS [24]. Modeling such structures requires one to couple the optical field to the photocarrier population and, in many cases, also to the local temperature [24].

For the purposes of this chapter, simpler models suffice. In one limit, the semiconductor nonlinearity is primarily dispersive; i.e., it behaves like a Kerr medium, in which the refractive index changes linearly with optical intensity. If the cavity finesse is high enough, a single longitudinal mode dominates the cavity field, and its amplitude E in response to a planewave input field obeys the following equation [25–26]:

$$\frac{\partial E}{\partial t} = -E - i\theta E + E_{in} + i\nabla^2 E + i|E|^2 E. \quad (7)$$

In comparison with (3) there are some similarities, but with a notable difference in that E is complex-valued. The terms on the right side of (7) respectively represent loss, mistuning (θ), driving, diffraction and the Kerr nonlinearity. It is easy to show that the homogeneous (planewave) steady-state solution E_s is three-valued for $\theta > \sqrt{3}$, corresponding to optical bistability. Less obvious, it was shown by Lugiato and Lefever [25] that the homogeneous solution is unstable to pattern formation whenever $|E_s| > 1$. In one transverse dimension the pattern-forming bifurcation is supercritical if $\theta < 41/30$, but subcritical for larger θ .

Another interesting limit of semiconductor nonlinearity is when it behaves as a saturable absorber (or amplifier). The simplest analogy to (7) is

$$\frac{\partial E}{\partial t} = -E - i\theta E + E_{in} + i\nabla^2 E - \frac{2C}{1 + |E|^2} E. \quad (8)$$

This differs from (7) only in the last term. $2C$ describes the density of the medium: if $|E|$ is small, it provides an additional loss, but this absorption gets bleached as $|E|$ increases. The pattern-forming behavior of this model has been extensively analyzed and simulated by our Strathclyde group [27–28]. In particular, if one writes $E = E_s(1 + A)$, a Swift–Hohenberg equation for the real part of A is found at threshold [27].

Figure 4 summarizes key results from these investigations [28], which involve quasi-exact methods of finding the fully nonlinear patterns of large amplitude. On the left is a bifurcation diagram for the critical wavevector k_c , which is precisely $\sqrt{-\theta}$ in this model [27]. The bifurcation is subcritical for these parameters, so that a pattern with this wavevector (or with a range of other k 's) coexists with a stable homogeneous solution down to about 75% of the threshold value of $I = |E_s|^2$. As mentioned before [see Figure 2(b)], a low-amplitude subcritical roll solution is always unstable (to growth or decay). At I_{sn} this branch bends around to form a high-amplitude roll solution. The existence and stability of the latter has been studied in detail over a wide range of pattern wavevectors [28], and the results are summarized in the right panel of Figure 4. As mentioned in the caption, there is a considerable domain over which a given roll pattern is stable “against all-comers.” This domain is known as the Busse Balloon [4]. Below it, the rolls are unstable to breaking up into spots, forming a hexagonal pattern. Another instability is found for $K < k_c$, in which region the wave vector is “too small for comfort.” The stripes can then develop a transverse instability, called the zigzag instability (Figure 5).

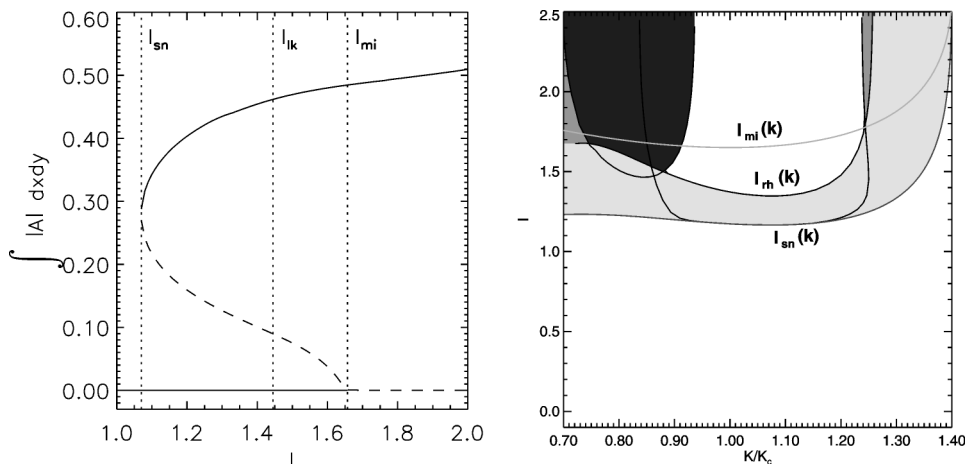


Figure 4 (Left) Bifurcation diagram for roll solution to (8) for $C = 5.4$, $\theta = -1.2$ as a function of $I = |E_s|^2$. The roll solution (with $K = k_c = \sqrt{-\theta}$) is characterized by the integral of $|A|$ (see text), and bifurcates subcritically for I about 1.65: (Right) Existence and stability of roll solutions to (8) as a function of wavevector. The rolls exist above $I_{sn}(K)$, but are unstable to hexagon formation below $I_{rh}(K)$. In the dark region to the left they are zigzag unstable (see Figure 5), in the small shaded regions they are Eckhaus unstable (see text), while in the unshaded domain they are absolutely stable. The curve $I_{mi}(K)$ marks the threshold value of I at which the homogeneous solution becomes unstable to a roll of wavevector K .



Figure 5 Dynamics of roll patterns in the model (8), confirming predictions of Figure 4: (left to right) zigzag instability; stable roll; roll-hexagon transition in early stages; roll-hexagon transition almost complete.

Or, if K is either too large or too small, extra stripes can appear or stripes can disappear, as appropriate. This is the Eckhaus instability [4].

Figure 5 shows a few examples from direct simulation of (8), confirming the predictions of the quasi-exact method used to generate Figure 4. On the left, the parameters are set in the zigzag unstable domain, and the panel shows the evolved zigzag pattern. The next panel confirms that rolls in the Busse Balloon are indeed stable. The two panels on the right are stages in the evolution of the roll-hexagon instability: this is a very typical dynamics in the field of pattern formation. The stripes break up, and clump into spots. Here the final pattern is a regular hexagon, but in the real world the hexagons will be imperfect to a greater or lesser degree. This transition involves a spontaneous breaking of the translational symmetry along the stripe direction, just as the instability by which the homogeneous solution forms stripes breaks its (perfect) translational symmetry. Both transitions are termed modulational instabilities (which explains the terminology I_{mi} in Figure 4).

A similar analysis has been carried out for hexagonal patterns [28], with broadly similar results, but is not elaborated here. Instead, note that Figure 4 demonstrates that (because of the subcritical bifurcation) there is a kind of bistability between a stable homogeneous solution and a stable roll pattern over a substantial range of I . What would happen at a mutual boundary between them? Usually one or other will “win,” invading and annihilating the other phase. In other words, the front between the two solutions will usually move spontaneously. Often, however, there is a parameter range over which the front “locks,” and the two states can stably coexist [29]. Pairs of such locked fronts can lock together, forming “islands” of pattern in a “sea” of the homogenous phase [30]. The smallest possible island is a single stripe or spot. If it is stable, then the aforementioned proposal to encode information using an irregular pattern becomes a realistic possibility. The following section explores that issue.

3 Solitonlike self-localized structures

This section outlines some ideas and results about self-localized spots, or clusters of spots, in nonlinear optical systems of the sort that support spontaneous pattern formation. Figure 6 shows the basic scenario, which works in both theory (top)

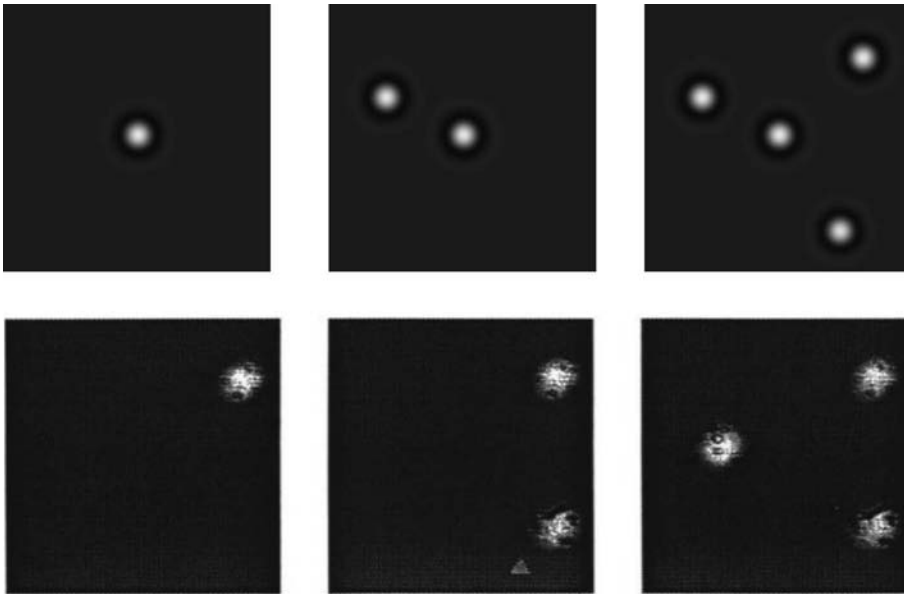


Figure 6 Cavity solitons in a saturable absorber resonator. (Top) Simulation in the model of (8) with $C = 5.4$ and $\theta = -1.2$. Successive localized address pulses write stable 2D cavity solitons in the transverse cross-section of the resonator (panel shows the transmitted intensity). (Bottom) Experimental realization using aberchrome in a cavity driven by a doubled Nd:YAG laser, addressed by pulses from an independent Ar^+ laser [12]. (Courtesy of Carl Weiss, Physikalisch-Technische Bundesanstalt, Germany.)

and experiment (bottom). The model (8) is configured so that the cavity is in an absorbing state under planewave excitation. Application of a short, localized pulse locally bleaches the absorber. This increases the finesse of the cavity, and the external driving field may be coupled in with enough extra efficiency to sustain the local bleaching. For the right parameters, a stable self-localized structure evolves, an island of transparency in the otherwise opaque cavity.

As Figure 6 shows, this can work in experiment as well as in theory, and the system acts as a spatial memory of the pulse locations. Such structures are now usually known as cavity solitons, and are described in detail in several of the cited reviews [10]–[12], [31]. An optical soliton is a pulse or beam in which nonlinearity balances dispersion and/or diffraction, so that the soliton propagates without spreading—it is self-localized. In a cavity there is no significant propagation, but the structures of Figure 6 would still spread with time, through diffraction, were that effect not compensated by nonlinearity. So, the term “cavity soliton” seems appropriate. One might enquire exactly how nonlinearity compensates diffraction here. Ordinary solitons usually occur in Kerr-type mediums in which the refractive index is intensity dependent, and the self-confinement arises through self-lensing. No such lens exists in a saturable absorber. Instead, in Figure 6 the bleached region acts as a kind of “gain guide,” because the extra loss in the wings of the beam acts to clip the beam and oppose diffraction. Extra light enters the cavity at the

spot, while light diffracting away from the spot enters a highly lossy region and is absorbed. Thus, a dynamic equilibrium is reached, which is self-sustaining.

By their nature, cavity solitons have to sit on homogeneous backgrounds. Unless the background is stable, the soliton cannot be stable. Furthermore, they have to coexist with the zero-soliton state: you can choose whether or not to create a soliton at a given location. All of these properties point to solitons being associated with a subcritical bifurcation. This is indeed the case [34], and in fact their bifurcation diagrams are practically identical to that for the subcritical roll in Figure 4. In particular, there is an unstable lower-branch soliton solution that acts as a separatrix in terms of writing a soliton [28]. If an address pulse is too weak, it simply dissipates into the background. If the strength of address is gradually increased, the response begins to evolve towards the shape and height of the unstable soliton. There is then a critical behavior. If the height is just too small, it decays; but if it exceeds the unstable soliton's amplitude, it evolves all the way to the stable upper-branch soliton [28].

Cavity solitons have now been predicted and observed for a variety of non-linear mediums [10]-[12], [31]. These include atomlike mediums [32]-[35] like the model (8), and also Kerr mediums [28], [36], and semiconductors [22], [24], [37]-[41], which (as mentioned) are of particular interest. Very similar solitonlike structures occur in the feedback systems previously mentioned [41-42]. Figure 7 shows single and clustered solitons observed in such a feedback system [39].

Underlying the structures in Figure 7 is one of the most interesting features of cavity solitons (generalizing the term to include similar dissipative solitons in feedback systems); namely, that they can move, or be moved, around in the transverse plane. This occurs because the transverse translational invariance of such systems, manifest in (8), implies the existence of a zero-frequency eigenmode (see e.g., Ref. 43), which corresponds to translation of the soliton. The presence of a local gradient (e.g., of the drive field amplitude or phase) tends to excite this mode and cause the soliton to move. The solitons in Figure 7 see gradients of a nature such as to drive them away from the center of the beam, but then they meet an opposing force due to the amplitude profile of the beam and stop. In the presence of other solitons, there are also gradients due to their overlap, leading to a tendency to cluster at favored separations, also evident in Figure 7.

Figure 8 is a perspective view of an equilateral cluster of three cavity solitons in the model of (8), analyzed in a recent work [43]. It shows that the bright peak is surrounded by weak rings alternately darker and brighter than the background. Some evidence of such rings is visible in Figure 7. A second soliton will in general

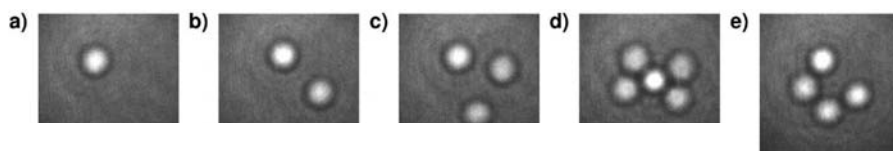


Figure 7 Dissipative solitons observed in a sodium vapor with single feedback mirror. [Reprinted figure with permission from B. Schäpers et al., *Phys. Rev. Lett.* **85** (2000). Copyright (2000) by the American Physical Society, <http://link.aps.org/abstract/PRL/v85/p748>.]

experience a force due to the field gradient associated with these rings. One sitting at an extremum of a ring will experience no force, however. Thus, there are a series of equilibrium soliton separations, which turn out to be alternately stable and unstable [43], very much as the experiments suggest [42]. A cluster such as that in Figure 8 (solitons at the closest unstable separation) is a kind of “soliton molecule,” and as such has natural modes of excitation. The modes of the triangular cluster are also shown in Figure 8. Three of these (b–d) are neutral: two translations and a rotation. The other three (e–g) correspond to dilation or distortion of the triangle, and are the source of its instability.

Clusters of various multiplicities have been studied, and their stability predicted and verified [43]. Figure 9 is an example simulation, showing the evolution of an

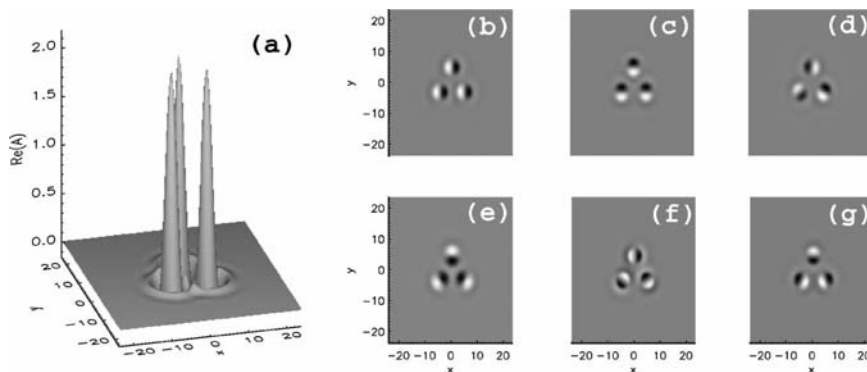


Figure 8 Unstable cavity soliton cluster: perspective view (a), and its normal modes (b–g), of which the upper three are neutral. [Reprinted figure with permission from A.G. Vladimirov et al., *Phys. Rev. E* **65** (2002). Copyright (2002) by the American Physical Society, <http://link.aps.org/abstract/PRE/v65/e046606>.]

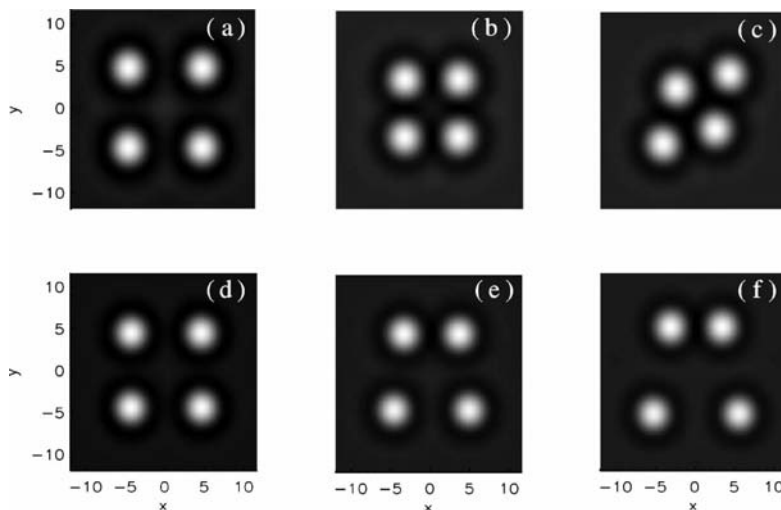


Figure 9 Simulations of dynamical evolution of unstable square clusters of cavity solitons [43]: (a–c) the square contracts, then distorts to form a stable rhombus; with a stronger background field, (d–e) a square evolves to stable trapezoid. Parameters as Figure 6.

unstable square cluster. In case (a) the first stage is evolution towards a close-packed square (b), but this is itself unstable to distortion into a close-packed rhombus. For different parameters (d–f), a stable trapezoid forms (compare panel (e) of Figure 7).

Lastly, consider again the issue of encoding information onto a pattern through controlled irregularity. Suppose a pattern is, in some sense, an infinitely extended regular cluster of solitons. Suppose also that the cluster remains stable when any number of solitons are removed. Then each soliton site can be used as a pixel, “on” when it holds a soliton, “off” when missing. On that basis, a supercluster of N independent solitons can encode 2^N bits of information. Figure 10 shows steps towards this goal [44]. In the simulation (top), effective independence of the hexagon spots in (a) is shown by switching off one, (b), or a group of three, (c). The simulation [44] is based on a semiconductor model, in the bistable regime, so that one can have “dark solitons” on an intense homogenous background. Experimentally [44],

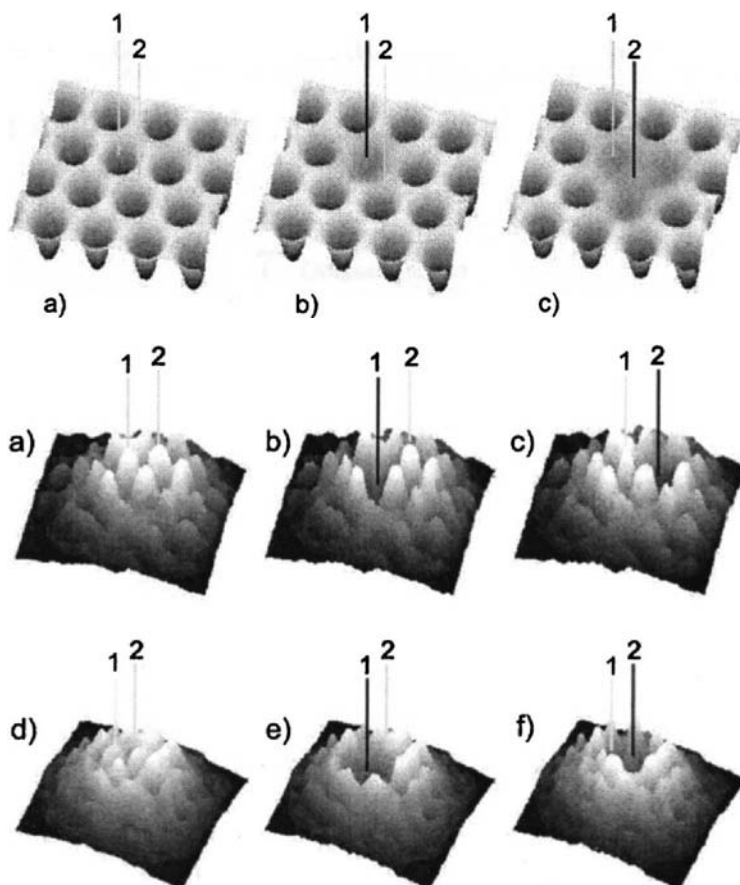


Figure 10 Simulations (top) and experiments (middle, lower) showing localized switching in semiconductor resonators [44] (see text). (Courtesy of Victor Taranenko, Physikalisch-Technische Bundesanstalt, Germany.)

in a semiconductor resonator, the output is viewed in reflection. This negates the input state, so that a high cavity field corresponds to a low reflected field and *vice versa*. Spikes in reflection thus correspond to dark-soliton-like intracavity states. In (a) there is a spontaneous hexagonal pattern under the Gaussian envelope of the driving beam. In (b, c) single spikes are switched off by a short, localized, address pulse aimed at the spike: compare (b) of the simulation above. In the lowest panels, a broader address pulse switches off several spikes, as in (c) of the simulation. Note that, in all these diagrams, the address pulse is over and a new stable equilibrium has been reached. This experiment, then, links and summarizes all the major themes of this article: patterns, solitons, nonlinear dynamics and informatics.

4 Conclusions

Stable patterns and solitonlike dissipative structures are closely linked phenomena with a universal character. They exist in driven optical cavities containing any of a wide variety of nonlinear materials. This class includes semiconductor microresonators, which is promising for possible applications. Qualitatively very similar structures occur in feedback systems. As self-organized dissipative structures, cavity and feedback solitons have useful practical properties like stability and uniqueness. They are mirror-confined in the longitudinal direction, but self-confined, with particlelike properties such as motion under forces, in the transverse plane. They thus provide a means of trapping and manipulating optical pulses and encoding information and images. These solitons can bind into clusters through mutual interactions, leading to a picture of a pattern as an infinitely extended cluster. It is possible in some cases to selectively remove cluster elements, thus enabling each site to be used as an on-off pixel. Patterns, solitons and clusters will be of continuing interest, both in optics and in optoelectronic technology, in the coming years.

Acknowledgments

Thanks are due to many colleagues and collaborators for provision of material and of scientific insights. Particular thanks are due to the PIANOS collaborators and to Thorsten Ackemann, Carl Weiss and Victor Taranenko for additional source material. Financial support from EPSRC grants GR/M 31880 and GR/M 19727, and ESPRIT LTR project 28235 (PIANOS) [24] is gratefully acknowledged.

References

- [1] T. Shin, D. Kraemer, J. Pryor, L. Liu, J. Ruglia, L. Howe, S. Buck, K. Murphy, L. Lyons and M. Westhusin, A cat cloned by nuclear transplantation, *Nature* **415** (2002), 859.
- [2] H.G. Solari, M.A. Natiello and G.B. Mindlin, *Nonlinear dynamics: A two way trip from physics to maths*, Institute of Physics Publishing, Bristol, UK, 1996.

- [3] J. von Hardenberg, E. Meron, M. Shachak and Y. Zarmi, Diversity of vegetation patterns and desertification, *Phys Rev Lett* **87** (2001), 198101.
- [4] M.C. Cross and P.C. Hohenberg, Pattern-formation outside of equilibrium, *Rev Mod Phys* **65** (1993), 851–1112.
- [5] N. Savage, A bright spot: The evolution of the laser market, *OSA Opt Photon News* **13** (5) (2002), 24–27 (May issue).
- [6] P. Kner, D. Sun, J. Boucart, P. Floyd, R. Nabiev, D. Davis, W. Yuen, M. Jansen and C.J. Chang-Hasnain, VCSELS go the distance, *OSA Opt Photon News* **13** (3) (2002), 45–47 (March issue).
- [7] J. Partanen, Lasers for solid imaging, *Opt Photon News* **13** (5) (2002), 45–48 (May issue).
- [8] J.W. Goodman, *Introduction to Fourier optics*, McGraw-Hill, London, UK, 1968.
- [9] G. Stegeman, N. Christodoulides and M. Segev, Optical spatial solitons: Historical perspectives, *IEEE J Sel Top Quant Electron* **6** (2000), 1419–1427.
- [10] S. Trillo and W. Torruellas (eds), *Spatial solitons*, Springer, Heidelberg, Germany, 2001.
- [11] A.D. Boardman and A.P. Sukhorukov (eds), *Soliton-driven photonics*, Kluwer, Amsterdam, The Netherlands, 2001.
- [12] *OSA Optics and Photonics News*, Feature Issue Spatial Solitons, February 2002.
- [13] L.A. Lugiato, Spatiotemporal structures—I, *Phys Rept* **219** (1992), 293–310.
- [14] C.O. Weiss, Spatiotemporal structures—II. Vortices and defects in lasers, *Phys Rept* **219** (1992), 311–338.
- [15] P.K. Jakobsen, J.V. Moloney, A.C. Newell and R. Indik, Space-time dynamics of wide gain-section lasers, *Phys Rev A* **45** (1992), 8129–8137.
- [16] W.J. Firth, Spatial instabilities in a Kerr medium with a single feed-back mirror, *J Modern Opt* **37** (1990), 151–154.
- [17] G. D'Alessandro and W.J. Firth, Hexagonal spatial patterns in a simple optical system, *Phys Rev Lett* **66** (1991), 1597–1600.
- [18] A. Aumann, E.G. Westhoff, R. Herrero, T. Ackemann and W. Lange, Interplay of dispersion and absorption in a new optical pattern-forming system, *J Opt B: At Mol Opt* **1** (1999), 166–172.
- [19] E. Große Westhoff, PhD thesis, University of Münster, Münster, Germany, 2002.
- [20] T. Ackemann, E. Große Westhoff, M. Pesch, D. Rudolph and W. Lange, Optical pattern formation far beyond threshold, *ICONO 2001: Nonlinear optical phenomena and nonlinear dynamics of optical systems* (K.N. Drabovich, N.S. Kazak, V.A. Makarov and A.P. Voitovich, eds), Proc SPIE **4751** (2002), 370–381.
- [21] R. Herrero, E.G. Westhoff, A. Aumann, T. Ackemann, Y.A. Logvin and W. Lange, Twelfefold quasiperiodic patterns in a nonlinear optical system with continuous rotational symmetry, *Phys Rev Lett* **82** (1999), 4627–4630.

- [22] V.B. Taranenko, I. Ganne, R. Kuszelewicz and C.O. Weiss, Patterns and localized structures in bistable semiconductor resonators, *Phys Rev A* **61** (2000), 063818.
- [23] T. Ackemann, S. Barland, J.R. Tredicce, M. Cara, S. Balle, R. Jaeger, M. Grabherr, M. Miller and K.J. Ebeling, Spatial structure of broad area vertical-cavity regenerative amplifiers, *Opt Lett* **25** (2000), 814–816.
- [24] Published and conference material (including movies) relevant to PIANOS and to the present article, can be found at www.pianos-int.org.
- [25] L.A. Lugiato and R. Lefever, Spatial dissipative structures in passive optical systems, *Phys Rev Lett* **58** (1987), 2209–2211.
- [26] W.J. Firth, G.K. Harkness, A. Lord, J.M. McSloy, D. Gomila and P. Colet, Dynamical properties of 2D Kerr cavity solitons, *J Opt Soc Am B* **19** (2002), 747–752.
- [27] A.J. Scroggie and W. J. Firth, Spontaneous pattern formation in an absorptive system, *Europhys Lett* **26** (1994), 521–526.
- [28] G.K. Harkness, W.J. Firth, G.-L. Oppo and J.M. McSloy, Computationally-determined existence and stability of transverse structures: I. Periodic optical patterns, *Phys Rev E* **66** (2002), 046605.
- [29] Y. Pomeau, Front motion, metastability and subcritical bifurcations in hydrodynamics, *Physica D* **23** (1986), 3–11.
- [30] P. Coullet, C. Riera and C. Tresser, Stable static localized structures in one dimension, *Phys Rev Lett* **84** (2000), 3069–3072.
- [31] W.J. Firth and G.K. Harkness, Cavity solitons, *Asian J Phys* **7** (1998), 665–677.
- [32] N.N. Rosanov and G.V. Khodova, Autosolitons in bistable interferometers, *Opt Spectrosc (USSR)* **65** (1988), 449–450.
- [33] M. Tlidi, P. Mandel and R. Lefever, Localized structures and localized patterns in optical bistability, *Phys Rev Lett* **73** (1994), 640–643.
- [34] W.J. Firth and A.J. Scroggie, Optical bullet holes: robust controllable localized states of a nonlinear cavity, *Phys Rev Lett* **76** (1996), 1623–1626.
- [35] M. Brambilla, L.A. Lugiato and M. Stefani, Interaction and control of optical localized structures, *Europhys Lett* **34** (1996), 109–114.
- [36] D. W. McLaughlin, J.V. Moloney and A.C. Newell, Solitary waves as fixed points of infinite-dimensional maps in an optical bistable ring cavity, *Phys Rev Lett* **51** (1983), 75–78.
- [37] M. Brambilla, L.A. Lugiato, F. Prati, L. Spinelli and W.J. Firth, Spatial soliton pixels in semiconductor devices, *Phys Rev Lett* **79** (1997), 2042–2045.
- [38] D. Michaelis, U. Peschel and F. Lederer, Multistable localized structures and superlattices in semiconductor optical resonators, *Phys Rev A* **56** (1997), R3366–R3369.
- [39] L. Spinelli, G. Tissoni, M. Brambilla, F. Prati and L.A. Lugiato, Spatial solitons in semiconductor microcavities, *Phys Rev A* **58** (1998), 2542–2559.
- [40] V.B. Taranenko, I. Ganne, R. Kuszelewicz and C.O. Weiss, Spatial solitons in a semiconductor microresonator, *Appl Phys B* **72** (2000), 377–380.

- [41] A. Schreiber, B. Thuring, M. Kreuzer and T. Tschudi, Experimental investigation of solitary structures in a nonlinear optical feedback system, *Opt Commun* **136** (1997), 415–418.
- [42] B. Schäpers, M. Feldmann, T. Ackemann and W. Lange, Interaction of localized structures in an optical pattern-forming system, *Phys Rev Lett* **85** (2000), 748–751.
- [43] A.G. Vladimirov, J.M. McSloy, D.V. Skryabin and W.J. Firth, Two-dimensional clusters of solitary structures in driven optical cavities, *Phys Rev E* **65** (2002), 046606.
- [44] V.B. Taranenko, C.O. Weiss and B. Schäpers, From coherent to incoherent hexagonal patterns in semiconductor resonators, *Phys Rev A* **65** (2002), 013812.

Additional references not directly cited

- [45] N.N. Rosanov, Transverse patterns in wide-aperture nonlinear optical systems, *Progress in optics XXXV* (E. Wolf, ed), North-Holland, Amsterdam, The Netherlands, 1996, 1–60.
- [46] H. Haken, *Synergetics, an introduction*, Springer, Berlin, Germany, 1997.
- [47] G. Nicolis, *Introduction to nonlinear science*, Cambridge University Press, Cambridge, UK, 1995.
- [48] F.T. Arecchi, Space-time complexity in nonlinear optics, *Physica D* **51** (1991), 450–464.
- [49] W.J. Firth, Pattern formation in passive nonlinear optical systems, *Self-organization in optical systems and applications in information technology* (M. A. Vorontsov and W. B. Miller, eds), Springer, Berlin, Germany, 1995, 69–96.
- [50] L.A. Lugiato, M. Brambilla and A. Gatti, Optical pattern formation, *Advances in atomic, molecular and optical physics* **40** (B. Bederson and H. Walther, eds), Academic Press, London, UK, 1998, 229–306.
- [51] M. Segev, B. Crosignani and A. Yariv, Spatial solitons in photorefractive media, *Phys Rev Lett* **68** (1992), 923–926.
- [52] O. Thual and S. Fauve, Localized structures generated by subcritical instabilities, *J Phys* **49** (1988), 1829–1923.
- [53] G. Dewel, P. Borckmans, A. de Wit, B. Rudovics, J.-J. Perrand, B. Dulos, J. Boissonade and P. de Kepper, Pattern selection and localized structures in reaction-diffusion system, *Physica A* **213** (1995), 181–198.
- [54] M. Saffman, D. Montgomery and D.Z. Anderson, Collapse of a transverse-mode continuum in a self-imaging photorefractively pumped ring resonator, *Opt Lett* **19** (1994), 518–520.
- [55] G. Tissoni, L. Spinelli, M. Brambilla, T. Maggipinto, I.M. Perrini and L.A. Lugiato, Cavity solitons in passive bulk semiconductor microcavities. I. Microscopic model and instabilities, *J Opt Soc Am B* **16** (1999), 2083–2094.

- [56] G. Tissoni, L. Spinelli, M. Brambilla, T. Maggipinto, I.M. Perrini and L.A. Lugiato, Cavity solitons in passive bulk semiconductor microcavities. II. Dynamical properties and control, *J Opt Soc Am B* **16** (1999), 2095–2105.



William J. Firth is Professor of Natural Philosophy and Head of the Department of Physics at the University of Strathclyde, UK. He has published more than 200 journal articles; has contributed chapters to a dozen research books; has reviewed for many journals, and has been the Co-Editor for *Physical Review E* from 2002 till the present. His current research interests lie in nonlinear dynamics and spontaneous spatial structure formation in optical and other systems.



John M. McSloy has very recently completed his Ph.D. and is currently a Research Assistant in the Department of Physics at the University of Strathclyde, UK. He has published six journal articles and contributed chapters to one research book. His current research interests lie in spatial structure formation and optical solitons in nonlinear systems.

Statistical Approaches to Scattering

Walid Tabbara, Véronique Rannou, and Stefano Salio

Abstract

Statistical and probabilistic methods have frequently been used to model and analyze electromagnetic scattering problems. This approach may in some cases be the most realistic one. In others, when combined with deterministic methods, it allows to extract from the available data information that is otherwise difficult to obtain. The aim of this chapter is to provide an introduction to some statistically based approaches to the analysis of scattering problems, and to illustrate their efficiency and limitations. A general methodology of prediction of the behavior of a complex system can thus be established which, among many advantages, allows a significant reduction of the computational cost.

1 Introduction

What might be the motivation for the introduction of randomness into the analysis of coupling of an electromagnetic wave to a system, since both the system and the mediums can be described in a deterministic way? A consideration of the following two cases provides a rough answer.

- 1) We are faced with a situation in which there are no means to predict which configuration of the given system will occur. We are thus led to observe the occurrences of the configurations and count the number of realizations of each one. Next, a statistical data analysis is conducted in order to define the probabilities of occurrence of the various configurations.
- 2) Although dealing with a deterministic problem that can be analyzed numerically or analytically by means of well-known methods, the large number of configurations of the system one has to investigate leads to prohibitive computational costs. It is then more convenient to consider the configurations as the realizations of a random experiment and introduce probabilistic models for the observations.

Many problems in the field of electromagnetic wave scattering can be analyzed as stated above in the second case, such as

- *propagation problems in complex mediums*, including, for example, propagation through rain, scattering from a forest canopy or from soil, and scattering from a random distribution of particles in a host medium;
- *propagation problems in urban environments* concerning the statistical modeling of the amplitudes of signals scattered by obstacles; and
- *electromagnetic compatibility problems*, such as statistical distribution of the field amplitude in a reverberating chamber [1], the probability law of the voltage induced by an incoming wave on a cable [2], or potential problems [3].

2 Elements of the statistical vocabulary

The aim of this section is to introduce some elements of the statistical vocabulary that are needed for the understanding of this chapter. To achieve this goal, we refer to examples in the field of electromagnetic compatibility (EMC). There one has to deal with the coupling of an electromagnetic wave to a complex system. This is in no way a limitation of the tools and methods to be described, as they can be used without any significant modification to analyze scattering by any complex medium. When we talk of voltage and current in EMC, we may as well consider reflection, transmission or absorption coefficients in the frame of other applications.

The starting point in statistics [4-5] is the *population*, which is the set of all possible configurations of a system. From the population, one extracts a subset of configurations called the *sample*, and analysis is then conducted on each element

of the sample. In a final step, the results obtained for this subset must be extended in some way to the whole population. By doing so, we predict the behavior of a configuration that does not belong to the sample from the behavior of those that are part of it.

The parameter selected to measure the coupling of the illuminating wave to the system is called the *observable*. It is a function of some other parameters, such as frequency, dimensions of the system, permittivity, conductivity, etc., called *factors*. A configuration of the system is defined by considering a combination of values of the factors. The latter and the observable are considered as random variables. For some simple systems, it is possible to determine analytically the probability law of the observable from those of the factors. In that case, one usually employs the *Monte Carlo* technique where a large number of combinations of the values of the factors are chosen at random and the value of the observable is computed, or measured, for each of these combinations. A *histogram*, also called bar graph, of the values of the observable is then drawn where the length of each bar is equal to the frequency of occurrence of the corresponding value of the observable. The histogram is then fitted with a function known as the *probability density function* (pdf) of the observable. The knowledge of the pdf permits the computation of the probability of occurrence of any value of the observable.

Various techniques are used to build a sample of values of the observable. Here we make use of the one called *experiment design* [6]. For each factor, we select a small number of values over its interval of definition—such a number being called the *level of the factor*. Let us consider a system controlled by means of three factors, f_1 , f_2 and f_3 , the level of which are respectively n_1 , n_2 and n_3 . The selected values of f_i , ($i = 1, 2, 3$), are $\{f_{i,k}\}$, $k = (1, \dots, n_i)$. Next, we consider the set of all the combinations of values of the factors, called triplets:

$$(f_{1,i}, f_{2,j}, f_{3,k}), \quad i = 1 \dots n_1, \quad j = 1 \dots n_2, \quad k = 1 \dots n_3.$$

This set is called a complete factorial experiment design. Each triplet represents one configuration of the system. Let $V(f_1, f_2, f_3)$ be the observable, a function of the three factors, and let $\{V_{ijk} = V(f_{1,i}, f_{2,j}, f_{3,k})\}$ be its values for each triplet. This set of values of the observable is our database. We consider now a new configuration of the system that does not belong to the database; the associated value of the observable is then predicted from the ones in the database by means of an interpolation technique called *kriging* [7], first introduced in the field of geostatistics [8]. The interpolating function goes through the data points, and kriging provides for each predicted (interpolated) value a measure of its accuracy by mean of a standard deviation. This method can also be used to interpolate the pdf obtained by the Monte Carlo approach [9].

Problems of practical interest involve a large number of factors, each having more or less influence on the observable. In order to reduce the computational cost we only need to keep the most important factors. Moreover, the factors may not be independent parameters; i.e., the observable is also a function of products like $f_i f_j$.

Thus, there is a need to point out the correlation between them. To achieve these two goals one may resort to the *analysis of variance* (ANOVA) and to *regression* techniques [10]. The ANOVA, which allows one to investigate the influence of one factor at a time, is a fast and simple method, but it does not provide information on the correlations between factors. Regression is a more elaborate approach that may be used in place of ANOVA and is able to give correlation information. It may also be used to interpolate the values of the observable in the database. The interpolating function is then a linear combination of functions of the factors called regressors, (these are usually polynomials in many variables) and the coefficients of the linear combination are obtained by applying a mean-square minimization to some cost function.

Other statistical and probabilistic tools exist but they are not addressed here. One must keep in mind that the main difficulty in statistics lies in the interpretation of the results obtained when statistical techniques are applied.

3 The statistical approach

The statistical approach is very suitable for analysis of scattering and coupling of waves to systems problems, especially when one has to contend with reliability and quality control. Thus, the behavior of a system can be characterized by means of a probability of failure, or the probability for an observable to have values below or above a given level. Furthermore, the need for statistical methods is motivated, as stated in the second motivational case in Section 1, by the large number of possible configurations of the system being dealt with. This makes it difficult to guess which configuration one will have to consider. Such variability can be attributed to the following reasons:

- *Inner variability of the system.* This is due to the only approximate knowledge of some factors of the system, such as the values of the permittivity, the dimensions or the randomness in the location of some parts of the system—as, for example, particles in a host medium or the path of cables in a vehicle. These uncertainties in the factors make the observable behave as a random variable.
- *Environmental variability.* Here, the characteristics of the environment of a system are in general not known in advance; for example, in the case of a random illumination of the system or an unpredictable variation of the electrical parameters of the environment. A probabilistic model of the latter is then necessary, and the randomness of the environment will lead to unpredictable values of the observable.

A consequence of the variability is the high computational or experimental cost induced by a classical analysis of an observable as a function of a large number of factors. To achieve this task, one has to study the evolution of the observable for a significant number of values of each factor taken independently. This means that a large number of system configurations are considered. In order to overcome these difficulties, we suggest the following approach:

- 1) Select the main factors by mean of ANOVA and regression techniques.
- 2) Build a small sample of configurations of the system by means of the experiment design approach. As a configuration is defined by a combination of values of the factors, in experiment design 2 to 5 values are chosen for each factor. When three factors are considered, this leads to a study of 8 to 125 configurations for a complete factorial experiment design. Incomplete factorial design may also be considered, which reduces the computational time but also the accuracy of the method. Then, for each configuration in the sample, the value of the observable is computed (measured) by standard techniques; and this set of values provides the database.
- 3) When one considers a configuration that does not belong to the database, it can be defined by a combination of values of the factors. The value of the observable, for this configuration, is predicted by interpolating the values in the database by mean of kriging. With this technique we study the evolution of an observable with respect to all the factors simultaneously rather than one at a time as in a classical parametric analysis.

What is the budget of the statistical approach? The main cost is induced by the construction of the database, whether this is done numerically, experimentally or by combining the two. In the case where numerical simulation is used, time-consuming methods such as finite differences, finite elements or integral representations are involved. On the other hand, ANOVA, regression and kriging do not increase significantly the cost of the method. It must be noted that the statistical approach does not replace any standard modeling method. Rather, it works in association with the latter, enhancing its capabilities and making it possible to retrieve information that is otherwise difficult to obtain. As a final remark, this approach applies to a probability density function where the interpolation is done on the parameters of this function [9].

To illuminate the statistical approach, it is applied to two EMC problems in the following section. The first deals with the prediction of the crosstalk between the wires of a cable placed above a ground plane, and it is used to illustrate the steps followed in the application of the statistical approach. The practical application of this scenario is that of a bundle of wires running over the chassis of a car. The second problem addresses the coupling of the field radiated by a randomly located dipole to a transmission line. This is a first-order solution to the coupling of mobile phone signals to transmission lines placed in a room.

4 Application I: Crosstalk

4.1 Geometrical representation of the cable

A cable 2 m long and made of 9 wires is placed above a perfectly conducting ground plane, as shown in cross-section in Figure 1. The cable diameter is 7.9 mm, the wire diameter is 1.5 mm, and the thickness of its dielectric jacket is 0.4 mm

with a relative permittivity of 2.6. The medium between the wires is air. The center of the central wire (conductor 1) is located at a distance of 7.3 mm above the ground plane. If applied to a car scenario, the cable is not a uniform structure along its length, and the wires are twisted in such a disordered way that it is not simple to describe its path by means of a deterministic equation. Among the methods used to model the geometry of the cable is the *Random Midpoint Displacement* (RMD) algorithm [11] based on the use of fractals. The disorder in the wires position is described by the fractal dimension D [12], which when equal to 1 leads to a uniform cable, and nonuniformity increases with D . An example of a cable of 3 wires with $D = 1.8$ is shown in Figure 2.

4.2 Electrical parameters of the cable

Identical resistive loads are placed between the wire terminations and the ground plane. Four cases are considered: $10\ \Omega$, $100\ \Omega$, $1,000\ \Omega$ and $10,000\ \Omega$. At the lowest value we experience an inductive coupling between the wires, at the two highest ones we have a capacitive coupling, and a combination of the latter two occurs at the second value of the impedance. The sole exception is for the excited termination (left end termination of wire 1), where a $50\text{-}\Omega$ resistance is placed so

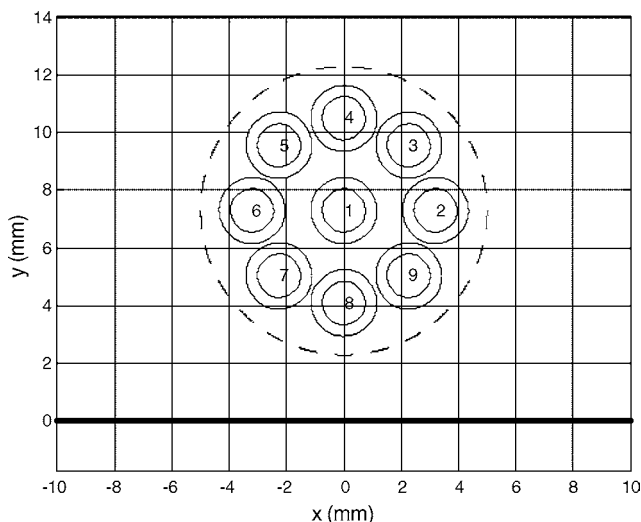


Figure 1 Cross-section of the 9-wire cable.

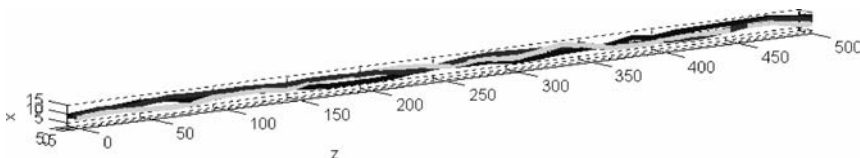


Figure 2 A 3-wire cable generated by the RMD algorithm; fractal dimension $D = 1.8$.

as to act as the output impedance of an ideal sinusoidal 1-V generator operating over the frequency range 1 kHz to 500 MHz.

The crosstalk voltage is the voltage observed at an end termination impedance of a wire, induced by the signal propagating in a wire in its neighborhood. Here, the crosstalk voltage is determined at the left-end termination resistance of wire 2, and numerical and experimental results have been obtained that are in good agreement. Measurements were performed at the European Air Defense and Space (EADS) Common Center of Research, and numerical simulations were carried out using either a nonuniform, multiconductor transmission lines code [13] or the CRIPTE code based on the BLT equation [14]. The computation time over the whole frequency range is less than a minute on a standard personal computer for the first numerical code, and of the order of few minutes on a workstation for the second code. A typical example of the crosstalk voltage is shown in Figure 3.

At low frequencies the crosstalk increases linearly with frequency, with a slope of 20 dB/decade; thus, over this subrange one needs to predict the voltage at only one frequency. On the other hand, at the higher end of the frequency range, between 100 MHz and 500 MHz, many resonances occur, the sources of which are difficult to trace back exactly. In practical applications it is useless to predict the value of the voltage at each frequency; and in order to comply with a given standard, it is sufficient to determine the maximum value of the voltage. The mean value of the envelope may in some cases provide useful information.

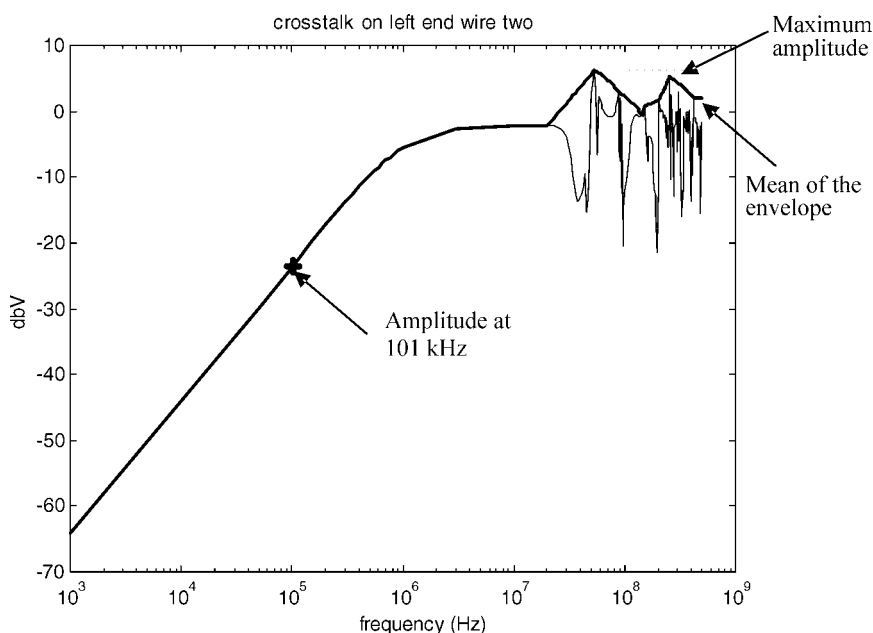


Figure 3 Crosstalk voltage at the left end of wire 2. The three observables considered here are the voltage amplitude at 101 kHz (Ob1); the mean value of the voltage envelope over the frequency range; the maximum voltage of the envelope over the range 100 MHz to 500 MHz (Ob2); and the mean voltage of the envelope (Ob3) over the same range.

4.3 Observables and factors

Three observables have been selected according to their ability to carry the requested information for the system to comply with existing standards; see Figure 3. These are the voltage at the frequency of 101 kHz (Ob1), the maximum voltage of the envelope over the range 100 MHz to 500 MHz (Ob2), and the mean voltage of the envelope (Ob3) over the same range.

We emphasize that as a result of applying ANOVA and regression, the main factors are the height (h) of the cable (distance of the center of wire 1 to the ground plane), the fractal dimension (D) and the loads (Z) at the ends of the wires. The levels of the factors are respectively equal to 3, 3 and 4. The selected values for each of the three factors are $h = 7.3$ mm, 12.3 mm, 22.3 mm; $D = 1, 1.1, 1.5$; and $Z = 10 \Omega, 100 \Omega, 1,000 \Omega, 10,000 \Omega$.

The height values are not equally spaced in order to accurately model the higher influence of the ground plane for small values of h , especially at low frequencies and for inductive coupling. The uneven sampling of the values of D ensures a good transition between a uniform and a nonuniform cable. Thus, we see that the choice of the sampled values of the factors is based on physical observations and reflects the rate of change of the observable with respect to each factor.

4.4 Experiment design

The aforementioned choice of factor levels leads to a complete factorial experiment design of size $3 \times 3 \times 4 = 36$. This means that we have to conduct 36 numerical simulations (or 36 measurements, or a combination of both) for each of the 3 observables in order to build the corresponding database, as defined in Section 2. For reasons that will become clearer later on, we have added 12 other data points defined for the following combinations of the values of the factors: ($h = 10.3, D = 1.35, Z = 10, 100, 1000, 10,000$), ($h = 17.3, D = 1.1, Z = 10, 100, 1000, 10,000$), ($h = 17.3, D = 1.35, Z = 10, 100, 1000, 10,000$). The new experiment design can be represented in a three-dimensional space (h, D, Z) as in Figure 4, where only 12 of the 48 points (indicated by stars) are shown. The values of the observables are computed at the 48 points (48 configurations of the system) and stored in three different databases. The three axes show the values of h, D and $\log(Z)$.

The “box” of Figure 4 represents the volume that contains all possible configurations of the system, each one represented by a point with a coordinate vector (h, D, Z), the components of which vary over the segments (7.3 mm, 22.3 mm), (1, 1.5) and (10 Ω , 10,000 Ω), respectively. Let us now consider one of the previously defined observables; we know its values at 48 points inside the “box,” and we want to predict its value at a new point (a new configuration of the system) by interpolating between the already available 48 values. The interpolated values are to be mapped into the two planes drawn in bold in Figure 4, or along lines in these same planes. This interpolation is to be achieved using kriging.

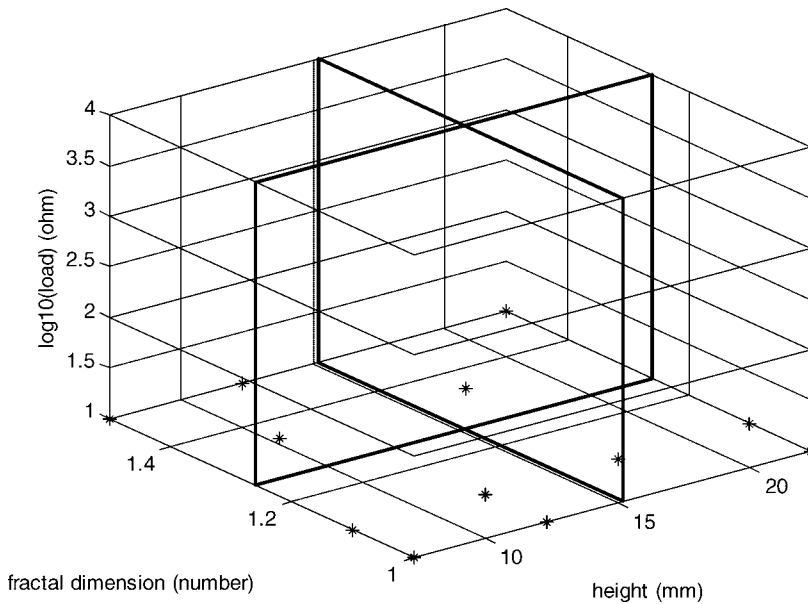


Figure 4 Three-dimensional representation of the experiment design. Each asterisk is one of the 48 configurations of the system (only 12 are shown here).

4.5 A brief presentation of kriging

Let $\mathbf{x} = (h, D, Z)$ be the vector comprising the factors controlling our system, and $y(\mathbf{x})$ the value of the observable y for the configuration defined by $\mathbf{x}$. Kriging is a multifactor parametric approach to system modeling, and it provides the best linear unbiased predictor (BLUP) of $y(\mathbf{x})$. The latter satisfies

$$y(\mathbf{x}) = \mathbf{r}^t(\mathbf{x})\mathbf{p} + \omega(\mathbf{x}), \quad (1)$$

where $\mathbf{r}^t(\mathbf{x})\mathbf{p}$ is a regression model [10] (t stands for transpose), $\mathbf{p}$ is the vector of unknown parameters, and $\mathbf{r}(\mathbf{x})$ is a vector function of the factors called the regressor vector. Finally, $\omega(\mathbf{x})$ is a realization of Gaussian random process with zero mean and covariance function

$$\text{Cov}[\omega(\mathbf{x}_i), \omega(\mathbf{x}_j)] = \sigma_\omega^2 C\omega(\mathbf{x}_i, \mathbf{x}_j), \quad (2)$$

wherein $C\omega$ is a decreasing function of $\|\mathbf{x}_i - \mathbf{x}_j\|$ (Euclidian norm), $\mathbf{x}_i$ and $\mathbf{x}_j$ are the factor vectors of two configurations of the experiment design. For any new configuration, defined by its factor vector $\mathbf{x}$, that does not belong to the database, $y(\mathbf{x})$ is sought as a linear combination of the data $y(\mathbf{x}_i)$ $i = 1, \dots, n$ (here $n = 48$):

$$y(x) = \sum_{i=1}^n \lambda^i y(\mathbf{x}_i). \quad (3)$$

Therein, λ_i are unknowns that satisfy the constraint of nonbias of $y(\mathbf{x})$:

$$\sum_{i=1}^n \lambda^i = 1. \quad (4)$$

The maximum likelihood estimate of $\mathbf{p}$ is obtained by minimizing the mean square error

$$\text{mse} = E[\boldsymbol{\lambda}'\mathbf{Y} - \mathbf{r}'(\mathbf{x})\mathbf{p} - \omega(\mathbf{x})]^2. \quad (5)$$

In this expression, $\mathbf{Y}$ is a matrix with columns containing the vectors $\mathbf{y}(\mathbf{x}_i)$, $\boldsymbol{\lambda}$ is the vector made of λ_i and E is the mathematical expectation. When λ_i and $\mathbf{p}$ are replaced by their estimates, the square root of mse is the standard deviation $\sigma(\mathbf{x})$ of the predicted value of $y(\mathbf{x})$, which is a measure of the accuracy of the prediction. This standard deviation is equal to zero when $\mathbf{x} = \mathbf{x}_i$, which means that the interpolator goes through the data points. One may also estimate σ_ω^2 and the parameters of $C\omega$. For more details about kriging the reader may consult the literature [7-8]. In the application considered here the regression model of (1) is reduced to a constant p_0 ; such a simplification is called ordinary kriging.

5 Transmission-line coupling

The configuration is shown in Figure 5 and consists of a single-wire transmission line of length $L = 1.5$ m, wire diameter $d = 1$ mm, located at height h cm above

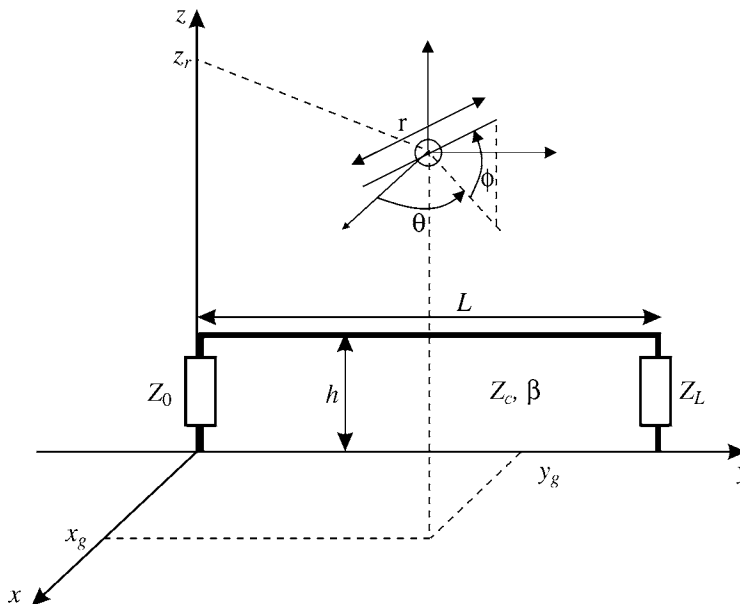


Figure 5 Transmission line over a ground plane. Source: randomly located dipole emitting at 900 MHz.

an infinitely extended perfectly conducting ground plane, loaded by impedances $Z_0 = 50 \, \Omega$ and Z_L . The characteristic impedance of the line is Z_c , and β is its propagation constant. The transmission line is illuminated by an elementary dipole source emitting at the frequency of 900 MHz. The location of the source is defined by the coordinates of its center (x_s, y_s, z_s) , and its orientation by angles θ and ϕ , which are uniformly distributed random variables over the ranges $(0, \pi/2)$ and $(0, 2\pi)$, respectively.

The current I at load Z_L is computed by means of the method of moments applied to the integral representation of the scattered field [15]. As the source is randomly placed, the induced current I will be a random variable, and we will determine its pdf using a Monte Carlo approach.

The observable is the 95% quantile of the current I_{95} , defined by¹

$$\text{Probability}(I < I_{95}) = 0.95. \quad (6)$$

The factors are the height h and the impedance Z_L with a level of three for each one, leading to two one-dimensional complete experiments designs, i.e., two databases of three elements each. All other parameters of the transmission line have fixed values. Kriging is to be used to interpolate the values of the observable, one factor at a time.

6 Results

6.1 Coupling to a cable

Kriging is applied to the database of each observable separately, and the 48 values in each database are interpolated to predict the values of the related observable at 8000 points (i.e., configurations) obtained by combining 20 values for each of the three factors h , D and Z . A first representation of the results is by means of maps on plane sections defined by a constant value of one of the factors. Let us consider a height $h = 15.19 \, \text{mm}$, which does not belong to any of the databases and plot the map of the values of Ob1 (see Section 3.1.3) in the plane $h = 15.9 \, \text{mm}$; see Figure 6.

The map on the left side of Figure 6 shows the predicted values of Ob1 at $h = 15.9 \, \text{mm}$ for combinations of values of D (vertical axis) and $\log Z$ (horizontal axis). We thus have quick access to the variations of Ob1 as a function of the factors D and $\log Z$. For example, if we look at this evolution along a line at $\log Z = 2$ (i.e., $100 \, \Omega$), we see no significant changes in the “color.” We can therefore conclude that Ob1 is not strongly influenced by I , the disorder of the bundle of wires; a statement that is in good agreement with experimental results. The map on the right side of Figure 6 shows the values of the standard deviation σ provided by kriging for each predicted value. We thus have a point-to-point correspondence between the two maps. The lower the value of σ , the higher the accuracy of the prediction,

¹The p th quantile, x_p , of a random variable X is such that $P(X < x_p) = p\%$.

with $\sigma = 0$ at the data points. High values of σ mean that the point where the prediction is made is not close enough to data points, and one can then improve the prediction by adding more configurations to the experiment design and applying kriging to the new database. This was the reason for adding 12 new configurations to the original 36 ones, as was discussed in Section 3.4. Similar comments apply to Ob2 (Figure 7) and Ob3 (Figure 8).

Let us now consider, in the aforementioned plane sections, the variation of each observable as a function of $\log Z$ along the line $D = 1.26$, a value that does not belong to the databases.

In each of the frames in Figure 9, the curves display the predicted values of the observable $+$ (upper curve) or $-$ (lower curve) σ . All predicted values must fall in the strip bounded by the two curves. The width of the strip reflects the accuracy of the prediction and is influenced by the choice of the covariance function in (2). A Gaussian covariance smoothes the results too much; spherical (used here) and exponential covariances are more suitable for the present problem. See Ref. 8 for more information on covariance functions.

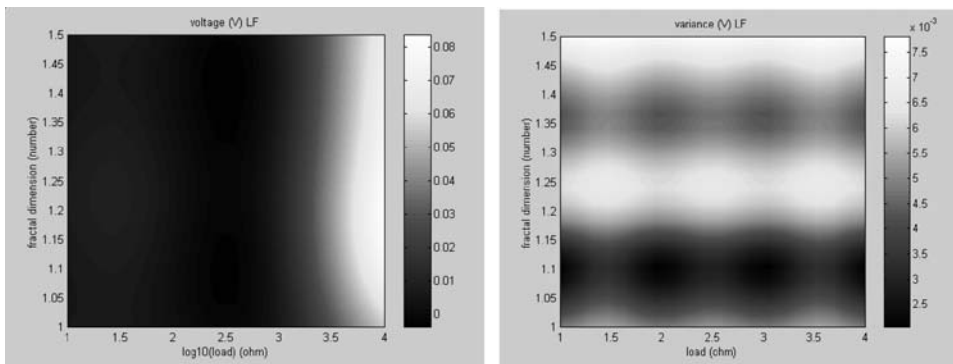


Figure 6 Maps of the predicted values of Ob1 (left), and of the related σ (right). Plane cut at $h = 15.9$ mm.

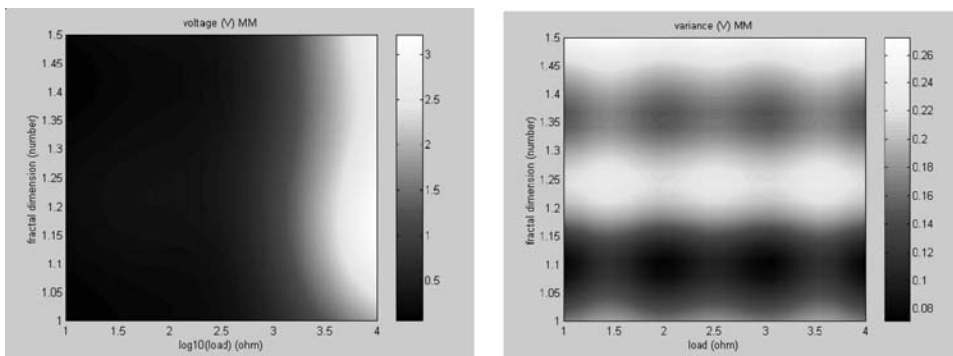


Figure 7 Maps of the predicted values of Ob2 (left) and of the related σ (right). Plane cut at $h = 15.9$ mm.

To check the validity of the approach we have computed the exact values of the observables for three configurations that do not belong to the databases and that are defined by the following values of the factors (h , D , Z) as (15.9 mm, 1.26, 42.8 Ω), (15.9 mm, 1.26, 545 Ω), (15.9 mm, 1.26, 4833 Ω).

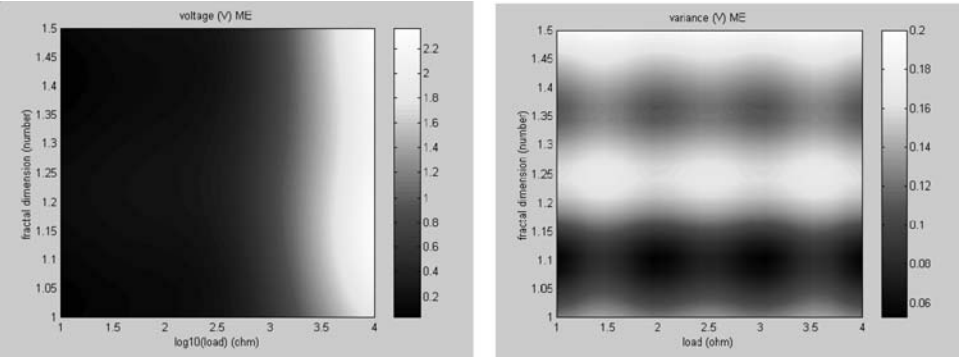


Figure 8 Maps of the predicted values of Ob3 (left), and of the related σ (right). Plane cut at $h = 15.9$ mm.

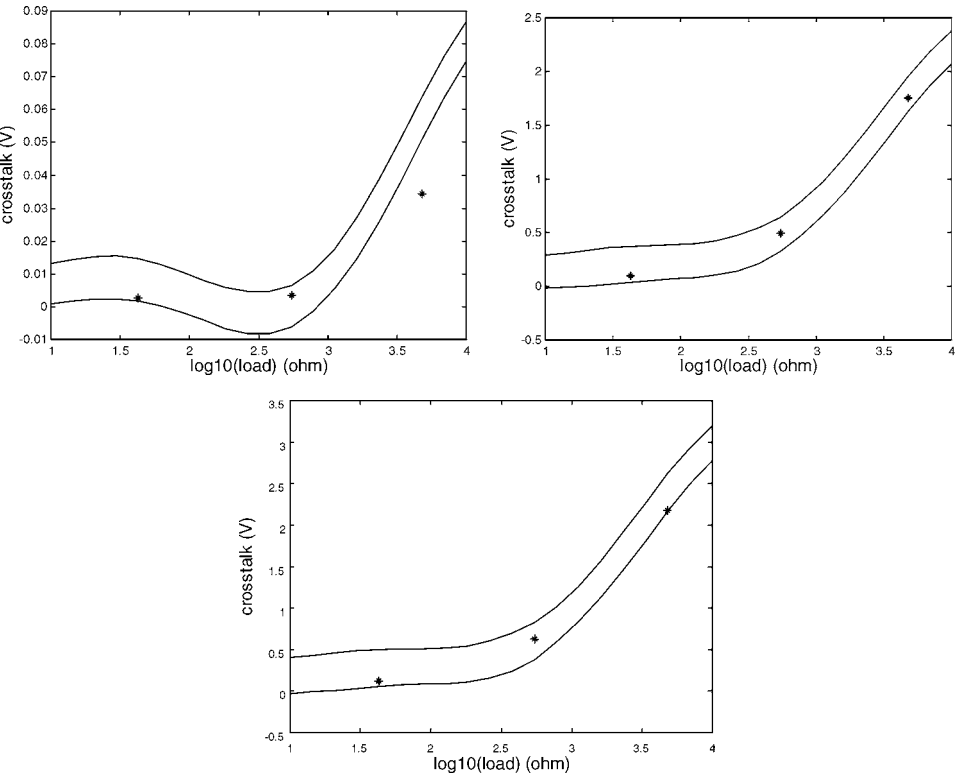


Figure 9 (upper left) Ob1 $\pm \sigma$, (upper right) Ob2 $\pm \sigma$, (lower center) Ob3 $\pm \sigma$ as a function of $\log(Z)$; $h = 15.9$ mm, $D = 1.26$, (*) test points, exact values.

The results are reported in Figure 9 where the points are indicated an asterisk (*). As long as they fall in the prediction strip we can consider that we have achieved our goals using kriging. We cannot expect more from this technique, besides reducing the width of the strip by increasing the size of each database and therefore incur a higher computational cost. Thus, one must look for a trade-off between accuracy and computational cost.

As a final check, we have considered a configuration defined by the factors $h = 17.56$ mm, $D = 1.29$ and $Z = 100 \Omega$, and we have randomly generated, with the RMD algorithm, 50 cables with these values of the factors. For each of the latter we have computed the exact values of the voltage over the whole frequency range, the results—the hairy structure—are displayed in Figure 10 along with the previously determined prediction strips of the three observables. It is apparent that most of the 50 realizations fit within the limits given by kriging or are below the maximum value predicted by this technique.

6.2 Coupling to a transmission line

The histogram of the values of the current I at Z_L , set to 100Ω , is shown in Figure 11 along with the fitted pdf—here, a beta distribution:

$$\beta(I) = \frac{\Gamma(\mu + \nu)}{\Gamma(\mu)\Gamma(\nu)} I^{\mu-1} (I - 1)^{\nu-1}, \quad (7)$$

where $\Gamma(x)$ is the Gamma function.

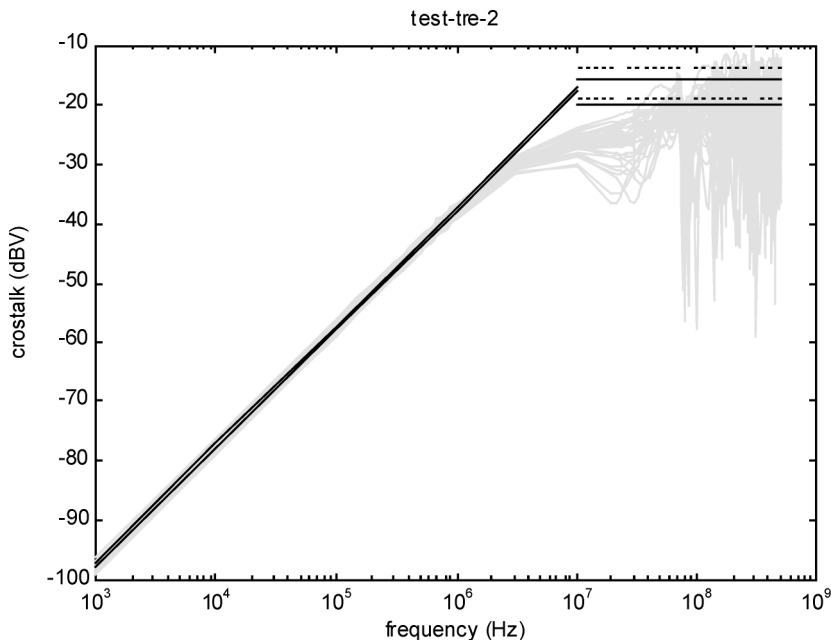


Figure 10 Position of the exact crosstalk voltage of 50 cables (hairy structure) with respect to the prediction strips determined by kriging; $h = 17.59$ mm, $D = 1.29$, and $Z = 100 \Omega$.

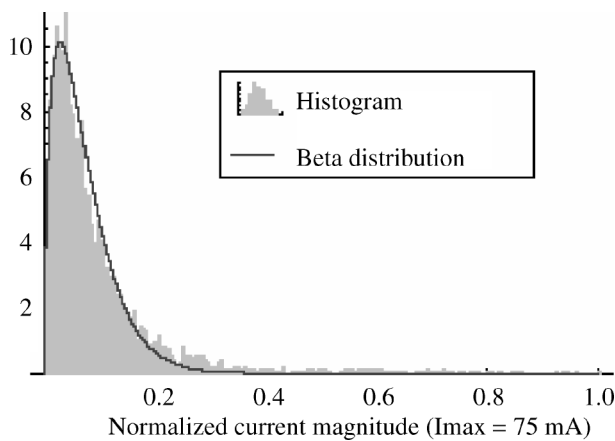


Figure 11 Histogram of the current I and the fitted beta distribution.

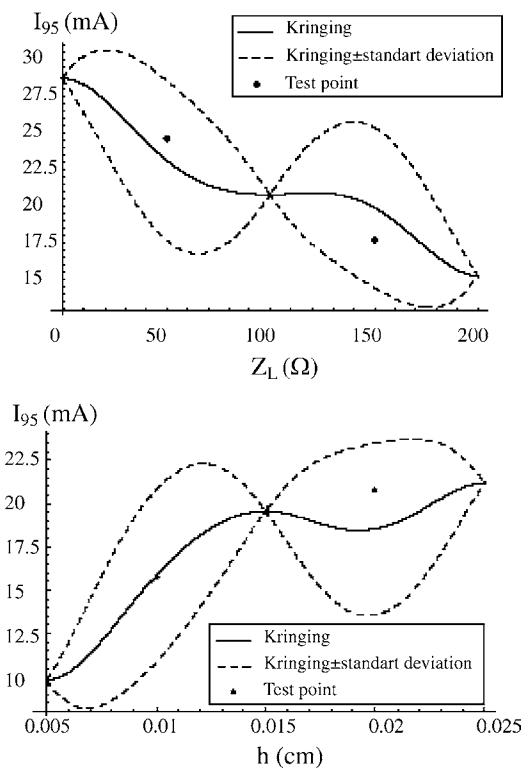


Figure 12 Kriging of I_{95} as a function of Z_L (left) and h (right); The test points (exact values of I_{95}) fall within the so-called “sausage of accuracy,” leading one to consider the predicted (interpolated) values as acceptable.

The histogram was constructed using seven values for each of the three coordinates of the center of the dipole and five values for each of the two angles θ and ϕ , leading to 8575 values of the current. We have also observed that the beta distribution

provides a good fit for all considered values of L , h and Z_L , with different values for μ and ν in each case. It is also possible to chose as observables the two parameters μ and ν of the beta distribution and apply the kriging approach. We could then interpolate the pdf of the current.

Application of kriging to the quantile I_{95} for each factor separately provides the results shown in Figure 12.

7 Conclusion

Statistical tools, when used along with classical modeling techniques to investigate the coupling of an electromagnetic wave to a structure, enhance their efficiency by allowing one to gain information that is otherwise difficult to obtain. They also significantly reduce the cost of a parametric study, whether done numerically or experimentally. This is particularly true of kriging when considered as an interpolation method, which in addition provides an estimate of the accuracy of the interpolated values. Many features of kriging have not been considered here, and would be of considerable interest in EMC problems. Among them are the simultaneous handling of two or more observables and the evaluation of their correlation, as well as the processing of noisy data [8].

The statistical tools presented here, as well as others, have not yet been fully exploited in wave scattering problems, and it is expected that their efficiency increases with the complexity of the problem one is investigating. Among the most promising statistical methods for the analysis of scattering problems are the analysis of variance (ANOVA) and experiment design [4]-[6], [16].

Acknowledgments

The assessment of statistical and probabilistic tools in the study of electromagnetic wave coupling to a structure has been funded by the DGA, a department of the French Ministry of Defense, France Telecom and Renault. The results of the first application were obtained in the frame of the European research program BRITE-EURAM “AUTO-EMC,” and those of the second one are part of a collaboration with France Telecom. This research has also considerably benefited from our collaboration with Professors Marc Hélier and Flavio Canavero, and Drs. Françoise Brouaye, Helene Roussel and Jean Lefebvre.

References

- [1] F. Hoëppe, P.-N. Gineste, B. Demoulin and F. Flourens, Concept de groupe de modes appliqué au calcul de la distribution du champ électromagnétique dans des cavités réverbérantes, *10ème Colloque International sur la CEM*, Clermont-Ferrand, France (March 14-16, 2000), 312–317.
- [2] W.J. Karzas, *Back door coupling of RF (microwave) energy to spacecraft interior cabling*, Interaction Notes 513, ARL, Kirtland AFB, New Mexico, 1994.

- [3] R.M. Bevensee, Probabilistic potential theory applied to electrical engineering problems, *Proc IEEE* **61** (1973), 423–437.
- [4] J. Devore and N. Farnum, *Applied statistics for engineers and scientists*, Duxbury Press, Pacific Grove, CA, US, 1999.
- [5] T.H. Wonnacott and R.J. Wonnacott, *Introductory statistics*, 5th ed, Wiley, New York, NY, USA, 1990.
- [6] G.E.P. Box, W.G. Hunter and J.S. Hunter, *Statistics for experimenters*, Wiley, New York, NY, USA, 1978.
- [7] J. Lefebvre, H. Roussel, E. Walter and W. Tabbara, Prediction from wrong models: The kriging approach, *IEEE Antennas Propagat Mag* **38** (1996), 35–45.
- [8] H. Wackernagel, *Multivariate geostatistics*, Springer, Berlin, Germany, 1995.
- [9] V. Rannou, F. Brouaye, M. Helier and W. Tabbara, Kriging the quantile: Application to a simple transmission line model, *Inverse Problems* **18** (2002), 37–48.
- [10] T. Ryan, *Modern regression methods*, Wiley, New York, NY, US, 1997.
- [11] S. Salio, J. Lefebvre, F. Canavero and W. Tabbara, Statistical description of signal propagation on random bundles of wires, *13th int Zurich symp EMC*, Zurich, Switzerland (February, 1999), 499–504.
- [12] G.A. Edgar, *Measure, topology and fractal geometry*, Springer, Berlin, Germany, 1990.
- [13] C.R. Paul, *Analysis of multiconductor transmission lines*, Wiley, New York, NY, US, 1994.
- [14] F.M. Tesche, M.V. Ianoz and T. Karlsson, *EMC analysis methods and computational models*, Wiley, New York, NY, USA, 1997.
- [15] C.A. Balanis, *Advanced engineering electromagnetics*, Wiley, New York, NY, USA, 1989.
- [16] N. Ribière-Tharaud, M. Hélier, D. Lecointe, R. Chotard and J.-Ch. Bolomey, CDF and quantile: relevant observables related to the common-mode current of a realistic bundle of cables, *14th int Zurich symp EMC*, Zurich, Switzerland (February, 2001), 93–96.



Walid Tabbara, born in Beirut (Lebanon) in 1946, he obtained the Doctorat ès Sciences (Ph.D.) from the University of Paris VI in 1976. He has been a full professor at the same university since 1981. His research is conducted at the Département de Recherche en Électromagnétisme (DRE) of Supélec. From 1987 to 1996 he acted as a Deputy Director of the Laboratoire des Signaux et Systèmes (Supélec-CNRS). From 1984 to 1990 he acted as the French delegate to Commission B of URSI. He is a member of URSI, the Electromagnetics Academy and, the editorial boards of the *Journal of Electromagnetic Waves*

and Applications (JEWA) and of the *Journal of Computer Applications in Engineering Education*. He has organized and/or chaired sessions at national and international conferences. After working for a long period in the field of inverse scattering in electromagnetics and acoustics, his present fields of interest are electromagnetic compatibility (EMC), remote sensing, indoor propagation and high-power microwaves (HPM). His present and past research has led to the publication of 48 papers in international journals and to 158 presentations at international and national conferences.



Veronique Rannou obtained her Ph.D. in 2001 from the University of Paris VI, for her theoretical and experimental work on statistical analysis of electromagnetic waves coupling to transmission lines. She is the author of one paper in *Inverse Problems* and 6 presentations at international conferences. She is now with the French agency of aeronautical research (ONERA) acting as a research engineer in the field of low-frequency radars.



Stefano Salio received a M.S. degree in Electronic Engineering from the Polytechnic University of Turin, Italy, in 1996, and a Ph.D. degree in Electronic and Communications Engineering from the same university in 2000. He is currently with an engineering enterprise, with activity focused on the research and development of localization techniques of mobile terminals over mobile networks. His research interests involves signal processing, advanced localization algorithms and signaling protocols. He has held a teaching assistant positions at the Polytechnic University of Turin. He has published several contributions to international EMC conferences, especially in the field of statistical analysis of crosstalk prediction in and lightning effects on nonuniform bundles of wires.

Elastic Orthonormal Beams and Localized Fields

George N. Borzdov

Abstract

Special types of elastic fields can be generated as beams defined by a given set of orthonormal scalar functions on a two-dimensional or three-dimensional beam manifold. The proposed method enables one to obtain sets of orthonormal beams and various families of localized fields in both isotropic and anisotropic solids. This can also be applied to sound beams in liquids and, by way of illustration, the fields defined by the spherical harmonics are considered. The families of orthonormal beams can be used as functional bases for complex elastic fields, and it is shown that they can induce chiral inhomogeneities in isotropic elastic mediums. The presented localized elastic fields include storms, whirls, and tornadoes, i.e., the localized fields for which the time-averaged energy flux is identically zero, azimuthal, and spiral, respectively. It is shown that the presented fields can be combined into a complex field structure such as an acoustic diffraction grating, which makes them promising tools to control laser radiation.

1 Introduction

A considerable amount of attention has recently been focused on new solutions to the wave equation in a source-free space that describe scalar and vector fields with a high degree of two-dimensional and three-dimensional spatial localization (2D and 3D localized fields). In particular, they include Bessel [1]–[4], Bessel–Gauss [5] and Weber [6] beams possessing high power confinement near the axis and maintaining it over a long propagation range. Regular beams, i.e., superpositions of homogeneous plane waves propagating in a linear medium or free space, are exact solutions of the wave equation. Bessel beams are a typical example of regular beams conserving their shape across any plane orthogonal to the direction of propagation, i.e., having a transverse intensity distribution independent of the propagation distance. Such beams are said to be nondiffracting (or diffraction-free). The total energy flux of a Bessel beam is infinite, for which there can be no physical realization. However, a truncation of this exact solution can give an approximate solution with finite energy flux and good propagation properties [6].

In 1983, Brittingham [7] proposed the problem of searching for specific electromagnetic waves having a three-dimensional pulse structure, being nondispersive for all time, and moving at light velocity in straight lines. This pioneering work lent impetus to a successful search for so-called localized wave solutions both in electrodynamics and acoustics. The obtained solutions include, in particular, focus wave modes [7]–[9], modulated, moving Gaussian pulses [10], electromagnetic missiles [11], acoustic [12–13] and electromagnetic [14–15] directed energy pulse trains (ADEPTs and EDEPTs), moving modified Bessel–Gauss pulses [16], moving pulses with Gaussian localization in both longitudinal and transverse directions [17], etc. It was shown that such moving localized fields can be realized physically; they appear to be launchable from finite-aperture antennas [14–15], [18]. Moreover, the existence of ADEPTs was confirmed with experiments using arrays of ultrasonic transducers in water [12–13]. To realize in full measure localized-wave effects, the use of an array that has independently addressable elements was suggested [19]. An optical implementation of localized wave fields also has attracted considerable interest (see, for example, Refs. 20–21 and references therein). In particular, it was shown [21] that wavelength dispersion of the cone angle of an axicon (axially symmetric optical glass element with two refracting surfaces—conical and flat) and circular diffraction gratings can be used to generate a good approximation to focus wave modes in the optical region.

Planewave expansions provide the most general and fruitful technique (Fourier analysis) for designing and investigating various diffraction-free beams and 3D localized fields. In this framework, it was shown [22] that the source-free focus wave modes can be treated as planewave superpositions of the Whittaker type; i.e., they are composed of homogeneous plane waves propagating in opposite directions. Moreover, the space-time Fourier transform provides a powerful tool for constructing localized solutions to various homogeneous partial differential equations with constant coefficients, such as the wave equation, the damped wave equation and the Klein–Gordon equation [23–24].

In recent works [25]-[28], we have introduced a specific type of electromagnetic field in free space and complex mediums—planewave superpositions defined by a set of orthonormal scalar functions on a 2D or 3D manifold (beam manifold $\mathcal{B}$). The proposed mathematical technique makes it possible to obtain a set of orthonormal beams, normalized to either the energy flux through a given plane (beams with two-dimensional $\mathcal{B}$) or the total energy transmitted through this plane (beams with three-dimensional $\mathcal{B}$), and various families of localized fields. As an illustration, when the fields defined by the spherical harmonics are treated. In this case, the beam manifold is two-dimensional (the unit sphere), and the spherical harmonics specify intensities and initial phases of partial harmonic plane waves (eigenwaves).

The families of orthonormal electromagnetic beams, forming convenient functional bases, can be very useful in the characterization of both complex mediums and fields [29-30]. In particular, this provides a means to generalize the free-space techniques [31]-[34] for characterizing complex mediums to the case of incident beams.

The families of localized electromagnetic fields defined by the spherical harmonics and called, for the sake of brevity, storms, whirls and tornadoes, also possess very interesting properties [26], [28], [35]. They have a small (about several wavelengths) clearly defined core region with maximum intensity of field oscillations and unique spatial distributions of energy densities, energy fluxes and polarization states. For an electromagnetic storm, the time-averaged energy flux vector $\mathbf{S}$ is identically zero at all points. For both whirls and tornadoes, the cylindrical components of $\mathbf{S}$ are independent of the azimuthal angle. Whirls and tornadoes have circular and spiral energy flux lines, respectively. The solutions, which describe electromagnetic whirls moving without dispersion with speed $0 < V < c$ (c is the speed of light in vacuum), finite-energy evolving electromagnetic whirls, and weak gravitational orthonormal beams and localized fields [26], [35], have been found as well.

The mathematical formalism proposed in Ref. 26 can be applied to any scalar, vector, or tensor field that is linear. However, the special features of its application, as well as the properties of the resulting exact solutions to the corresponding wave equation, substantially depend on the properties of the partial eigenwaves. By now, this formalism has been illustrated for the cases of electromagnetic waves in free space [26]-[28], [30], [35] and complex mediums [26]-[28], weak gravitational fields [35], elastic waves in isotropic and anisotropic solids [36]-[39], and sound waves in an ideal liquid [40]. In particular, it has been shown [36], [41] that the presented localized fields can be combined as constructive elements to obtain a complex field structure with desirable properties, such as one-, two-, and three-dimensional elastic [36] and electromagnetic [41] field gratings. They also can be used to induce chiral inhomogeneities in isotropic mediums [42]. All this makes them promising tools to control laser and particle beams.

Another technique for obtaining localized exact solutions of linear field equations, based on the use of plane wave expansion, some notions and tools of the

differential geometry, and the relevant rotation group, was recently proposed [41]. It can be considered as a generalization of the earlier-presented approach [26] that has extended potential for characterizing and designing various localized fields.

In this chapter we discuss the design of ultrasonic orthonormal beams and localized fields in solids and liquids, and we also illustrate some of their characteristic properties.

2 Basic relations

To compose a field from eigenwaves, it is necessary first to calculate their parameters. In this section, we derive necessary relations for ultrasonic eigenwaves in solids and liquids.

2.1 Elastic eigenwaves in an anisotropic medium

2.1.1 Dispersion equation

Let us consider a nondissipative linear elastic medium described by Hooke's law [43]

$$\sigma_{jk} = c_{jklm} \frac{\partial u_m}{\partial x^l}, \quad (1)$$

where $\underline{\underline{\sigma}}$ is the stress tensor, $\mathbf{u}$ is the displacement vector, and the stiffness parameters c_{jklm} satisfy the symmetry relations

$$c_{jklm} = c_{kjl m} = c_{j k m l} = c_{l m j k} = c_{j k l m}^*, \quad j, k, l, m = 1, 2, 3. \quad (2)$$

Here and throughout, summation over repeated indexes is carried out from 1 to 3, and c_{jklm}^* is the complex conjugate of c_{jklm} . For an eigenwave

$$\mathbf{u} = \mathbf{u}_0 \exp[i(\mathbf{k} \cdot \mathbf{r} - \omega t)], \quad (3)$$

with wavevector $\mathbf{k}$ and angular frequency ω , the elastodynamics equation [43]

$$\frac{\partial \sigma_{jk}}{\partial x^k} = \varrho \frac{\partial^2 u_j}{\partial t^2} \quad (4)$$

becomes

$$\underline{\underline{d}}(\mathbf{k}, \mathbf{k}) \mathbf{u}_0 = \varrho \omega^2 \mathbf{u}_0. \quad (5)$$

Here, ϱ is the mass density, and $\underline{\underline{d}}(\mathbf{a}, \mathbf{b})$ denotes the dyadic with the components $[\underline{\underline{d}}(\mathbf{a}, \mathbf{b})]_{jm} = c_{jklm} a_k b_l$. Hence, the wavevector surface is defined by the dispersion equation

$$|\underline{\underline{d}}(\mathbf{k}, \mathbf{k}) - \varrho \omega^2 \underline{\underline{I}}| = 0, \quad (6)$$

where $\underline{\underline{I}}$ is the unit dyadic, and $|\underline{\underline{A}}|$ denotes the determinant of a dyadic $\underline{\underline{A}}$. To calculate the determinant of a sum of two or three dyadics, we use the identity [43]

$$|\underline{\underline{A}} + \underline{\underline{B}}| = |\underline{\underline{A}}| + \text{Tr}(\overline{\underline{\underline{A}}} \underline{\underline{B}}) + \text{Tr}(\underline{\underline{A}} \overline{\underline{\underline{B}}}) + |\underline{\underline{B}}|, \quad (7)$$

where $\underline{\underline{A}}$ and $\underline{\underline{B}}$ are arbitrary dyadics, $\overline{\underline{\underline{A}}}$ is the adjoint dyadic ($\underline{\underline{A}} \overline{\underline{\underline{A}}} = \overline{\underline{\underline{A}}} \underline{\underline{A}} = |\underline{\underline{A}}| \underline{\underline{I}}$), and $\text{Tr}(\underline{\underline{A}})$ denotes the trace of $\underline{\underline{A}}$.

2.1.2 Wavevectors

There are two major approaches to solve the dispersion equation. In the framework of the first, the angular frequency ω and the unit wavevector $\hat{\mathbf{k}} = \mathbf{k}/k$ are given, and so it is sufficient to find the wavenumber k or the phase velocity $v_p = \omega/k$. In this case, (6) and (7) result in the bicubic equation

$$v_p^6 - v_p^4 \text{Tr}(\underline{\underline{A}}) + v_p^2 \text{Tr}(\overline{\underline{\underline{A}}}) - |\underline{\underline{A}}| = 0, \quad (8)$$

where $\underline{\underline{A}} = d(\hat{\mathbf{k}}, \hat{\mathbf{k}})/\varrho$. This approach is convenient to study the properties of eigenwaves in an infinite medium.

The second approach is applied to solve a boundary value problem or to compose a family of orthonormal beams that are normalized to the energy flux through some given plane σ_0 with the unit normal $\mathbf{q}$. It is based on the use of the wavevector surface parametrization by the tangential component $\mathbf{t}$ of $\mathbf{k} = \mathbf{t} + \xi \mathbf{q}$, where $\mathbf{t} \cdot \mathbf{q} = 0$. This parametrization is convenient because the wavevectors of the incident, reflected, and refracted eigenwaves have the same tangential component but different normal components. When $\mathbf{q}$ and $\mathbf{t}$ are given, the normal component ξ of $\mathbf{k}$ is defined by the full sixth-order equation in ξ [37], [43]

$$|\xi^2 \underline{\underline{A}} + \xi \underline{\underline{B}} + \underline{\underline{C}}| \equiv \sum_{n=1}^6 a_n \xi^n + |\underline{\underline{C}}| = 0, \quad (9)$$

where

$$a_1 = \text{Tr}(\overline{\underline{\underline{C}}} \underline{\underline{B}}), \quad a_2 = \text{Tr}(\overline{\underline{\underline{B}}} \underline{\underline{C}} + \overline{\underline{\underline{C}}} \underline{\underline{A}}), \quad (10)$$

$$a_3 = |\underline{\underline{B}}| + \text{Tr}[\underline{\underline{A}} \underline{\underline{B}} \underline{\underline{C}} + \underline{\underline{C}} \underline{\underline{B}} \underline{\underline{A}} + \underline{\underline{A}} \text{Tr}(\underline{\underline{B}}) \text{Tr}(\underline{\underline{C}})] - \text{Tr}[\underline{\underline{A}} \underline{\underline{B}} \text{Tr}(\underline{\underline{C}}) + \underline{\underline{B}} \underline{\underline{C}} \text{Tr}(\underline{\underline{A}}) + \underline{\underline{C}} \underline{\underline{A}} \text{Tr}(\underline{\underline{B}})], \quad (11)$$

$$a_4 = \text{Tr}(\overline{\underline{\underline{A}}} \underline{\underline{C}} + \overline{\underline{\underline{B}}} \underline{\underline{A}}), \quad a_5 = \text{Tr}(\overline{\underline{\underline{A}}} \underline{\underline{B}}), \quad a_6 = |\underline{\underline{A}}|, \quad (12)$$

$$\underline{\underline{A}} = \underline{\underline{d}}(\mathbf{q}, \mathbf{q}), \quad \underline{\underline{C}} = \underline{\underline{d}}(\mathbf{t}, \mathbf{t}) - \varrho \omega^2 \underline{\underline{I}}, \quad (13)$$

$$\underline{\underline{B}} = \underline{\underline{B}}_1 + \underline{\underline{B}}_2, \quad \underline{\underline{B}}_1 = \underline{\underline{d}}(\mathbf{t}, \mathbf{q}), \quad \underline{\underline{B}}_2 = \underline{\underline{d}}(\mathbf{q}, \mathbf{t}). \quad (14)$$

The roots ξ_j ($j = 1, 2, \dots, 6$) of this equation specify all six wavevectors $\mathbf{k}_j = \mathbf{t} + \xi_j \mathbf{q}$, which have the same given tangential component $\mathbf{t}$.

2.1.3 Amplitude orthogonality

After calculation of the wavevector $\mathbf{k}$, one can substitute it in (5) and rewrite the latter as an eigenvector equation with zero eigenvalue: $\underline{\underline{\chi}}\mathbf{u} = \mathbf{0}$, where $\underline{\underline{\chi}} = \underline{\underline{A}} - v_p^2 \underline{\underline{L}}$, and $|\underline{\underline{\chi}}| = 0$. This equation specifies the eigenwave polarization in terms of the displacement vector $\mathbf{u}$ as follows. If $\overline{\underline{\chi}} \neq 0$, the amplitude subspace is one-dimensional, and the eigenwave amplitude is given by $\mathbf{u} = \overline{\underline{\chi}}\mathbf{p}$, where $\mathbf{p}$ is an arbitrary vector. However, if $\underline{\underline{\chi}}$ is a dyad ($\overline{\underline{\chi}} = 0$), the amplitude subspace becomes two-dimensional, and $\mathbf{u}$ is an arbitrary vector normal to $\mathbf{n}_u = \mathbf{p}\underline{\underline{\chi}}$.

To compose the orthonormal beams, it is necessary to use the six-dimensional eigenwave amplitude [36-37]

$$\mathbf{W}_0 = \begin{pmatrix} \mathbf{u} \\ \mathbf{f} \end{pmatrix}, \quad \mathbf{f} = \underline{\underline{\sigma}}\mathbf{q} = i\underline{\underline{d}}(\mathbf{q}, \mathbf{k})\mathbf{u} = i\mathbf{e}_j c_{jklm} q_k k_l u_m, \quad (15)$$

where $\mathbf{k}$ and $\mathbf{u}$ are specified as before, and $\mathbf{e}_j$ are the Cartesian basis vectors. Let ξ_1 be an arbitrary root of (9). Substituting $\mathbf{k}_1 = \mathbf{t} + \xi_1\mathbf{q}$ in (5) and (15), we obtain

$$R\mathbf{W}_1 = \xi_1\mathbf{W}_1, \quad (16)$$

where

$$\mathbf{W}_1 = \begin{pmatrix} \mathbf{u}_1 \\ \mathbf{f}_1 \end{pmatrix}, \quad R = \begin{pmatrix} -\underline{\underline{A}}^{-1}\underline{\underline{B}}_2 & -i\underline{\underline{A}}^{-1} \\ i(\underline{\underline{B}}_1\underline{\underline{A}}^{-1}\underline{\underline{B}}_2 - \underline{\underline{C}}) & -\underline{\underline{B}}_1\underline{\underline{A}}^{-1} \end{pmatrix}. \quad (17)$$

For real values of $\mathbf{t}$, the dyadics $\underline{\underline{A}}$, $\underline{\underline{B}}_1$, $\underline{\underline{B}}_2$ and $\underline{\underline{C}}$ have the properties

$$\underline{\underline{A}}^\dagger = \underline{\underline{A}}, \quad \underline{\underline{C}}^\dagger = \underline{\underline{C}}, \quad \underline{\underline{B}}_1^\dagger = \underline{\underline{B}}_2, \quad (18)$$

where $\underline{\underline{A}}^\dagger$ is the Hermitian conjugate of $\underline{\underline{A}}$. As a result, the matrix R satisfies the identity

$$R^\dagger = Q_0 R Q_0, \quad Q_0 = \begin{pmatrix} 0 & -i\underline{\underline{I}} \\ i\underline{\underline{I}} & 0 \end{pmatrix}. \quad (19)$$

Hence, when $\xi_k \neq \xi_j^*$, (16) results in the amplitude orthogonality relation

$$\mathbf{W}_j^\dagger Q_0 \mathbf{W}_k \equiv i(\mathbf{f}_j^* \cdot \mathbf{u}_k - \mathbf{u}_j^* \cdot \mathbf{f}_k) = 0, \quad j \neq k. \quad (20)$$

Such orthogonality relations are of great importance in composing various families of orthonormal beams [26], [36-37], [44].

2.2 Elastic eigenwaves in an isotropic medium

In an isotropic elastic medium, Hooke's law (1) becomes [43]

$$\underline{\underline{\sigma}} = \lambda_L (\text{div } \mathbf{u}) \underline{\underline{I}} + 2\mu_L \underline{\underline{\gamma}}, \quad (21)$$

where λ_L and μ_L are the so-called Lamé constants, while $\underline{\underline{\gamma}}$ is the Cauchy strain tensor with the components

$$\gamma_{jk} = \frac{1}{2} \left(\frac{\partial u_j}{\partial x_k} + \frac{\partial u_k}{\partial x_j} \right). \quad (22)$$

In this medium, the eigenwave $\mathbf{u}$ of equation (3) is also described by

$$\underline{\underline{\gamma}} = \frac{ik}{2} (\hat{\mathbf{k}} \otimes \mathbf{u} + \mathbf{u} \otimes \hat{\mathbf{k}}), \quad (23)$$

$$\underline{\underline{\sigma}} = ik [\lambda_L (\hat{\mathbf{k}} \cdot \mathbf{u}) \underline{\underline{I}} + \mu_L (\hat{\mathbf{k}} \otimes \mathbf{u} + \mathbf{u} \otimes \hat{\mathbf{k}})], \quad (24)$$

$$\mathbf{f} = ik [\lambda_L (\hat{\mathbf{k}} \cdot \mathbf{u}) \mathbf{q} + \mu_L (\mathbf{u} \cdot \mathbf{q}) \hat{\mathbf{k}} + \mu_L (\hat{\mathbf{k}} \cdot \mathbf{q}) \mathbf{u}], \quad (25)$$

$$\underline{\underline{\Delta}} = v_2^2 \underline{\underline{I}} + (v_1^2 - v_2^2) \hat{\mathbf{k}} \otimes \hat{\mathbf{k}}, \quad (26)$$

where $\otimes$ symbolizes the tensor product, and $v_1 = \sqrt{(\lambda_L + 2\mu_L)/\varrho}$ and $v_2 = \sqrt{\mu_L/\varrho}$ are the velocities of longitudinal and transverse elastic waves, respectively.

In subsequent sections, we characterize time-harmonic eigenwave superpositions in elastic solids by the time-averaged kinetic (w_K) and elastic (w_E) energy densities, and the time-averaged energy flux density vector $\mathbf{S}$. They can be calculated as [43]

$$w_K = \frac{1}{4} \varrho \omega^2 |\mathbf{u}|^2, \quad w_E = \frac{1}{4} \text{Re} (\sigma_{jk} \gamma_{jk}^*), \quad \mathbf{S} = \frac{\omega}{2} \text{Re} (i \underline{\underline{\sigma}}^* \mathbf{u}), \quad (27)$$

where Re denotes the real part of the corresponding complex scalars and vectors.

2.3 Sound eigenwaves in an ideal liquid

This chapter is primarily concerned with elastic fields in solids, but most of the theory can be readily extended to sound waves in an ideal liquid as follows. In the linear approximation, the velocities $\mathbf{v}$ of fluid particles are far less in magnitude than the sound velocity c_0 , and variations of pressure $p' = p - p_0$ and density $\varrho' = \varrho - \varrho_0$ are far smaller than the equilibrium values p_0 and ϱ_0 . Therefore, for an eigenwave, the continuity equation and the Euler equation reduce to [44]

$$\omega \varrho' = \varrho_0 \mathbf{k} \cdot \mathbf{v}, \quad \omega \varrho_0 \mathbf{v} = p' \mathbf{k}, \quad (28)$$

where $p' = c_0^2 \varrho'$. The compatibility condition of this system gives the dispersion equation $\mathbf{k}^2 - \omega^2/c_0^2 = 0$ that has two different solutions

$$\mathbf{k}_j = \mathbf{t} + \xi_j \mathbf{q}, \quad j = 1, 2, \quad \xi_{1,2} = \pm \sqrt{\omega^2/c_0^2 - \mathbf{t}^2}, \quad (29)$$

while the tangential component $\mathbf{t}$ of $\mathbf{k}$ is assumed given. Note that $\mathbf{t}^2 = \mathbf{t} \cdot \mathbf{t}$ does not necessarily equal $\mathbf{t} \cdot \mathbf{t}^*$. From (28) and (29) we obtain

$$R\mathbf{W}_1 = \xi_1 \mathbf{W}_1, \quad \mathbf{W}_1 = \begin{pmatrix} p'_1 \\ \mathbf{q} \cdot \mathbf{v}_1 \end{pmatrix} \equiv \begin{pmatrix} p'_1 \\ p'_1 \xi_1 / (\omega \varrho_0) \end{pmatrix}, \quad (30)$$

where

$$R = \begin{pmatrix} 0 & \omega \varrho_0 \\ A_0 & 0 \end{pmatrix}, \quad A_0 = \frac{1}{\varrho_0} \left(\frac{\omega}{c_0^2} - \frac{\mathbf{t}^2}{\omega} \right). \quad (31)$$

The second eigenwave is described by the similar relations. For real values of $\mathbf{t}$, R satisfies the identity

$$R^\dagger = \varrho_0 R \varrho_0, \quad \varrho_0 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}. \quad (32)$$

Therefore, when $\xi_1^* \neq \xi_2$, (30) results in the orthogonality relation

$$\mathbf{W}_1^\dagger \varrho_0 \mathbf{W}_2 \equiv \mathbf{q} \cdot (p_1'^* \mathbf{v}_2 + p_2' \mathbf{v}_1^*) = 0. \quad (33)$$

This relation is used to compose various families of ultrasonic orthonormal beams [40], [44].

2.4 Superpositions of eigenwaves

To compose an ultrasonic field of eigenwaves in a linear medium, one must specify all their properties. In the case of elastic waves in an anisotropic elastic medium, polarizations of eigenwaves are specified by the medium itself, and one has to set propagation directions, intensities, and initial phases only.

The eigenwave superpositions (beams) defined by a set of orthonormal functions (u_n) on a two- or three-dimensional real manifold $\mathcal{B}_u$ can be written as [26]

$$\mathbf{W}_n(\mathbf{r}, t) = \int_{\mathcal{B}} \exp\{i[\mathbf{k}(b) \cdot \mathbf{r} - \omega(b)t]\} u_n(b) v(b) \mathbf{W}(b) d\mathcal{B}, \quad (34)$$

where b is a point of the beam manifold $\mathcal{B}$, which is a subset of $\mathcal{B}_u$ with nonvanishing values of function $\mathbf{W}' = v(b) \mathbf{W}(b)$. The amplitude function $\mathbf{W} = \mathbf{W}(b)$ must

be given in an explicit normalized form for each specific ultrasonic field (see Sections 2.1–2.3). Intensities and phases of all eigenwaves forming $\mathbf{W}_n$ are specified by the same complex scalar function u_n satisfying the condition

$$\langle u_m | u_n \rangle \equiv \int_{\mathcal{B}_u} u_m^*(b) u_n(b) d\mathcal{B} = \delta_{mn}, \quad (35)$$

where δ_{mn} is the Kronecker delta. The so-called beam state function [26] $v = v(b)$ is used to specify general properties of the whole family of beams $\mathbf{W}_n$.

By setting functions $\mathbf{k} = \mathbf{k}(b)$, $\mathbf{W} = \mathbf{W}(b)$, and $v = v(b)$ in various ways, one can obtain a multitude of specific fields (see, for example, Refs. 26–41, among them orthonormal beams satisfying the condition

$$\langle \mathbf{W}_m | Q | \mathbf{W}_n \rangle \equiv \int_{\sigma_0} \mathbf{W}_m^\dagger(\mathbf{r}, t) Q \mathbf{W}_n(\mathbf{r}, t) d\sigma_0 = N_Q \delta_{mn}, \quad (36)$$

where σ_0 is either a two- or a three-dimensional manifold, Q is some Hermitian operator, and N_Q is the normalization constant.

The time-harmonic beams with two-dimensional $\mathcal{B}$ can be written as

$$\mathbf{W}_n(\mathbf{r}, t) = \exp(-i\omega t) \int_{\mathcal{B}} \exp[i\mathbf{k}(b) \cdot \mathbf{r}] u_n(b) v(b) \mathbf{W}(b) d\mathcal{B}. \quad (37)$$

They become orthonormal, provided the following conditions are met [26]:

- i) σ_0 is a plane with unit normal $\mathbf{q}$, passing through the point $\mathbf{r} = \mathbf{0}$;
- ii) the tangential component $\mathbf{t}(b)$ of $\mathbf{k}(b)$ is real for all $b \in \mathcal{B}$, and the mapping $b \mapsto \mathbf{t}(b)$ is one-to-one (injective); and
- iii) $\mathcal{B} = \mathcal{B}_u$, with the beam state (orthonormalizing) function given by

$$v(b) = \frac{1}{2\pi} \sqrt{\frac{N_Q J(b)}{g(b) \mathbf{W}^\dagger(b) Q \mathbf{W}(b)}}. \quad (38)$$

Here, $J(b) = D(t^j)/D(b^l)$ is the Jacobian determinant of the mapping $b \mapsto \mathbf{t}(b)$, calculated in terms of the local coordinate systems $(b^l, l = 1, 2)$ and $(t^j, j = 1, 2)$, and $d\mathcal{B} = g(b) db^1 db^2$, where $g(b)$ is defined by the metric properties of $\mathcal{B}$.

In the case of elastic waves in solids (see Section 2.1), with $Q = (\omega/4) Q_0$ and Q_0 of (19), the condition $\langle \mathbf{W}_n | Q | \mathbf{W}_n \rangle = N_Q$ normalizes the time-averaged elastic beam energy flux through the plane σ_0 :

$$\langle \mathbf{W}_n | Q | \mathbf{W}_n \rangle = -\frac{1}{4} \int_{\sigma_0} (\mathbf{v}^* \cdot \mathbf{f} + \mathbf{v} \cdot \mathbf{f}^*) d\sigma_0 = N_Q, \quad (39)$$

where the relation $\mathbf{v} = -i\omega \mathbf{u}$ between the velocity $\mathbf{v}$ and the displacement $\mathbf{u}$ has been taken into account. Similarly, in the case of sound waves in liquids (see Section 2.3), with $Q = Q_0/4$ and Q_0 of (32), this condition normalizes the time-

averaged sound energy flux as

$$\langle \mathbf{W}_n | Q | \mathbf{W}_n \rangle = \frac{1}{4} \int_{\sigma_0} (p'^* \mathbf{v} + p' \mathbf{v}^*) \cdot \mathbf{q} d\sigma_0 = N_Q. \quad (40)$$

We assume in the following sections that $\mathbf{q} = \mathbf{e}_3$.

2.5 Fields defined by spherical harmonics

The presented general relations make it possible to apply the described approach to both ultrasonic fields in solids and liquids. We restrict our illustrations to the time-harmonic elastic fields in an isotropic solid, defined by the spherical harmonics Y_j^s as

$$\begin{aligned} \mathbf{W}_j^s(\mathbf{r}, t) = & \exp(-i\omega t) \int_0^{2\pi} d\varphi \int_0^{\theta_2} \exp[i\mathbf{k}(\theta, \varphi) \cdot \mathbf{r}] \\ & \times Y_j^s(\theta, \varphi) v(\theta, \varphi) \mathbf{W}(\theta, \varphi) \sin \theta d\theta. \end{aligned} \quad (41)$$

For these fields, $\mathcal{B}_u$ is a unit sphere ($\mathcal{B}_u = S^2$), $\mathcal{B}$ is its zone with $\theta \in [0, \theta_2]$ and $\varphi \in [0, 2\pi]$, and $d\mathcal{B} = \sin \theta d\theta d\varphi$. The value $Y_j^s(\theta, \varphi)$ of the function Y_j^s at given θ and φ specifies the magnitude and the phase of an eigenwave with the unit wave normal

$$\hat{\mathbf{k}} = \mathbf{e}_r = \sin \theta' (\mathbf{e}_1 \cos \varphi + \mathbf{e}_2 \sin \varphi) + \mathbf{e}_3 \cos \theta', \quad (42)$$

where $\mathbf{e}_r$ is the radial basis vector of the spherical coordinate system, $\theta' = \kappa_0 \theta$, and parameter κ_0 satisfies the condition $0 < \kappa_0 \leq 1$. These fields are composed of plane waves propagating in the solid angle $\Omega = 2\pi(1 - \cos \kappa_0 \theta_2)$. Further illustrations for various ultrasonic fields can be found elsewhere [38]–[40], [42].

2.5.1 Classification

The ultrasonic fields treated in this chapter can be classified in the following manner [40].

- 1) Orthonormal beams with $\theta_2 = \pi/2$ and $\kappa_0 = 1$ (Type I). They are formed from eigenwaves propagating into a given half-space. In this case, the beam manifold is the northern hemisphere S_N^2 of the unit sphere S^2 ($\mathcal{B} = S_N^2$).
- 2) Orthonormal beams with $\theta_2 = \pi$ and $\kappa_0 \leq 1/2$ (Type II). For these beams, the beam manifold is the unit sphere ($\mathcal{B} = S^2$).
- 3) Three-dimensional standing waves with $\theta_2 = \pi$, $\kappa_0 = 1$, and $\mathcal{B} = S^2$. They are formed from eigenwaves of all possible propagation directions. The family of these waves consists of storms, defined by the zonal spherical harmonics Y_j^0 , and whirls defined by the other Y_j^s ($s \neq 0$).
- 4) Localized fields with $\pi/2 < \theta_2 < \pi$ and $\kappa_0 = 1$. These fields include tornadoes, which are defined by Y_j^s with $s \neq 0$.

By integrating the time-harmonic localized fields with respect to the frequency as

$$\widehat{\mathbf{W}}_j^s(\mathbf{r}, t) = \frac{1}{2\Delta\omega} \int_{\omega_-}^{\omega_+} \mathbf{W}_j^s(\mathbf{r}, t, \omega) d\omega, \quad (43)$$

where $\Delta\omega = (\omega_+ - \omega_-)/2 \ll (\omega_+ + \omega_-)/2$, one can also obtain finite-energy evolving ultrasonic storms, whirls, and tornadoes.

2.5.2 Series expansion

The fields under consideration consist of homogeneous eigenwaves. They can be expanded in series by using the Rayleigh formula [45]

$$e^{i\mathbf{k}\cdot\mathbf{r}} = 4\pi \sum_{l=0}^{+\infty} i^l j_l(kr) \sum_{m=-l}^l Y_l^{m*}(\hat{\mathbf{k}}) Y_l^m(\hat{\mathbf{r}}), \quad (44)$$

where

$$\hat{\mathbf{r}} = \mathbf{r}/r = \sin\gamma (\mathbf{e}_1 \cos\psi + \mathbf{e}_2 \sin\psi) + \mathbf{e}_3 \cos\gamma, \quad (45)$$

$$Y_l^m(\hat{\mathbf{k}}) \equiv Y_l^m(\theta', \varphi), \quad Y_l^m(\hat{\mathbf{r}}) \equiv Y_l^m(\gamma, \psi), \quad (46)$$

$$Y_l^m(\theta, \varphi) = N_{lm} P_l^{|m|}(\cos\theta) e^{im\varphi}, \quad (47)$$

$$N_{lm} = \sqrt{\frac{(2l+1)(l-|m|)!}{4\pi(l+|m|)!}}. \quad (48)$$

Here, $\hat{\mathbf{k}}$ is given by (42), while spherical $P_l^m(\cos\theta)$ and $j_l(kr)$ are the associated Legendre and Bessel functions, respectively [46].

In an anisotropic medium, substitution of the Rayleigh formula in (41) yields an expansion of the form [26]

$$\mathbf{W}_j^s(\mathbf{r}, t) = e^{-i\omega t} \sum_{l=0}^{+\infty} i^l \sum_{m=-l}^l Y_l^m(\hat{\mathbf{r}}) \mathbf{W}_{lj}^{ms}(r), \quad (49)$$

where vector coefficient functions $\mathbf{W}_{lj}^{ms}$ depend only on radius r . In an isotropic medium, the field is completely characterized by coordinate-independent vector coefficients $\mathbf{W}_{lj}^{ms}$.

To find these coefficients, we extensively use a family of functions $I_j^{sm}[f]$ defined by the relations:

$$\begin{aligned} & \int_0^{2\pi} d\varphi \int_0^{\pi/2} \exp\{i[\mathbf{k}\mathbf{r} \cdot \mathbf{e}_r(\theta, \varphi) + n\varphi]\} Y_j^s(\theta, \varphi) f(\theta) \sin\theta d\theta \\ &= \exp[i(s+n)\psi] I_j^{s+s+n}[f](r, \gamma), \end{aligned} \quad (50)$$

$$I_j^{sm}[f] = I_j^{sm}[f](r, \gamma) = \sum_{l=|m|}^{+\infty} i^l j_l(kr) P_l^{|m|}(\cos \gamma) \mathcal{P}_{jl}^{sm}[f], \quad (51)$$

$$\mathcal{P}_{jl}^{sm}[f] = 8\pi^2 N_{js} N_{lm}^2 \int_0^{\pi/2} P_j^{|s|}(\cos \theta) P_l^{|m|}(\cos \theta) f(\theta) \sin \theta d\theta, \quad (52)$$

where $f = f(\theta)$ is a scalar function of the polar angle θ , and n is an integer.

Thus, the complex-valued scalar function $I_j^{sm}[f] = I_j^{sm}[f](r, \gamma)$ is specified by the spherical harmonic Y_j^s , the integer m , and the scalar function f . The notations emphasize the fact that $I_j^{sm}[f]$ at fixed r and γ as well as coefficient $\mathcal{P}_{jl}^{sm}[f]$ are functionals of f . For any given f , $I_j^{sm}[f]$ is a function of r and γ , whereas $\mathcal{P}_{jl}^{sm}[f]$ is a constant. We omit the arguments (r, γ) where appropriate. The real and imaginary parts of $I_j^{sm}[f]$ can be separated as

$$I_j^{sm}[f] = i^{|m|} (J_{j0}^{sm}[f] + i J_{j1}^{sm}[f]), \quad (53)$$

where

$$\begin{aligned} J_{jp}^{sm}[f] &= J_{jp}^{sm}[f](r, \gamma) \\ &= \sum_{v=0}^{+\infty} (-1)^v j_{|m|+2v+p}(kr) P_{|m|+2v+p}^{|m|}(\cos \gamma) \mathcal{P}_{j|m|+2v+p}^{s|m|}[f]. \end{aligned} \quad (54)$$

Additional information on these functions can be found elsewhere [26].

2.6 Photoelasticity in an isotropic medium

Let us consider a photoelastic homogeneous isotropic medium characterized by refractive index n_0 in the absence of elastic fields. In the presence of elastic waves, this medium becomes inhomogeneous, nonstationary and anisotropic; i.e., it is described by the permittivity tensor field $\underline{\underline{\epsilon}} = \underline{\underline{\epsilon}}(\mathbf{r}, t)$. Generally, the induced anisotropy is very small, so that the inverse tensor $\underline{\underline{\epsilon}}^{-1}$ can be written as [47]

$$\underline{\underline{\epsilon}}^{-1} = \frac{1}{n_0^2} \underline{\underline{I}} + \underline{\underline{A}}, \quad (55)$$

where

$$\underline{\underline{A}} = p_{12}(\text{div } \mathbf{u}) \underline{\underline{I}} + 2p_{44} \underline{\underline{\gamma}}, \quad (56)$$

and p_{12} and p_{44} are the photoelastic constants. Hence, the permittivity tensor $\underline{\underline{\epsilon}}$ and the deformation tensor $\underline{\underline{\gamma}}$ have the same system of eigenvectors, while their eigenvalues ϵ_j and γ_j are related as

$$\frac{1}{\epsilon_j} = \frac{1}{n_0^2} + p_{12} \text{div } \mathbf{u} + 2p_{44} \gamma_j, \quad j = 1, 2, 3. \quad (57)$$

In the case of shear waves, i.e., superpositions of transverse elastic eigenwaves, $\text{div } \mathbf{u} = \gamma_{11} + \gamma_{22} + \gamma_{33} = 0$. As a result, the perturbation $\underline{\underline{A}}$ of $\underline{\underline{\epsilon}}^{-1}$, caused by these waves, reduces to $\underline{\underline{A}} = 2p_{44}\underline{\underline{\gamma}}$. Since the perturbation $\underline{\underline{A}}$ is small, the local values of refractive indexes for a plane optical wave with unit wave normal $\mathbf{n}$ can be approximated as

$$n_{\pm} = n_0 - n_0^3 p_{44} \beta_{\pm}, \quad (58)$$

where $\beta_{\pm}$ are the eigenvalues of the tensor

$$\underline{\underline{\beta}} = -\mathbf{n}^{\times} \underline{\underline{\gamma}} \mathbf{n}^{\times}. \quad (59)$$

Here, $\mathbf{n}^{\times}$ is an antisymmetric tensor dual to $\mathbf{n}$. Since $\mathbf{n}^{\times} \mathbf{E} = \mathbf{n} \times \mathbf{E}$ and $\underline{\underline{I}} = \mathbf{e}_j \otimes \mathbf{e}_j$, $\mathbf{n}^{\times}$ can also be denoted by $\mathbf{n} \times \underline{\underline{I}} \equiv (\mathbf{n} \times \mathbf{e}_j) \otimes \mathbf{e}_j$. The components of $\mathbf{n}^{\times}$ are uniquely defined by the components $n_j = \mathbf{n} \cdot \mathbf{e}_j$ of $\mathbf{n}$ and can be written in the matrix form as

$$\begin{pmatrix} 0 & -n_3 & n_2 \\ n_3 & 0 & -n_1 \\ -n_2 & n_1 & 0 \end{pmatrix}. \quad (60)$$

In subsequent sections, we present some graphic illustrations for the case of shear waves in fused quartz that has parameters $n_0 = 1.46$, $v_1 = 5969$ m/s, $v_2 = 3753$ m/s, $\varrho = 2.203$ g/cm³, $p_{12} = 0.27$ and $p_{44} = -0.0745$. In all illustrations, the elastic waves are time-harmonic fields of frequency 0.5 GHz. They are described in terms of the dimensionless coordinates $x' = x^1/\lambda$, $y' = x^2/\lambda$, $z' = x^3/\lambda$ and $R' = R/\lambda$, where $\mathbf{r} = x^i \mathbf{e}_i$; R is the radial cylindrical coordinate, and λ is an appropriate wavelength.

3 Superpositions of longitudinal eigenwaves

3.1 Orthonormal beams of Type I

Let us first consider orthonormal beams $\mathbf{W}_j^s$ with $\theta_2 = \pi/2$ and $\kappa_0 = 1$, composed of planewaves propagating into a solid angle $\Omega = 2\pi$. A longitudinal elastic eigenwave has the one-dimensional amplitude subspace and its displacement vector $\mathbf{u}$ can be written as $\mathbf{u} = (\mathbf{u} \cdot \hat{\mathbf{k}}) \hat{\mathbf{k}}$. With $\kappa_0 = 1$, we have $\theta' = \theta$, and the amplitude function becomes

$$\mathbf{W}(\theta, \varphi) \equiv \begin{pmatrix} \mathbf{u} \\ \mathbf{f} \end{pmatrix} = \begin{bmatrix} \mathbf{e}_r \\ ik(\lambda_L \mathbf{e}_3 + 2\mu_L \cos \theta \mathbf{e}_r) \end{bmatrix}, \quad (61)$$

where $k = 2\pi/\lambda = \omega/v_1$ is the wavenumber. In this case, the orthonormalizing function $v = v(\theta, \varphi)$ reduces to a constant.

From (41) and (61) we find the displacement $\mathbf{u}$ and the force density $\mathbf{f}$ of the ultrasonic beam $\mathbf{W}_j^s$ as [38]

$$\mathbf{u} = v_1 e^{i(s\psi - \omega t)} \{ \mathbf{e} I_j^{s-1}[\sin] + \mathbf{e}^* I_j^{s+1}[\sin] + \mathbf{e}_3 I_j^{s,s}[\cos] \}, \quad (62)$$

$$\begin{aligned} \mathbf{f} = ikv_1 e^{i(s\psi - \omega t)} \{ \mu_L \mathbf{e} I_j^{s-1}[\sin \circ 2] + \mu_L \mathbf{e}^* I_j^{s+1}[\sin \circ 2] \\ + \mathbf{e}_3 I_j^{s,s}[\lambda_L + 2\mu_L \cos^2] \}, \end{aligned} \quad (63)$$

where

$$v_1 = \frac{1}{\pi} \sqrt{\frac{N_Q}{\varrho v_1^3}}, \quad \mathbf{e} = (\mathbf{e}_R + i\mathbf{e}_A)/2, \quad (64)$$

$$\mathbf{e}_R = \mathbf{e}_1 \cos \psi + \mathbf{e}_2 \sin \psi, \quad \mathbf{e}_A = -\mathbf{e}_1 \sin \psi + \mathbf{e}_2 \cos \psi, \quad (65)$$

$$\mathbf{r} = R\mathbf{e}_R + z\mathbf{e}_3, \quad R = r \sin \gamma, \quad z = r \cos \gamma; \quad (66)$$

R , ψ , and z are the cylindrical coordinates and r , γ , and ψ are the spherical coordinates of the point with radius vector $\mathbf{r}$. Here, $f \circ g$ denotes the composition of functions f and g . In particular, to calculate $I_j^{s \pm 1}[\sin \circ 2]$, it is necessary to set $f(\theta) = \sin 2\theta$ and $m = s \pm 1$ in (51) and (52). Similarly, to find $I_j^{s,s}[\lambda_L + 2\mu_L \cos^2]$, one must set $f(\theta) = \lambda_L + 2\mu_L \cos^2 \theta$ and $m = s$. The vector fields $\mathbf{u}$ and $\mathbf{f}$, the tensor fields $\underline{\underline{\gamma}}$ and $\underline{\underline{\sigma}}$, as well as other characteristics of the beam $\mathbf{W}_j^s$ defined the spherical harmonic Y_j^s , naturally, depend on j and s . Since this dependence is clearly indicated in the right side of the corresponding expressions [see, for example, (62) and (63)], instead of notations of the form $\mathbf{u}_j^s$, $\underline{\underline{\gamma}}_j^s$, w_K^s , and so on, here and throughout we use more convenient, simplified notations: $\mathbf{u}$, $\underline{\underline{\gamma}}$, w_K , and so on.

Using (23), (24), and (61), we can calculate the tensors $\underline{\underline{\gamma}}$ and $\underline{\underline{\sigma}}$ for all partial eigenwaves and then replace $\mathbf{W}(\theta, \varphi)$ in (41) by the obtained tensor amplitude functions $\underline{\underline{\gamma}}(\theta, \varphi)$ and $\underline{\underline{\sigma}}(\theta, \varphi)$. This gives, on integration, the deformation and the stress tensor fields of the beam $\mathbf{W}_j^s$ as

$$\underline{\underline{\gamma}} = ikv_1 e^{i(s\psi - \omega t)} \underline{\underline{\gamma}}_0, \quad \underline{\underline{\sigma}} = ikv_1 e^{i(s\psi - \omega t)} (\lambda_L \underline{\underline{I}} I_j^{s,s}[1] + 2\mu_L \underline{\underline{\gamma}}_0), \quad (67)$$

where

$$\begin{aligned} \underline{\underline{\gamma}}_0 = \underline{\underline{\rho}} I_j^{s-2}[\sin^2] + \underline{\underline{\rho}}^* I_j^{s+2}[\sin^2] + \underline{\underline{\rho}}_1 I_j^{s,s}[\sin^2] + \underline{\underline{\rho}}_2 I_j^{s,s-1}[\sin \circ 2] \\ + \underline{\underline{\rho}}_2^* I_j^{s+1}[\sin \circ 2] + \underline{\underline{\rho}}_3 I_j^{s,s}[\cos^2], \end{aligned} \quad (68)$$

$$\underline{\underline{\rho}} = \mathbf{e} \otimes \mathbf{e}, \quad \underline{\underline{\rho}}_1 = \mathbf{e} \otimes \mathbf{e}^* + \mathbf{e}^* \otimes \mathbf{e} = \frac{1}{2}(\underline{\underline{I}} - \underline{\underline{\rho}}_3), \quad (69)$$

$$\underline{\underline{\rho}}_2 = \frac{1}{2}(\mathbf{e} \otimes \mathbf{e}_3 + \mathbf{e}_3 \otimes \mathbf{e}), \quad \underline{\underline{\rho}}_3 = \mathbf{e}_3 \otimes \mathbf{e}_3. \quad (70)$$

Upon substitution of $\mathbf{u}$, $\underline{\underline{\gamma}}$, and $\underline{\underline{\sigma}}$ into (27), one can find the normalized kinetic $w'_K = w_K/w_0$ and elastic $w'_E = w_E/w_0$ energy densities, as well as the normalized energy flux density vector $\mathbf{S}' = \mathbf{S}/S_0$, where $S_0 = N_Q/\lambda^2$ and $w_0 = S_0/v_1$. In particular, the normal component $S'_N = S_3/S_0$ of $\mathbf{S}'$ is given by [36], [38]

$$S'_N = \frac{1}{\lambda_L + 2\mu_L} \sum_{p=0}^1 \{ \mu_L (J_{jp}^{ss-1}[\sin] J_{jp}^{ss-1}[\sin \circ 2] + J_{jp}^{ss+1}[\sin] J_{jp}^{ss+1}[\sin \circ 2]) + 2J_{jp}^{ss}[\cos] J_{jp}^{ss}[\lambda_L + 2\mu_L \cos^2] \}. \quad (71)$$

It is essential that the functions $I_j^m[f]$ and $J_{jp}^m[f]$ are independent of the azimuthal angle ψ . They depend on r and γ or, in other words, R and z . On the other hand, vectors $\mathbf{e}_R$, $\mathbf{e}_A$ and, hence, basis elements $\mathbf{e}$ and $(\underline{\underline{\rho}}, \underline{\underline{\rho}}_1, \underline{\underline{\rho}}_2)$ for vector $(\mathbf{u}, \mathbf{f})$ and tensor $(\underline{\underline{\gamma}}, \underline{\underline{\sigma}})$ fields, respectively, depend on ψ only. Because of the factor $\exp[i(s\psi - \omega t)]$, the obtained expressions describe the fields rotating with the angular velocity $\omega_{rot} = \omega/s$, provided $s \neq 0$. The radial S'_R , the azimuthal S'_A , and the normal S'_N cylindrical components of $\mathbf{S}'$ as well as both energy densities w'_K and w'_E are independent of ψ . For the beams defined by the zonal spherical harmonics ($s = 0, j = 0, 1, \dots$), both $\mathbf{u}$ and $\mathbf{S}$ lie in the meridional planes.

All properties of the beams under consideration, in particular, spatial distribution of energy fluxes substantially depend on the indexes j and s of the spherical harmonic Y_j^s . Let us consider, for example, the function $S'_N = S'_N(R, z)$ for the lower modes with $j = 1, 2, 3$ and $s = 0, \dots, j$. For $j = 0, s = 0$ and $j = 1, s = 0, 1$, this function reaches its maximum in the plane $z = 0$ at $R = 0$. For $j = 2, s = 1, 2$ and $j = 3, s = 2, 3$, it also peaks in the plane $z = 0$, but at some distance $R_{\max} \neq 0$ from the z axis. The larger j is, the larger is $R_{\max}$. At $j = 2, 3$ and $s \neq 0$, S'_N vanishes on this axis. However, for $j = 2, 3$ and $s = 0$, it has two major maximums in two symmetrically placed planes $z = \pm z_{\max} \neq 0$ at $R = 0$. Figure 1 illustrates the transition stage ($j = 3, s = 1$) between these two types of S'_N localization.

3.2 Orthonormal beams of Type II

The spherical harmonics Y_j^s constitute a complete orthonormal system on the unit sphere S^2 . However, for the fields treated in the previous section, the beam manifold $\mathcal{B}$ is the northern hemisphere S_N^2 of S^2 . As a consequence, these fields form two separate sets of orthonormal beams, defined by the spherical harmonics Y_j^s with even and odd j , respectively. It may be advantageous to obtain a complete system of orthonormal elastic beams $\mathbf{W}_j^s$, defined by the whole set of spherical harmonics Y_j^s , for which $\langle \mathbf{W}_j^s | Q | \mathbf{W}_{j'}^{s'} \rangle = 0$ if $j' \neq j$ and/or $s' \neq s$. To this end, let

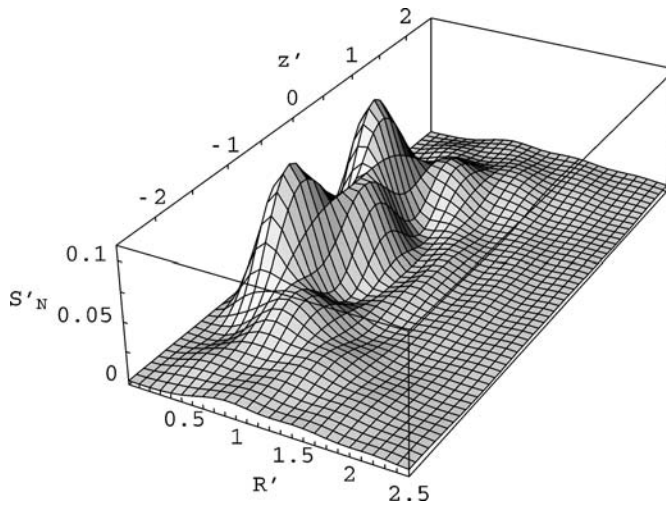


Figure 1 Normal component S'_N of the normalized energy flux vector of an elastic orthonormal beam composed of longitudinal eigenwaves as a function of cylindrical coordinates R' and z' ; $\theta_2 = \pi/2$; $\kappa_0 = 1$; $\Omega = 2\pi$; $j = 3$; $s = 1$.

us set $\theta_2 = \pi$, $\kappa_0 \leq 1/2$, and the amplitude function

$$\mathbf{W}(\theta, \varphi) \equiv \begin{pmatrix} \mathbf{u} \\ \mathbf{f} \end{pmatrix} = \begin{bmatrix} \mathbf{e}_r \\ ik(\lambda_L \mathbf{e}_3 + 2\mu_L \cos \theta' \mathbf{e}_r) \end{bmatrix}, \quad (72)$$

with $\mathbf{e}_r$ given by (42). In this case, the beam manifold is the unit sphere ($\mathcal{B} = S^2$), $\Omega = 2\pi(1 - \cos \kappa_0 \pi) \leq 2\pi$. Upon substitution of $\mathbf{W}(b) = \mathbf{W}(\theta, \varphi)$ into (38), we obtain the orthonormalizing function

$$\nu(\theta) = \frac{1}{\pi} \sqrt{\frac{\kappa_0 N_Q \sin \kappa_0 \theta}{2Q v_1^3 \sin \theta}}, \quad (73)$$

which is independent of the azimuthal angle φ . The smaller κ_0 is, the smaller is Ω , and the beam becomes more collimated but less localized in the transverse direction; see Figure 2. Conversely, if $\kappa_0 = 1/2$, $\Omega = 2\pi$, the beam has a pronounced core region. When $s \neq 0$ or $\kappa_0 \approx 1/2$, such beams resemble elastic tornadoes with spiral energy fluxes.

3.3 Localized fields

Let us now consider time-harmonic fields $\mathbf{W}_j^s$ with $\pi/2 \leq \theta_2 \leq \pi$, $\kappa_0 = 1$ ($\theta' = \theta$), and $2\pi \leq \Omega \leq 4\pi$. For the sake of simplicity, let us consider a special case when the function $\nu = \nu(\theta, \varphi)$ reduces to a constant. Similar electromagnetic and weak gravitational fields have been shown to possess many interesting properties [26], [35]. Taking into consideration peculiarities of energy transport, one can distinguish three types of such localized fields termed *storms* ($\theta_2 = \pi$, $s = 0$), *whirls*

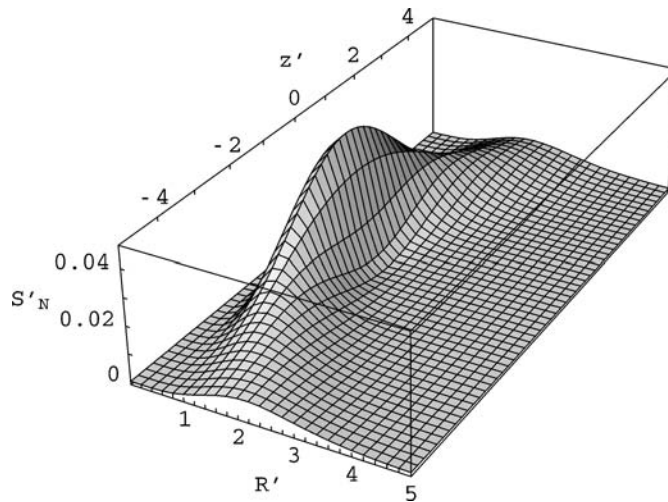


Figure 2 Normal component S'_N of the normalized energy flux vector of an elastic orthonormal beam composed of longitudinal eigenwaves as a function of R' and z' ; $\theta_2 = \pi$; $\kappa_0 = 0.3$; $\Omega = 2\pi[1 - \cos(0.3\pi)]$; $j = s = 2$.

($\theta_2 = \pi$, $s \neq 0$), and *tornadoes* ($\pi/2 \leq \theta_2 \leq \pi$, $s \neq 0$). For storms, the time-averaged energy flux vector $\mathbf{S}$ is identically zero at all points. Whirls and tornadoes have circular and spiral energy flux lines, respectively.

If $\theta_2 = \pi$, the fields under consideration span eigenwaves of all possible propagation directions. They are in effect three-dimensional standing waves with $\mathcal{B} = S^2$ and $\Omega = 4\pi$. Substituting the amplitude function $\mathbf{W}$ of (61) in the expansion (41), we obtain a standing wave [36], [38]

$$\mathbf{u} = i^{|s|+q} \sqrt{2} v_1 e^{i(s\psi - \omega t)} \{ (-1)^p [\mathbf{e}\beta(-s) J_{jp}^{ss-1}[\sin] + \mathbf{e}^* \beta(s) J_{jp}^{ss+1}[\sin]] + \mathbf{e}_3 J_{jq}^{ss}[\cos] \}, \quad (74)$$

where $\beta(s) = -1$ for $s = -1, -2, \dots$, and $\beta(s) = 1$ for $s = 0, 1, 2, \dots$. Here, $p = 1 - q = 0$ if $j + |s|$ is even, and $p = 1 - q = 1$ if $j + |s|$ is odd. The strain tensor $\underline{\underline{\gamma}}$ is given by

$$\underline{\underline{\gamma}} = i^{|s|+p+1} \sqrt{2} k v_1 e^{i(s\psi - \omega t)} \underline{\underline{\gamma}}_0, \quad (75)$$

where

$$\begin{aligned} \underline{\underline{\gamma}}_0 = & \underline{\underline{\rho}} \alpha(s) J_{jp}^{ss-2}[\sin^2] + \underline{\underline{\rho}}^* \alpha(-s) J_{jp}^{ss+2}[\sin^2] + \underline{\underline{\rho}}_1 J_{jp}^{ss}[\sin^2] \\ & + \underline{\underline{\rho}}_2 (-1)^q \beta(-s) J_{jq}^{ss-1}[\sin \circ 2] + \underline{\underline{\rho}}_2^* (-1)^q \beta(s) J_{jq}^{ss+1}[\sin \circ 2] \\ & + \underline{\underline{\rho}}_3 J_{jp}^{ss}[\cos^2] \end{aligned} \quad (76)$$

with $\alpha(1) = 1$ and $\alpha(s) = -1$ for $s \neq 1$. The energy flux density vector $\mathbf{S} = S_0 S'_A \mathbf{e}_A$ has the only nonvanishing (azimuthal) component given by

$$\begin{aligned} S'_A = \frac{2}{\lambda_L + 2\mu_L} \{ & \beta(-s) J_{jp}^{ss-1} [\sin] [\mu_L \alpha(s) J_{jp}^{ss-2} [\sin^2] - J_{jp}^{ss} [\lambda_L + \mu_L \sin^2]] \\ & + \beta(s) J_{jp}^{ss+1} [\sin] [J_{jp}^{ss} [\lambda_L + \mu_L \sin^2] - \mu_L \alpha(-s) J_{jp}^{ss+2} [\sin^2]] \\ & + \mu_L J_{jq}^{ss} [\cos] [\beta(s) J_{jq}^{ss+1} [\sin^2] - \beta(-s) J_{jq}^{ss-1} [\sin^2]] \}. \end{aligned} \quad (77)$$

If $s = 0$, (74) and (76) reduce to

$$\begin{aligned} \mathbf{u} &= i^q \sqrt{2} v_1 e^{-i\omega t} \{ (-1)^p \mathbf{e}_R J_{jp}^{01} [\sin] + \mathbf{e}_3 J_{jq}^{00} [\cos] \}, \\ \gamma_{=0} &= \frac{1}{2} (\mathbf{e}_A \otimes \mathbf{e}_A - \mathbf{e}_R \otimes \mathbf{e}_R) J_{jp}^{02} [\sin^2] \\ &+ \frac{1}{2} (\mathbf{e}_R \otimes \mathbf{e}_3 + \mathbf{e}_3 \otimes \mathbf{e}_R) (-1)^q J_{jq}^{01} [\sin^2] \\ &+ \rho_1 J_{jp}^{00} [\sin^2] + \rho_3 J_{jp}^{00} [\cos^2]. \end{aligned} \quad (79)$$

In this case, the displacement vectors lie in the meridional planes, and the time-averaged energy flux is everywhere zero.

Figures 3 and 4 illustrate properties of the three-dimensional standing waves. The nonvanishing cylindrical components of the displacement vector $\mathbf{u}$ of (78); i.e., the radial component u_R and the normal component $u_N = u_3$, are independent of the azimuthal angle ψ . Figure 3 illustrates the instantaneous field u_N for a storm

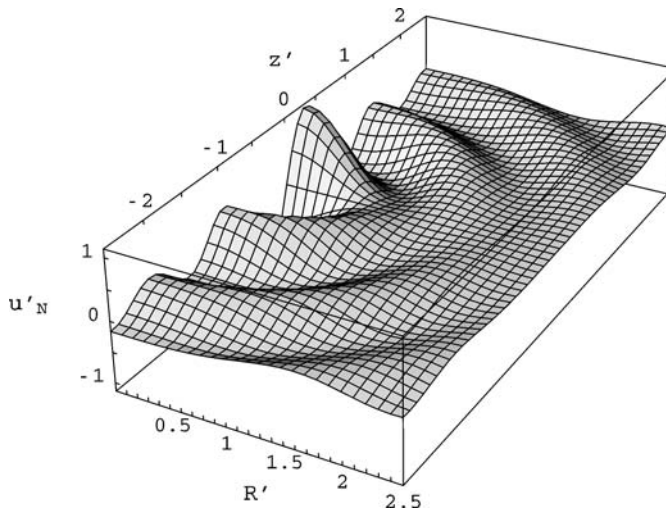


Figure 3 Normal component u'_N of the normalized instantaneous displacement field $\mathbf{u}' = (\text{Re } \mathbf{u})/u_n$ [$u_n = (2/\omega)\sqrt{w_0/\rho}$] of an elastic storm composed of longitudinal eigenwaves as a function of R' and z' ; $\theta_2 = \pi$; $\kappa_0 = 1$; $\Omega = 4\pi$; $j = s = 0$; $\omega t = \pi/2$.

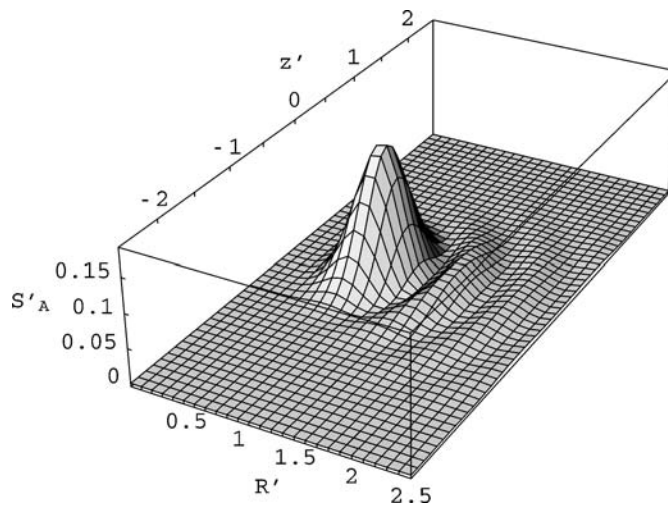


Figure 4 Azimuthal component S'_A of the normalized energy flux vector of an elastic whirl composed of longitudinal eigenwaves as a function of R' and z' ; $\theta_2 = \pi$; $\kappa_0 = 1$; $\Omega = 4\pi$; $j = s = 4$.

with $j = 0$. Figure 4 depicts the azimuthal energy flux of the elastic whirl with parameters $j = s = 4$. For the whirls defined by the sectorial spherical harmonics Y_j^j ($j \neq 0$), the azimuthal energy flux peaks in the plane $z = 0$. Calculations show that other whirls ($s \neq 0$, $s \neq j$) have two domains with large azimuthal energy fluxes that are symmetrical with respect to the plane $z = 0$, whereas in this plane $\mathbf{S} \equiv \mathbf{0}$. Since S'_A is independent of the azimuthal angle ψ , all whirls have circular energy flux lines.

Let us now consider a family of fields $\mathbf{W}_j^s$ with $\pi/2 < \theta_2 < \pi$ ($2\pi < \Omega < 4\pi$). Similar to storms and whirls, these fields are highly localized. However, the normal S'_N and the radial S'_R components of $\mathbf{S}$ do not vanish. As a result, lines of energy flux become spiral-like, provided that $s \neq 0$. That is why we refer to these unique localized fields as elastic tornadoes. The smaller θ_2 is, the larger is the normal component S'_N (see Figure 5) that defines the pitch of these spirals. The nonvanishing radial component S'_R results in variations of their radiuses. As θ_2 tends to π , the pitch of the spirals decreases, and a tornado reduces to a whirl with circular energy flux lines.

For the fields with $s = 0$, $\pi/2 < \theta_2 < \pi$, and ($2\pi < \Omega < 4\pi$), the lines of energy flux lie in meridional planes. These fields are intermediate in properties between the elastic storms and the beams with $s = 0$ and $\Omega = 2\pi$ (see Section 3.1).

4 Superpositions of transverse eigenwaves

In this section, we consider elastic fields composed of transverse eigenwaves propagating in an isotropic medium. These eigenwaves have two-dimensional ampli-

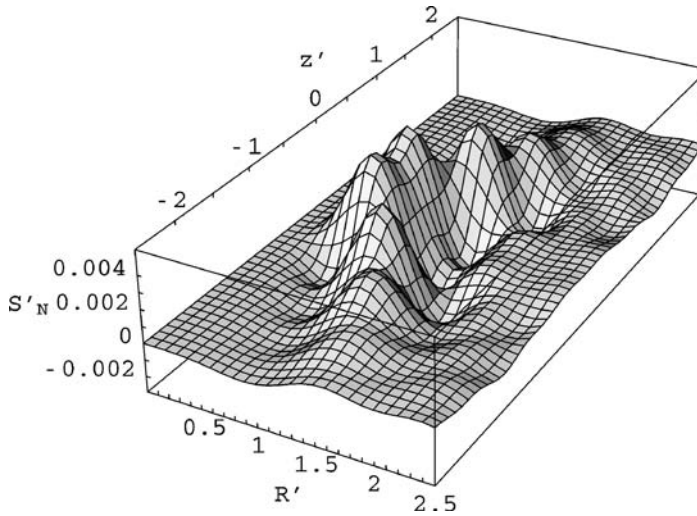


Figure 5 Normal component S'_N of the normalized energy flux vector of an elastic tornado composed of longitudinal eigenwaves as a function of R' and z' ; $\theta_2 = 5\pi/6$; $\kappa_0 = 1$; $\Omega = 2\pi(1 + \sqrt{3}/2)$; $j = 3$; $s = 2$.

tude subspaces. Therefore, we set two amplitude functions as

$$\mathbf{W}_M(\theta, \varphi) \equiv \begin{pmatrix} \mathbf{u} \\ \mathbf{f} \end{pmatrix} = \begin{pmatrix} \mathbf{e}_{\theta'} \\ ik\mu_L(-\sin\theta'\mathbf{e}_r + \cos\theta'\mathbf{e}_{\theta'}) \end{pmatrix}, \quad (80)$$

$$\mathbf{W}_A(\theta, \varphi) \equiv \begin{pmatrix} \mathbf{u} \\ \mathbf{f} \end{pmatrix} = \begin{pmatrix} \mathbf{e}_{\varphi} \\ ik\mu_L \cos\theta'\mathbf{e}_{\varphi} \end{pmatrix}, \quad (81)$$

where

$$\mathbf{e}_{\theta'} = \cos\theta'(\mathbf{e}_1 \cos\varphi + \mathbf{e}_2 \sin\varphi) - \mathbf{e}_3 \sin\theta', \quad (82)$$

$$\mathbf{e}_{\varphi} = -\mathbf{e}_1 \sin\varphi + \mathbf{e}_2 \cos\varphi. \quad (83)$$

Therein $\mathbf{e}_r$ is the spherical basis vector, see (42), and $k = 2\pi/\lambda = \omega/v_2$ is the wavenumber. This gives two different families of beams ($\mathbf{u}_M$ beams and $\mathbf{u}_A$ beams), which are composed of eigenwaves with the meridional and azimuthal orientation of $\mathbf{u}$, respectively.

4.1 Orthonormal beams of Type I

Let us first consider the family of orthonormal beams $\mathbf{W}_j^s$ with $\theta_2 = \pi/2$ and $\kappa_0 = 1$ ($\theta' = \theta$), which comprises eigenwaves propagating into a solid angle $\Omega = 2\pi$. The orthonormalizing function $\nu = \nu(\theta, \varphi)$ reduces to a constant for either of the amplitude functions. As a result, the displacement $\mathbf{u}$, the force density $\mathbf{f} = \underline{\underline{\sigma}}\mathbf{e}_3$, the

strain tensor $\underline{\underline{\gamma}}$ and the stress tensor $\underline{\underline{\sigma}}$ of the $\mathbf{u}_M$ beam have the form [36], [39]

$$\mathbf{u}_M = v_2 e^{i(s\psi - \omega t)} \{ \mathbf{e} I_j^{s s-1} [\cos] + \mathbf{e}^* I_j^{s s+1} [\cos] - \mathbf{e}_3 I_j^{s s} [\sin] \}, \quad (84)$$

$$\mathbf{f}_M = i k \mu_L v_2 e^{i(s\psi - \omega t)} \{ \mathbf{e} I_j^{s s-1} [\cos \circ 2] + \mathbf{e}^* I_j^{s s+1} [\cos \circ 2] - \mathbf{e}_3 I_j^{s s} [\sin \circ 2] \}, \quad (85)$$

$$\underline{\underline{\gamma}}_M = \frac{\underline{\underline{\sigma}}_M}{2\mu_L} = i k v_2 e^{i(s\psi - \omega t)} \underline{\underline{\gamma}}_0, \quad v_2 = \frac{1}{\pi} \sqrt{\frac{N_Q}{\varrho v_2^3}}, \quad (86)$$

where

$$\begin{aligned} \underline{\underline{\gamma}}_0 = & \frac{1}{2} \underline{\underline{\rho}} I_j^{s s-2} [\sin \circ 2] + \frac{1}{2} \underline{\underline{\rho}}^* I_j^{s s+2} [\sin \circ 2] + \underline{\underline{\rho}} I_j^{s s-1} [\cos \circ 2] \\ & + \underline{\underline{\rho}}^* I_j^{s s+1} [\cos \circ 2] + \frac{1}{2} (\underline{\underline{\rho}}_1 - \underline{\underline{\rho}}_3) I_j^{s s} [\sin \circ 2]. \end{aligned} \quad (87)$$

Similarly, for the $\mathbf{u}_A$ beam, we obtain

$$\mathbf{u}_A = i v_2 e^{i(s\psi - \omega t)} (\mathbf{e}^* I_j^{s s+1} [1] - \mathbf{e} I_j^{s s-1} [1]), \quad (88)$$

$$\mathbf{f}_A = k \mu_L v_2 e^{i(s\psi - \omega t)} (\mathbf{e} I_j^{s s-1} [\cos] - \mathbf{e}^* I_j^{s s+1} [\cos]), \quad (89)$$

$$\begin{aligned} \underline{\underline{\gamma}}_A = \frac{\underline{\underline{\sigma}}_A}{2\mu_L} = & k v_2 e^{i(s\psi - \omega t)} \{ \underline{\underline{\rho}} I_j^{s s-2} [\sin] - \underline{\underline{\rho}}^* I_j^{s s+2} [\sin] + \underline{\underline{\rho}} I_j^{s s-1} [\cos] \\ & - \underline{\underline{\rho}}^* I_j^{s s+1} [\cos] \}. \end{aligned} \quad (90)$$

Using the foregoing expressions, one can obtain the kinetic and elastic energy densities, and the energy flux density vector $\mathbf{S}$. Figures 6 and 7 illustrate the distinctions between the normal energy fluxes S'_N of $\mathbf{u}_M$ and $\mathbf{u}_A$ beams defined by the same spherical harmonic Y_0^0 .

The orthonormal elastic beams bear some similarities to the electromagnetic beams treated elsewhere, [26], [35]. In particular, the displacement fields for both $\mathbf{u}_M$ and $\mathbf{u}_A$ beams are described by the same functions as for the corresponding electric or magnetic field. As a result, Figures 1, 2, and 4 in [26] illustrate the properties of both electromagnetic and elastic beams.

4.2 Orthonormal beams of Type II

As in the case of the beams composed of longitudinal eigenwaves, to obtain a complete system of orthonormal beams, defined by the whole set of spherical harmonics, we set $\theta_2 = \pi$ and $\kappa_0 \leq 1/2$. Upon substitution of $\mathbf{W}(b) = \mathbf{W}_M(\theta, \varphi)$ and

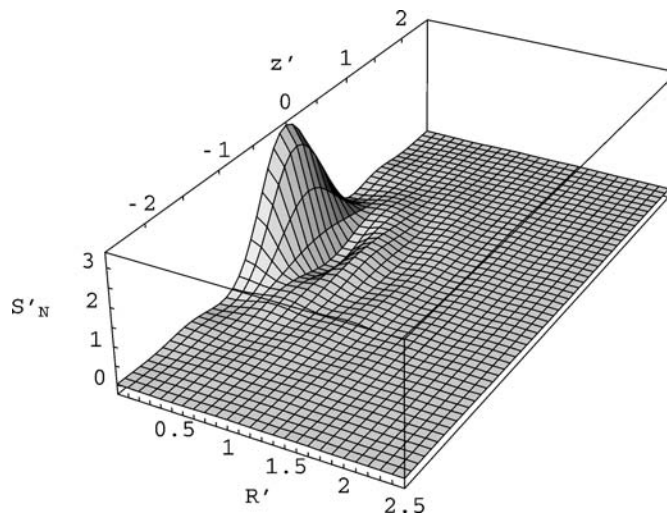


Figure 6 Normal component S'_N of the normalized energy flux vector of an elastic orthonormal $\mathbf{u}_M$ beam composed of transverse eigenwaves as a function of R' and z' ; $\theta_2 = \pi/2$; $\kappa_0 = 1$; $\Omega = 2\pi$; $j = s = 0$.

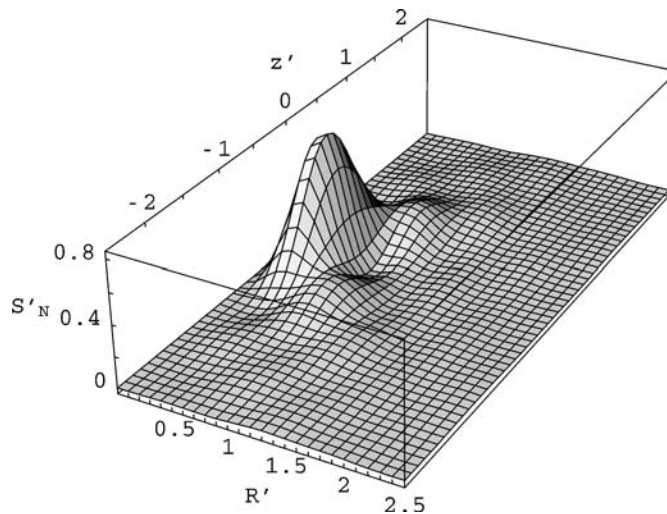


Figure 7 Normal component S'_N of the normalized energy flux vector of an elastic orthonormal $\mathbf{u}_A$ beam; the parameters are the same as those in Figure 6.

$\mathbf{W}(b) = \mathbf{W}_A(\theta, \varphi)$ into (38), we obtain the same orthonormalizing function for both $\mathbf{u}_M$ and $\mathbf{u}_A$ beams:

$$v(\theta) = \frac{1}{\pi} \sqrt{\frac{\kappa_0 N_Q \sin \kappa_0 \theta}{2Q v_2^3 \sin \theta}}. \quad (91)$$

Thus, the orthonormalizing function for the shear beams is independent of the azimuthal angle φ ; and the relations (73) and (91) for the beams comprising longitudinal and transverse eigenwaves show that the two types of beams differ in phase velocities only. The parameter κ_0 exerts primary control over the beam divergence and the dimensions of the beam cross-section at the plane $z = 0$. The smaller κ_0 is, the more collimated is the beam, but the cross-section becomes larger. Figure 8 and 9 show the energy distributions for $\mathbf{u}_M$ and $\mathbf{u}_A$ beams defined by the same spherical harmonic Y_1^0 . Further illustrations for such orthonormal beams can be found elsewhere [39].

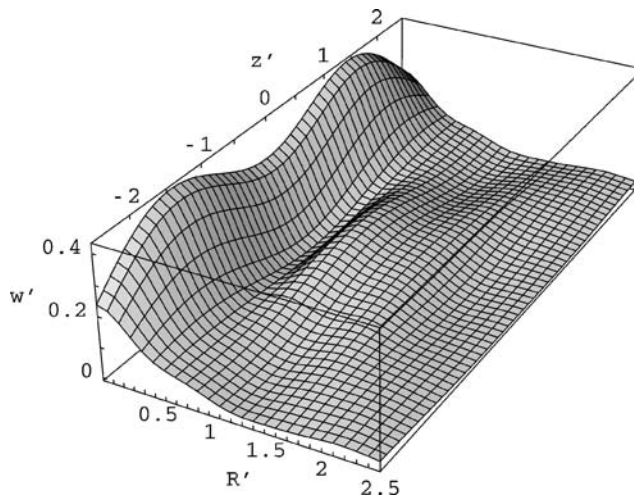


Figure 8 Normalized energy density $w' = (w_K + w_E)/w_0$ [$w_0 = S_0/v_2$] of an elastic orthonormal $\mathbf{u}_M$ beam composed of transverse eigenwaves as a function of R' and z' ; $\theta_2 = \pi$; $\kappa_0 = 0.3$; $\Omega = 2\pi[1 - \cos(0.3\pi)]$; $j = 1$; $s = 0$.

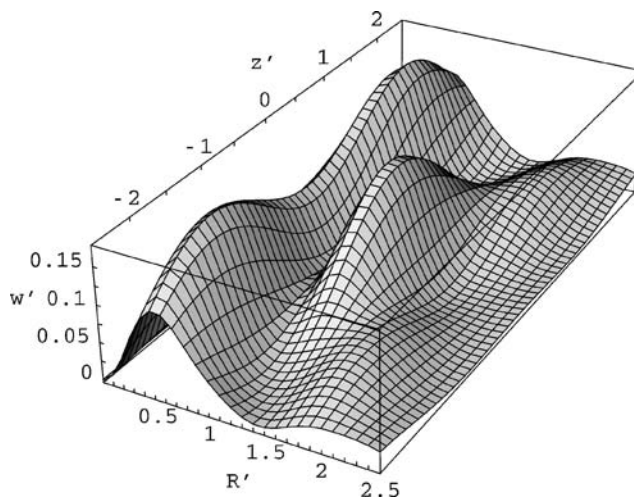


Figure 9 Normalized energy density $w' = (w_K + w_E)/w_0$ of an elastic orthonormal $\mathbf{u}_A$ beam; the parameters are the same as those in Figure 8.

4.3 Localized fields

In Section 3.3 we presented three families of localized elastic fields composed of longitudinal eigenwaves propagating in an isotropic medium: elastic storms, whirls, and tornadoes. In this section, we present relations for similar fields comprising transverse eigenwaves. They are described by the expansion (41) with $\pi/2 \leq \theta_2 \leq \pi$ and $\kappa_0 = 1$ ($\theta' = \theta$). As before, we assume that the beam state function $v = v(\theta, \varphi)$ reduces to a constant.

When $\theta_2 = \pi$, (41) describes three-dimensional standing waves [36], [39], namely, the $\mathbf{u}_M$ beam:

$$\mathbf{u}_M = \sqrt{2}v_2 e^{i(s\psi - \omega t)} \left\{ \mathbf{e} i^{|s-1|+q} J_{jq}^{ss-1}[\cos] + \mathbf{e}^* i^{|s+1|+q} J_{jq}^{ss+1}[\cos] - \mathbf{e}_3 i^{|s|+p} J_{jp}^{ss}[\sin] \right\}, \quad (92)$$

$$\begin{aligned} \underline{\underline{\sigma}}_M = 2\mu_L \underline{\underline{\gamma}}_M = i^{|s|+q+1} \sqrt{2}k v_2 \mu_L e^{i(s\psi - \omega t)} \left\{ \underline{\underline{\rho}} \alpha(s) J_{jq}^{ss-2}[\sin \circ 2] \right. \\ \left. + \underline{\underline{\rho}}^* \alpha(-s) J_{jq}^{ss+2}[\sin \circ 2] + 2\underline{\underline{\rho}}_2 (-1)^p \beta(-s) J_{jp}^{ss-1}[\cos \circ 2] \right. \\ \left. + 2\underline{\underline{\rho}}_2^* (-1)^p \beta(s) J_{jp}^{ss+1}[\cos \circ 2] + (\underline{\underline{\rho}}_1 - \underline{\underline{\rho}}_3) J_{jq}^{ss}[\sin \circ 2] \right\}, \quad (93) \end{aligned}$$

and the $\mathbf{u}_A$ beam:

$$\mathbf{u}_A = i\sqrt{2}v_2 e^{i(s\psi - \omega t)} \left\{ \mathbf{e}^* i^{|s+1|+p} J_{jp}^{ss+1}[1] - \mathbf{e} i^{|s-1|+p} J_{jp}^{ss-1}[1] \right\}, \quad (94)$$

$$\begin{aligned} \underline{\underline{\sigma}}_A = 2\mu_L \underline{\underline{\gamma}}_A = 2\sqrt{2}k v_2 \mu_L e^{i(s\psi - \omega t)} \left\{ \underline{\underline{\rho}} i^{|s-2|+p} J_{jp}^{ss-2}[\sin] \right. \\ \left. - \underline{\underline{\rho}}^* i^{|s+2|+p} J_{jp}^{ss+2}[\sin] + \underline{\underline{\rho}}_2 i^{|s-1|+q} J_{jq}^{ss-1}[\cos] \right. \\ \left. - \underline{\underline{\rho}}_2^* i^{|s+1|+q} J_{jq}^{ss+1}[\cos] \right\}. \quad (95) \end{aligned}$$

For $\mathbf{u}_M$ and $\mathbf{u}_A$ storms ($s = 0$), the time-averaged energy flux vector $\mathbf{S}$ is identically zero at all points. The whirls of both types have only azimuthal time-average energy fluxes ($\mathbf{S} = S_0 S'_A \mathbf{e}_A$). For $\mathbf{u}_M$ and $\mathbf{u}_A$ whirls, these fluxes are described by

$$\begin{aligned} S'_A = \beta(s) J_{jq}^{ss+1}[\cos] \left\{ \alpha(s) J_{jq}^{ss-2}[\sin \circ 2] - J_{jq}^{ss}[\sin \circ 2] \right\} \\ + \beta(-s) J_{jq}^{ss-1}[\cos] \left\{ J_{jq}^{ss}[\sin \circ 2] - \alpha(-s) J_{jq}^{ss+2}[\sin \circ 2] \right\} \\ + 2J_{jp}^{ss}[\sin] \left\{ \beta(-s) J_{jp}^{ss-1}[\cos \circ 2] - \beta(s) J_{jp}^{ss+1}[\cos \circ 2] \right\}, \quad (96) \end{aligned}$$

and

$$S'_A/2 = \beta(1+s) J_{jp}^{ss+1}[1] J_{jp}^{ss+2}[\sin] - \beta(1-s) J_{jp}^{ss-1}[1] J_{jp}^{ss-2}[\sin], \quad (97)$$

respectively.

When $s \neq 0$ and $\pi/2 < \theta_2 < \pi$, the field $\mathbf{W}_j^s$ of (41) is composed of eigenwaves propagating in the solid angle Ω ($2\pi < \Omega < 4\pi$). It is highly localized and has spiral energy flux lines similar to those of elastic tornadoes composed of longitudinal eigenwaves. The amplitude functions $\mathbf{W}$ of (80) and (81) specify two different polarization types ($\mathbf{u}_M$ and $\mathbf{u}_A$) of such elastic tornadoes.

For the fields defined by the zonal spherical harmonics ($s = 0$), energy flux lines lie in meridional planes. The perturbation $\underline{\underline{A}}$ of $\underline{\underline{\epsilon}}^{-1}$, caused by shear waves, is proportional to the deformation tensor $\underline{\underline{\gamma}}$. Figures 10 and 11 illustrate nonvanishing components of $\underline{\underline{A}}$ induced by the storm with $j = 1$.

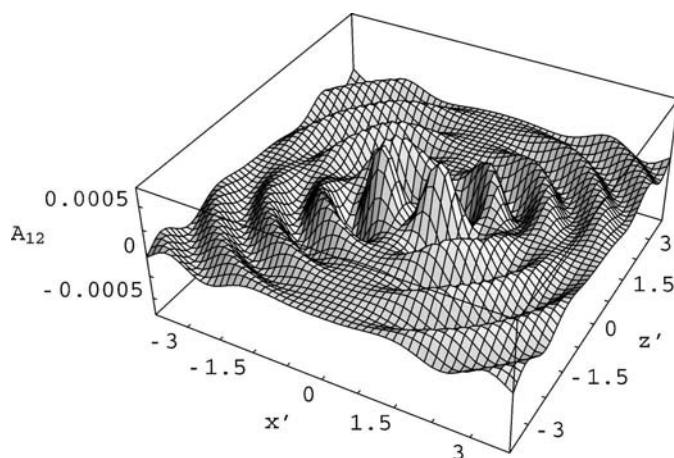


Figure 10 Component $A_{12} = A_{21} = (\underline{\underline{\epsilon}}^{-1})_{12}$ induced by an elastic $\mathbf{u}_A$ storm composed of transverse eigenwaves as a function of x' and z' ; $\theta_2 = \pi$; $\kappa_0 = 1$; $\Omega = 4\pi$; $\nu = \nu_2/\sqrt{2}$; $N_Q = 0.5W$; $j = 1$; $s = 0$; $y' = 0$; $t = 0$.

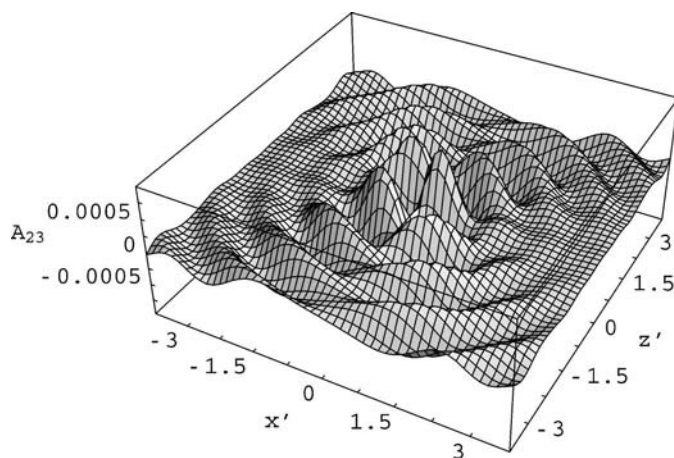


Figure 11 Component $A_{23} = A_{32} = (\underline{\underline{\epsilon}}^{-1})_{23}$; the parameters are the same as those in Figure 10.

5 Complex field structures

The presented elastic fields have a very small core region (about several wavelengths) with maximum intensity of field oscillations and unique spatial distributions of polarization states, energy densities, and energy fluxes. Outside the core, the intensity of oscillations rapidly decreases in all directions. By applying the relations from Section 2.3, one can obtain similar acoustic fields in liquids.

It was shown earlier that the beams defined by the zonal spherical harmonics Y_j^0 ($j = 0, 1, \dots$) have axially symmetric instantaneous fields. The beams, defined by Y_j^s with $s \neq 0$, constitute a family of rotating fields. It is then inevitable that they will induce chiral inhomogeneities in an isotropic elastic medium. Figures 12–15 illustrate chiral inhomogeneities induced by a orthonormal elastic beam

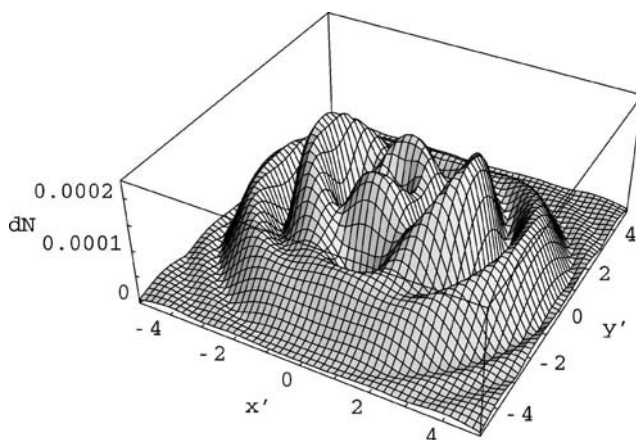


Figure 12 Refractive-index variation $dN = n_+ - n_0$ in the plane $z' = 0$, induced by an elastic $\mathbf{u}_A$ beam composed of transverse eigenwaves; $\theta_2 = \pi$; $\kappa_0 = 0.15$; $\Omega = 2\pi[1 - \cos(0.15\pi)]$; $N_Q = 5W$; $j = 2$; $s = 1$; $\mathbf{n} = \mathbf{e}_3$; $t = 0$.

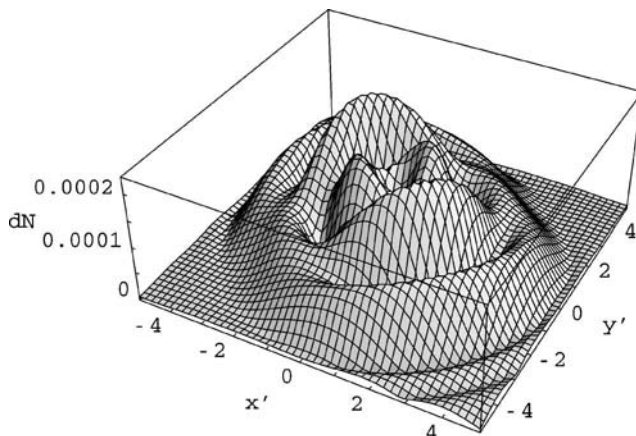


Figure 13 Refractive-index variation $dN = n_+ - n_0$ in the plane $z' = 0.125$; the other parameters are the same as those in Figure 12.

defined by the spherical harmonic Y_2^1 . They show induced refractive index variations $dN = n_+ - n_0$ in four cross-sections of the $\mathbf{u}_A$ beam with parameters $\theta_2 = \pi$ and $\kappa_0 = 0.15$. For sufficiently small values of z' ($|z'| < 2$), the spatial distribution $dN(x', y', z')$ in a plane $z' = \text{const}$ can be approximately found by the rotation of the distribution $dN(x', y', 0)$ in the plane $z' = 0$ through an angle $\psi_{rot} \approx 2\pi z'$ around the z' axis. Figure 15 illustrates the distribution $dN(x', y', 3)$ with already noticeable deviations from this approximation. Similar illustrations for the beam defined by Y_1^1 can be found elsewhere [42].

The three-dimensional localization makes it possible to use the presented fields as structural elements to form various complex elastic and sound fields. By way of

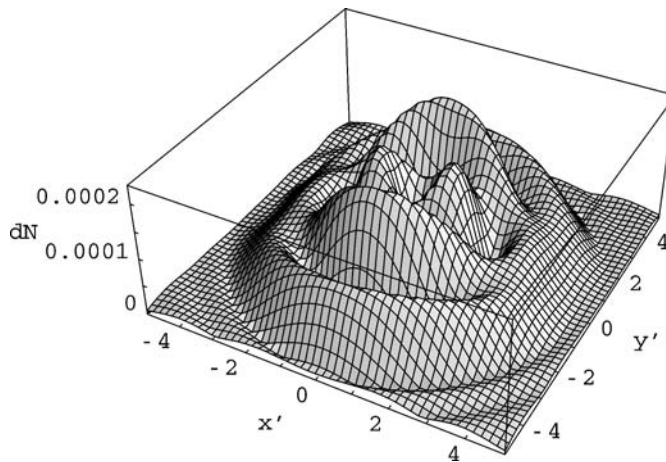


Figure 14 Refractive-index variation $dN = n_+ - n_0$ in the plane $z' = 0.25$; the other parameters are the same as those in Figure 12.

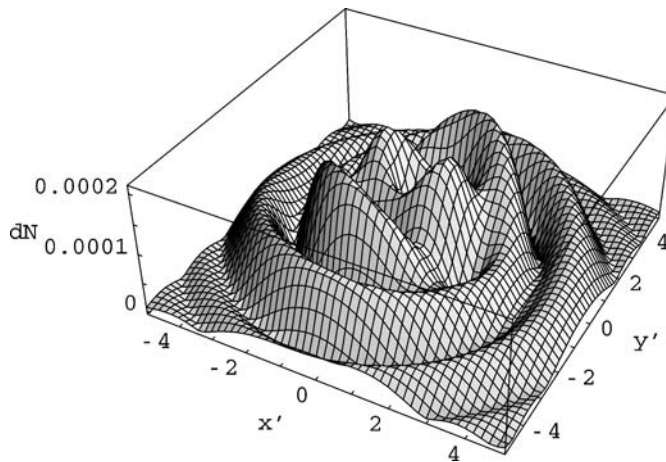


Figure 15 Refractive-index variation $dN = n_+ - n_0$ in the plane $z' = 3$; the other parameters are the same as those in Figure 12.

illustration, let us consider a field defined as

$$\mathbf{W}'_n(\mathbf{r}, t) = \sum_{n=M_1}^{N_1} \sum_{m=M_2}^{N_2} \sum_{l=M_3}^{N_3} \mathbf{W}_n(\mathbf{r} - \mathbf{a}_{nml}, t - \tau_{nml}), \quad (98)$$

$$\mathbf{a}_{nml} = n\mathbf{a}_1 + m\mathbf{a}_2 + l\mathbf{a}_3, \quad (99)$$

$$\tau_{nml} = n\tau_1 + m\tau_2 + l\tau_3, \quad (100)$$

where $\mathbf{W}_n(\mathbf{r}, t)$ is given by (34), $\mathbf{a}_j$ and τ_j are some prescribed spatial and temporal shifts. On the one hand, this field can be treated as a different state of the beam $\mathbf{W}_n(\mathbf{r}, t)$, obtained by the beam state transformation $v(b) \mapsto v'(b)$, where

$$v'(b) = v(b)G_1(b)G_2(b)G_3(b), \quad (101)$$

$$G_j(b) = \sum_{m_j=M_j}^{N_j} \exp\{i[\tau_j\omega(b) - \mathbf{a}_j \cdot \mathbf{k}(b)]\}, \quad j = 1, 2, 3. \quad (102)$$

On the other hand, the translated beams $\mathbf{W}_n(\mathbf{r} - \mathbf{a}, t - \tau)$ form a family of wavelets with $\mathbf{W}_n(\mathbf{r}, t)$ as the mother wavelet [48]. Hence, if $\mathbf{W}_n$ is one of the foregoing localized elastic fields, and the lengths of the shift vectors $\mathbf{a}_j$ ($j = 1, 2, 3$) are sufficiently large, the field $\mathbf{W}'_n$ will constitute a field grating. Figure 16 shows the spatial distribution of refractive index variation $dN = n_+ - n_0$ induced by a two-dimensional grating composed of 49 $\mathbf{u}_M$ beams of Type I. The presented results make it possible to calculate first the strain tensor $\underline{\gamma}$ and the stress tensor $\underline{\sigma}$ and thereafter the permittivity tensor and other optical properties of such an ultrasonic

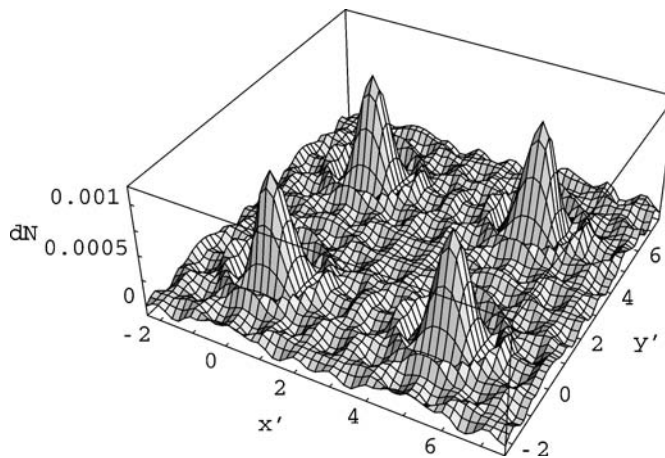


Figure 16 Refractive index variation $dN = n_+ - n_0$ in the plane $z' = 0$, induced by a two-dimensional grating composed of elastic $\mathbf{u}_M$ beams; $\theta_2 = \pi/2$; $\kappa_0 = 1$; $\Omega = 2\pi$; $N_Q = 0.5W$; $j = s = 0$; $M_1 = M_2 = -3$; $N_1 = N_2 = 3$; $M_3 = N_3 = 0$; $\mathbf{a}_1 = 5\lambda\mathbf{e}_1$; $\mathbf{a}_2 = 5\lambda\mathbf{e}_2$; $\mathbf{a}_3 = \mathbf{0}$; $\tau_1 = \tau_2 = \tau_3 = 0$; $\mathbf{n} = \mathbf{e}_3$; $\omega t = \pi/2$.

diffraction grating [42]. The latter can be a useful supplement to the plane wave ultrasonic diffraction gratings widely used to control laser radiation [47].

The great diversity of presented localized fields with different geometries of core regions provides a great scope for combining them, as constructive elements, into various one-, two-, or three-dimensional gratings and other complex geometrical structures, where each element has only reasonably small deviations from its initial form. This makes it possible to sculpture the morphology of complex ultrasonic fields and, hence, induced optical inhomogeneities. To some extent, this resembles the concept of the sculptured thin films [49-50], but applied instead to induced time-harmonic structures (sculptured fields).

6 Conclusion

New exact solutions of wave equations in linear isotropic and anisotropic elastic mediums, describing various families of orthonormal beams and localized fields, have become available by the method described in this chapter. Detailed illustrations in the form of unique elastic fields in an isotropic medium, defined by the spherical harmonics, were provided.

There are two different ways to compose families of orthonormal beams. The first results in two separate sets of orthonormal beams defined by the spherical harmonics Y_j^s with even and odd j , respectively. The second gives a complete system of orthonormal beams defined by the whole set of spherical harmonics Y_j^s . Owing to the orthonormality conditions, the families of orthonormal beams can be used as functional bases for complex elastic fields.

The presented localized fields include elastic storms, whirls, and tornadoes; i.e., highly localized fields for which time-averaged energy flux is identically zero at all points, azimuthal and spiral, respectively. Some of these fields can induce chiral inhomogeneities in an isotropic elastic medium. Owing to the three-dimensional localization, these fields can be combined into complex field structures, such as ultrasonic diffraction gratings which makes them promising tools to control laser radiation.

The presented elastic beams are composed of homogeneous eigenwaves with specifically set intensities, phases, polarizations, and radial propagation directions. As a possible source of such beams, one can imagine a large set of piezoelectric elements with digital control over intensities and phases of radiated ultrasonic beams. The partial beam from each element must propagate in the radial direction and, in the central domain, closely resemble an eigenwave. The larger (in terms of wavelengths) this domain is, the larger the grating that can be built. To obtain a reasonable approximation to the beam $\mathbf{W}_j^s$ and various complex acoustic structures, the source must consist of a sufficiently large number of elements (several thousand). Such structures with digital switching from one mode to another can be a useful multipurpose tool to control laser radiation, which makes it a kind of acoustic processor.

References

- [1] J. Durnin, Exact solutions for nondiffracting beams. 1. The scalar theory, *J Opt Soc Am A* **4** (1987), 651–654.
- [2] J. Durnin, J.J. Miceli, Jr. and J.H. Eberly, Comparison of Bessel and Gaussian beams, *Opt Lett* **13** (1988), 79–80.
- [3] A.M. Belsky, Components of field vectors and energy fluxes of nondiffracting electromagnetic beams, *Vestnik Belarusian State University*, Ser. 1, No. 2, (1995), 8–10.
- [4] A.M. Belsky, *Optics of coherent light beams*, Belarus State University, Minsk, Belarus, 2000.
- [5] F. Gori, G. Guattari and C. Padovani, Bessel–Gauss beams, *Opt Commun* **64** (1987), 491–495.
- [6] F. Bloisi and L. Vicari, Comparison of nondiffracting laser beams, *Opt Commun* **75** (1990), 353–357.
- [7] J.N. Brittingham, Focus waves modes in homogeneous Maxwell’s equations: Transverse electric mode, *J Appl Phys* **54** (1983), 1179–1189.
- [8] T.T. Wu and R.W.P. King, Comment on “Focus waves modes in homogeneous Maxwell’s equations: Transverse electric mode,” *J Appl Phys* **56** (1984), 2587–2588.
- [9] A. Sezginer, A general formulation of focus wave modes, *J Appl Phys* **57** (1985), 678–683.
- [10] R.W. Ziolkowski, Exact solutions of the wave equation with complex source locations, *J Math Phys* **26** (1985), 861–863.
- [11] T.T. Wu, Electromagnetic missiles, *J Appl Phys* **57** (1985), 2370–2373.
- [12] R.W. Ziolkowski, D.K. Lewis and B.D. Cook, Evidence of localized wave transmission, *Phys Rev Lett* **62** (1989), 147–150.
- [13] R.W. Ziolkowski and D.K. Lewis, Verification of the localized wave transmission effect, *J Appl Phys* **68** (1990), 6083–6086.
- [14] R.W. Ziolkowski, Localized transmission of electromagnetic energy, *Phys Rev A* **39** (1989), 2005–2033.
- [15] A.M. Shaarawi, I.M. Besieris and R.W. Ziolkowski, Localized energy pulse trains launched from an open, semi-infinite, circular waveguide, *J Appl Phys* **65** (1989), 805–813.
- [16] P.L. Overfelt, Bessel–Gauss pulses, *Phys Rev A* **44** (1991), 3941–3947.
- [17] A.P. Kiselev and M.V. Perel, Highly localized solutions of the wave equation, *J Math Phys* **41** (2000), 1934–1955.
- [18] R.W. Ziolkowski, I.M. Besieris and A.M. Shaarawi, Aperture realizations of exact solutions to homogeneous wave equations, *J Opt Soc Am A* **10** (1993), 75–87.
- [19] R.W. Ziolkowski, Localized wave physics and engineering, *Phys Rev A* **44** (1991), 3960–3984.
- [20] P. Saari and K. Reivelt, Evidence of X-shaped propagation invariant localized light waves, *Phys Rev Lett* **79** (1997), 4135–4138.

- [21] K. Reivelt and P. Saari, Optical generation of focus wave modes, *J Opt Soc Am A* **17** (2000), 1785–1790.
- [22] A.M. Shaarawi, R.W. Ziolkowski and I.M. Besieris, On the evanescent fields and the causality of the focus wave modes, *J Math Phys* **36** (1995), 5565–5587.
- [23] R. Donnelly and R.W. Ziolkowski, A method for constructing solutions for homogeneous partial differential equations: localized waves, *Proc Roy Soc Lond A* **437** (1992), 673–692.
- [24] R. Donnelly and R.W. Ziolkowski, Designing localized waves, *Proc Roy Soc Lond A* **440** (1993), 541–565.
- [25] G.N. Borzdov, New types of electromagnetic beams in complex media and vacuum, *Proceedings of 7th international symposium on recent advances in microwave technology* (C.C. Peñalosa and B.S. Rawat, eds), Málaga, Spain (December 13–17, 1999), 169–172.
- [26] G.N. Borzdov, Plane-wave superpositions defined by orthonormal scalar functions on two- and three-dimensional manifolds, *Phys Rev E* **61** (2000), 4462–4478.
- [27] G.N. Borzdov, New types of orthonormal electromagnetic beams in complex media and free space, *Proceedings of Bianisotropics 2000: 8th international conference on electromagnetics of complex media* (A.M. Barbosa and A.L. Topa, eds), Lisbon, Portugal, 2000, 55–58.
- [28] G.N. Borzdov, Localized electromagnetic fields in complex media and free space, *Proceedings of Bianisotropics 2000: 8th international conference on electromagnetics of complex media* (A.M. Barbosa and A.L. Topa, eds), Lisbon, Portugal, 2000, 59–62.
- [29] G.N. Borzdov, The application of orthonormal beams to characterizing complex media, *Proceedings of Bianisotropics 2000: 8th international conference on electromagnetics of complex media* (A.M. Barbosa and A.L. Topa, eds), Lisbon, Portugal, 2000, 11–14.
- [30] G.N. Borzdov, Orthonormal beams with applications to characterizing electromagnetic fields and complex mediums, *Arch Elektron Übertrag* **55** (2001), 224–232.
- [31] G.N. Borzdov, Lorentz-covariant surface impedance and characteristic matrix methods with applications to measurements of material parameters of linear media, *Opt Commun* **94** (1992), 159–173.
- [32] G.N. Borzdov, An intrinsic tensor technique in Minkowski space with applications to boundary value problems, *J Math Phys* **34** (1993), 3162–3196.
- [33] G.N. Borzdov, Inverse problem of reflection and transmission for a bianisotropic medium, *Advances in complex electromagnetic materials* (A. Priou, A. Sihvola, S. Tretyakov and A. Vinogradov, eds), Kluwer, Dordrecht, The Netherlands, 1997, 71–84.
- [34] G.N. Borzdov, Novel free-space techniques to characterize complex mediums, *Electromagnetic fields in unconventional materials and structures* (O.N. Singh and A. Lakhtakia, eds), Wiley, New York, NY, USA, 2000, 83–124.

- [35] G.N. Borzdov, Localized electromagnetic and weak gravitational fields in the source-free space, *Phys Rev E* **63** (2001), 036606.
- [36] G.N. Borzdov, Elastic orthonormal beams and localized fields with applications to control laser radiation, *Complex mediums II: Beyond linear isotropic dielectrics* (A. Lakhtakia, W.S. Weiglhofer and I.J. Hodgkinson, eds), Proc SPIE **4467**, Bellingham, WA, USA, 2001, 57–68.
- [37] G.N. Borzdov, Elastic and sound orthonormal beams and localized fields in linear mediums: I. Basic equations, *J Phys A: Math Gen* **34** (2001), 6249–6257.
- [38] G.N. Borzdov, Elastic and sound orthonormal beams and localized fields in linear mediums: II. Superpositions of longitudinal elastic plane waves in an isotropic medium, *J Phys A: Math Gen* **34** (2001), 6259–6267.
- [39] G.N. Borzdov, Elastic and sound orthonormal beams and localized fields in linear media: III. Superpositions of transverse elastic plane waves in an isotropic medium, *J Phys A: Math Gen* **34** (2001), 6269–6279.
- [40] G.N. Borzdov, Elastic and sound orthonormal beams and localized fields in linear media: IV. Superpositions of sound plane waves in an ideal liquid, *J Phys A: Math Gen* **34** (2001), 6281–6289.
- [41] G.N. Borzdov, Designing localized electromagnetic fields in a source-free space, *Phys Rev E* **65** (2002), 066612(16).
- [42] G.N. Borzdov, Localized electromagnetic and elastic localized fields with applications to control laser radiation, *Advances in electromagnetics of complex media and metamaterials* (S. Zouhdi, A. Sihvola and M. Arsalane, eds), Kluwer, Dordrecht, The Netherlands, 2002, 363–383.
- [43] F.I. Fedorov, *Theory of elastic waves in crystals*, Plenum Press, New York, NY, USA, 1968.
- [44] L.B. Felsen and N. Marcuvitz, *Radiation and scattering of waves*, IEEE Press, Piscataway, NY, USA, 1994 (reprinted from Prentice-Hall, Englewood Cliffs, NJ, USA, 1973).
- [45] L.C. Biedenharn and J.D. Louck, *Angular momentum in quantum physics*, Addison-Wesley, MA, USA, 1981.
- [46] J. Mathews and R.L. Walker, *Mathematical methods of physics*, W.A. Benjamin, NJ, USA, 1964.
- [47] E. Dieulesaint and D. Royer, *Ondes élastiques dans les solides. Application au traitement du signal*, Masson, Paris, France, 1974.
- [48] N.M. Astaf'eva, Wavelet analysis: basic theory and some applications, *Usp Fiz Nauk* **166** (1996), 1145–1170.
- [49] V.C. Venugopal and A. Lakhtakia, Sculptured thin films: conception, optical properties and applications, *Electromagnetic fields in unconventional materials and structures* (O.N. Singh and A. Lakhtakia, eds), Wiley, New York, NY, USA, 2000, 151–216.
- [50] A. Lakhtakia and R. Messier, The past, the present, and the future of sculptured thin films, *Introduction to complex mediums for optics and electromagnetics* (W.S. Weiglhofer and A. Lakhtakia, eds), SPIE Press, Bellingham, WA, USA, 2003.



George N. Borzdov received his diploma (Radio-physics and Electronics), a Ph.D. degree in Physics and Mathematics, and the degree of Doctor of Science in Theoretical Physics in 1971, 1977, and 1993, respectively. He is a professor of theoretical physics at the Belarus State University. He has published more than 60 journal articles and has contributed to four research books. He is a member of SPIE—The International Society for Optical Engineering. His current research interests include wave propagation in homogeneous and nonhomogeneous linear mediums, direct and inverse scattering problems, moving linear mediums and general mathematical techniques in complex mediums.

Polarimeter for Anisotropic Optically Active Materials

Toru Asahi and Jinzo Kobayashi

Abstract

The goal of polarimetry is to determine not only the linear and circular birefringences, but also the linear and circular dichroisms in an anisotropic optically active material. The measurements of circular effects are, however, exceedingly more difficult than those of linear ones, because the latter generally overwhelm the former. The long period of 170 years had elapsed from the discovery of optical activity by Arago until the high-accuracy universal polarimeter (HAUP) invented by Kobayashi in 1983 enabled one to measure optical activity and linear birefringence simultaneously along any direction in any material. The general HAUP is as an ultimate polarimeter capable of measuring all of the linear and circular effects in an anisotropic optically active material. General HAUP theory indicates that circular dichroism can be obtained exclusively through the accurate measurement of a characteristic angle introduced in the original HAUP method. In this chapter, the original and general HAUP theories are described and some examples from HAUP experiments are presented.

1 Introduction

Two complex susceptibility tensors of the second rank, the permittivity and the permeability tensors, are the most significant physical quantities to elucidate the optical properties of a material. The second-rank permittivity tensor ε_{ij} relates the electric field vector $\mathbf{E}$ with the electric displacement vector $\mathbf{D}$ by the constitutive equation

$$D_i = \varepsilon_{ij} E_j. \quad (1)$$

(Note that in this equation a summation symbol is omitted for brevity; i.e., the right side in (1) implies $\sum_{j=1}^3 \varepsilon_{ij} E_j$. Similar convention is used in subsequent equations.)

The magnetic field vector $\mathbf{H}$ and the magnetic induction vector $\mathbf{B}$ are connected through permeability tensor μ_{ij} in the same manner:

$$B_i = \mu_{ij} H_j. \quad (2)$$

An electromagnetic wave entering the material splits into two characteristic waves, which are called eigenwaves. Their polarization states depend on the anisotropies of ε_{ij} and μ_{ij} in the material if the depolarization effect is negligibly small. They are generally a pair of elliptically polarized waves with opposite senses or, in special cases, a pair of linearly polarized or circularly polarized ones. In this chapter we confine our description to the paramagnetic materials, for which we can safely regard permeability as a scalar constant which equals the permeability of a vacuum μ_0 .

On taking the character of a complex tensor into consideration, ε_{ij} can be represented with four independent terms as

$$\begin{aligned} \varepsilon_{ij} &= \operatorname{Re} \left[\left(\frac{\varepsilon_{ij} + \varepsilon_{ji}}{2} \right) + \left(\frac{\varepsilon_{ij} - \varepsilon_{ji}}{2} \right) \right] + \sqrt{-1} \operatorname{Im} \left[\left(\frac{\varepsilon_{ij} + \varepsilon_{ji}}{2} \right) + \left(\frac{\varepsilon_{ij} - \varepsilon_{ji}}{2} \right) \right] \\ &= \varepsilon_{ij}^{\prime sym} + \varepsilon_{ij}^{\prime an} + \sqrt{-1} (\varepsilon_{ij}^{\prime\prime sym} + \varepsilon_{ij}^{\prime\prime an}), \end{aligned} \quad (3)$$

where the superscripts “sym” and “an” designate the symmetric and antisymmetric components of ε_{ij} , respectively, and single and double primes mean real and imaginary parts of ε_{ij} , respectively. From the viewpoint of crystal optics, the anisotropies in $\varepsilon_{ij}^{\prime sym}$ and $\varepsilon_{ij}^{\prime\prime sym}$ lead to linear birefringence (LB) and linear dichroism (LD), respectively; while the existences of $\varepsilon_{ij}^{\prime an}$ and $\varepsilon_{ij}^{\prime\prime an}$ cause circular dichroism (CD) and circular birefringence (CB), respectively. While, LB and CB originate from the difference in refractive indexes, designated as n , for two linearly polarized eigenwaves and two circularly polarized eigenwaves, respectively; LD and CD are brought about by the difference in absorption coefficients, designated as m , for them, respectively. Here the complex-valued optical constant N is defined as $N = n - \sqrt{-1}m$. It is reasonable that the sign before m should be minus because

it is a physical quantity concerning dissipation. With this definition, LB, CB, LD, and CD can be expressed by $\Delta n^{lin} = n_s - n_f$, $\Delta n^{cir} = n_\ell - n_r$, $\Delta m^{lin} = m_s - m_f$, and $\Delta m^{cir} = m_\ell - m_r$, respectively, where s and f represent the slower and faster linearly polarized waves when neglecting the circular effects, and r and ℓ the right and left circularly polarized waves (RCP and LCP) when neglecting the linear effects.

LB and LD can be observed in any direction apart from the optic axes for a crystal belonging to a symmetry class in crystal systems lower than the cubic system. Since LB was discovered in Iceland spar (calcite) by Bartholinus in 1669, a large number of measurements and investigations of LB have been performed on anisotropic materials with a variety of polarimeters and ellipsometers which have been modified on the basis on compensation and rotating analyzer methods. LB studies have greatly contributed not only to science (e.g., solid-state physics and electro-optics) but also to industrial fields (e.g., development of optical instruments such as polarizers, modulators, and optical filters). On the other hand, LD was first observed in tourmaline by Biot in 1815. It has also been extensively investigated so far, and, in particular, various sheet polarizers with strong LD have been used practically as convenient optical filters [1].

A substance that exhibits CB and CD is called *optically active*. A basic phenomenon representing optical activity (OA) is the rotation of the vibration plane of linearly polarized light on transmission. The rotation of the vibration plane was first observed along the optic axis of a quartz plate by Arago in 1811 [2]. After the discovery of OA, Biot carried out excellent experimental and theoretical researches on the OA of the material in a solid or solution state, and he found the law of inverse squares for optical rotatory dispersion in 1817 [3], which was modified to Drude's formula afterward [4]. Fresnel described the significant theory for the rotation of polarization caused by OA by introducing the RCP and LCP planewaves with different velocities, thereby implying CB; and he also proved his theory by means of a prism composed of three pieces of quartz crystals in 1822 [5]. On the other hand, CD was first observed by Haidinger in 1847 in a sample of amethyst quartz [6]. Cotton made a historical study on CD in an optically active potassium chromium tartrate solution in 1895 [7], and discovered the so-called Cotton effect. The study of OA, or CB, in crystalline and solution states led to the proposal of the concept of *dissymmetry* by Pasteur [8], and the many subsequent works on OA and CD related to molecular chirality opened up a new science—stereochemistry—based on the tetrahedral model of an asymmetric carbon that was proposed by van't Hoff and Le Bel independently in 1874 [9-10]. The investigations of molecular chirality using OA and CD were focused mainly on organic molecules in solution, because OA and CD in the solution state (where linear effects vanish) were easily measured by means of an ordinary polarimeter. In particular, OA and CD data for biomaterials such as amino acids and proteins have been accumulated so far [11]-[14], which has greatly contributed to the progress in biochemistry.

In contrast, studies of OA and CD in the solid state, which physicists are deeply concerned with, have not advanced so much, because the relevant measurements are difficult except in the cases of measurement along the optic axes in an

anisotropic material and in any direction for an isotropic material. That is to say, the accurate extraction of OA and CD from measurements becomes next to impossible, as LB originating from anisotropy generally overwhelms OA and CD. Szivessy and Münster first measured OA successfully in an anisotropic crystal—quartz—in a direction perpendicular to its optic axis [15]. They not only determined quantitatively the anisotropy of gyration tensor of quartz, but also found that the sign of the tensor components in a direction perpendicular to the optic axis, g_{11} and g_{22} ($=g_{11}$), is opposite to that of the component along the optic axis, g_{33} . However, their experiments were performed using specific polarization states of an incident light, and did not give simultaneous measurements of LB and OA. Konstantinova proposed a polarimetric method for measuring OA in a crystal with an elaborate polarimeter [16], and Kobayashi *et al.* presented a simultaneous measuring method for LB and OA with an advanced polarimeter [17]. Nevertheless, their methods did not enable one to determine OA in a crystal belonging to a monoclinic or triclinic system. Thus, a general polarimeter for measuring OA accurately in any crystal has been considered desirable for a long time.

Kobayashi developed the high-accuracy universal polarimeter (HAUP) [18], which can provide us simultaneously with LB, OA, and the rotation angle Ψ of the optical indicatrix of any crystal. The HAUP appeared about 170 years later than the discovery of OA, and has been applied to various kinds of solids. As a result, OA has turned out to be a very significant tool for solid state physics. The reason is that the OA reveals very accurately the existence of helical trajectories of electrons instead of their density. Examples of what the HAUP successfully provided are as follows: (i) origin of the incommensurate state of ferroelectrics [19]–[25], (ii) a new way for exploring optical properties of the monoclinic crystals [26]–[28], (iii) determination of chirality of ferroelectric domains [29]–[31], (iv) invalidation of the anyon hypothesis as the origin of high-temperature superconductors [32], (v) elucidation of the phase transition in relaxor ferroelectric ceramics [33], (vi) discovery of the enormous rotatory power of helical polymers [34], (vii) a new method for the study of absolute asymmetric synthesis [35], (viii) first measurements on amino acid crystals [36–37], (viii) first quantitative evaluation of the structural differences in a protein between its liquid and solid states [38], and (ix) the discovery of such subtle phase transition that the change of electron density distribution cannot be detected by x rays [39].

Following numerous studies using the original [40–41] and extended HAUP theories [42–43], the HAUP theory was generalized for an ultimate polarimeter capable of measuring LB, OA (CB), LD, and CD simultaneously in an anisotropic optically active material [32], [44]. Several types of HAUPs [45]–[50] and other polarimeters [51]–[53] that are essentially based on HAUP theory have been constructed so far, and thereby the significance of OA studies in anisotropic materials has been demonstrated not only in chemistry but also physics.

In this chapter, the principle of HAUP is explained and some experimental results obtained by it are briefly described.

2 Optical activity

OA originates from a kind of internal perturbation to the permittivity ε ; namely, it is caused by a resonating effect in the eigenstates of this tensor [21], [54]. This is the reason why OA contains characteristic information which cannot be obtained from other physical quantities. OA provides us with microscopic information concerning the chirality of the structure and the specific bonding nature of constituent atoms. In this sense, knowledge of OA, if any, is indispensable; and the information obtained from x-ray and neutron scattering experiments is not always sufficient for deeper understanding in solid-state physics.

Strictly speaking, the reciprocal rotation of the vibration plane of linearly polarized light of an isotropic material is called natural optical activity (NOA), while that of an anisotropic material is called OA [55]. Recently, the terms NOA and OA are both used as the reciprocal rotation for any material under no magnetic field and they are regarded as the counterpart terms to Faraday rotation, which represents the nonreciprocal rotation induced by a magnetic field. In this chapter, the focus is mainly on OA, i.e., the reciprocal rotation of an anisotropic material.

In the crystalline state, optically active crystals belong to one of 15 classes: 1 (triclinic); 2, m (monoclinic); $mm2$, 222 (orthorhombic); 4, 422 , $\bar{4}$, $\bar{4}2m$ (tetragonal); 3, 32 (trigonal); 6, 622 (hexagonal); 23, 432 (cubic) among the 21 non-centrosymmetric crystal point groups. It should be noted that a crystal in the non-enantiomorphic (achiral) classes (m , $mm2$, $\bar{4}$, $\bar{4}2m$) can exhibit OA and CD, although it has mirror-symmetry elements. The method of mixing using KBr matrix or Nujol mull allows the detection of OA or CD in a solid material, in which crystallites belonging to an enantiomorphic class are randomly distributed, provided that artifacts can be completely eliminated from the measurement. However, the mixing method yields only the spatially averaged value of OA or CD. Furthermore, in the case of a material consisting of crystallites belonging to a non-enantiomorphic class, no information on OA or CD is obtained by using either of the mixing methods, because the spatially averaged value of OA or CD becomes equal to zero.

The permittivity tensor ε_{ij} changes as a function of angular frequency ω , a phenomenon universally known as frequency dispersion. Moreover, it must be stressed that ε_{ij} also depends on the wavevector. This phenomenon is called spatial dispersion [56-57]. When spatial dispersion is taken into account, the permittivity is written as

$$\varepsilon_{ij}(\omega, \mathbf{s}) = \varepsilon_{ij}(\omega) + \varepsilon_0 \alpha_{ijl}(\omega) s_l = \varepsilon_{ij}(\omega) + \varepsilon_0 \alpha_{ijl}(\omega) \frac{\omega}{c_0} u_l, \quad (4)$$

where $\mathbf{s}$ and $\mathbf{u}$ are the wavevector and the unit vector parallel to $\mathbf{s}$, respectively; $\alpha_{ijl}(\omega)$ is a third-rank polar tensor, ε_0 is the permittivity in vacuum, and c_0 is the velocity of light in vacuum. If the energy loss can be neglected, i.e., LD and CD both equal zero and time-reversal symmetry holds, the following symmetry

conditions can be applied to (4):

$$\varepsilon_{ij}(\omega, \mathbf{s}) = \varepsilon_{ij}^*(\omega, \mathbf{s}), \quad (5)$$

$$\varepsilon_{ij}(\omega, \mathbf{s}) = \varepsilon_{ji}(\omega, \mathbf{s}). \quad (6)$$

Here, the asterisk denotes the complex conjugate. As a result, the following character of each part of (4) is derived. The first term $\varepsilon_{ij}(\omega)$ is a real-valued symmetrical tensor, identified with $\varepsilon_{ij}^{\text{sym}}$ in (3), and

$$\alpha_{ijl}(\omega) = -\alpha_{jil}(\omega) = \alpha_{jil}^*(\omega). \quad (7)$$

The third-rank polar tensor $\alpha_{ijl}(\omega)$ is thus antisymmetrical with respect to ij and it is purely imaginary. Furthermore, $\alpha_{ijl}(\omega)$ is defined by using Levi–Civita antisymmetrical tensor e_{ijm} as

$$\alpha_{ijl}(\omega) = \sqrt{-1}\delta_{ijl}(\omega) = \sqrt{-1}e_{ijm}g_{ml}(\omega), \quad (8)$$

where $\delta_{ijl}(\omega)$ is a third-rank antisymmetrical polar tensor and $g_{ml}(\omega)$ a second-rank axial tensor. Substituting (8) into (4), we obtain

$$\varepsilon_{ij}(\omega, \mathbf{s}) = \varepsilon_{ij}(\omega) + \sqrt{-1}\varepsilon_0 \frac{\omega}{c_0} e_{ijm} g_{ml}(\omega) u_l = \varepsilon_{ij}(\omega) + \sqrt{-1}\varepsilon_0 \frac{\omega}{c_0} e_{ijm} G_m(\omega). \quad (9)$$

Here, $\mathbf{G}(\omega)$ is the axial vector called a gyration vector, which is related to the unit wave vector $\mathbf{u}$ as follows:

$$G_m(\omega) = g_{ml}(\omega) u_l. \quad (10)$$

Thus, the electric displacement $\mathbf{D}$ is expressed as

$$\begin{aligned} D_i &= \varepsilon_{ij}(\omega) E_j + \sqrt{-1}\varepsilon_0 \frac{\omega}{c_0} [e_{ijm} G_m(\omega) E_j - e_{ijm} G_j(\omega) E_m] \\ &= \varepsilon_{ij}(\omega) E_j + \sqrt{-1}\varepsilon_0 \frac{\omega}{c_0} [\mathbf{G}(\omega) \times \mathbf{E}]_i. \end{aligned} \quad (11)$$

Equation (11) leads to an important conclusion that the observable perturbation in the refractive index giving rise to OA is only associated with the component of $\mathbf{G}(\omega)$ parallel to $\mathbf{u}$. Accordingly, we have only to pay attention to the scalar $G(\omega)$ as defined in the following equation:

$$G(\omega) = \mathbf{G}(\omega) \cdot \mathbf{u} = G_i(\omega) u_i = g_{ij}(\omega) u_i u_j. \quad (12)$$

For simplicity, let us consider (11) of an isotropic material, i.e., NOA. Then, (12) is expressed as

$$\mathbf{G}(\omega) = G(\omega) \mathbf{u}, \quad (13)$$

and $\varepsilon_{ij}(\omega)$ becomes the scalar $\varepsilon(\omega)$. With Faraday's law of induction, (11) for an isotropic material is represented as follows:

$$\mathbf{D} = \varepsilon(\omega)\mathbf{E} + \sqrt{-1}\varepsilon_0 \frac{\omega}{c_0} G(\omega)(\mathbf{u} \times \mathbf{E}) = \varepsilon(\omega)\mathbf{E} + \varepsilon_0 G(\omega) \text{rot } \mathbf{E} \quad (14)$$

It is found that the second term of the right side in (14) appears by considering spatial dispersion of the permittivity and gives rise to NOA.

From a tensorial point of view, $G(\omega)$ is a pseudoscalar. It is called gyration and is related to optical rotatory power as

$$\rho(\omega) = \frac{\pi}{\lambda_0} [n_\ell(\omega) - n_r(\omega)] = \frac{\pi}{\lambda_0} \frac{G(\omega)}{\bar{n}}, \quad (15)$$

where λ_0 is the wavelength of light in vacuum, and $\bar{n}$ is the average refractive index of the material, defined by $\sqrt{n_s n_f}$. Generally, $g_{ij}(\omega)$, which is called the gyration tensor, possesses symmetric and antisymmetric components. However, its antisymmetric components cannot be detected in OA experiments because the measured quantity is $\rho(\omega)$ or $G(\omega)$, which is obtained from the summation of $g_{ij}(\omega)$ with respect to subscripts i and j ; and antisymmetrical components are canceled through the summation. When considering OA, we are allowed to regard $g_{ij}(\omega)$ as a second-rank, axial and symmetric tensor [54]-[67].

3 Principle of high-accuracy universal polarimeter (HAUP)

3.1 Original HAUP method

A polarimeter ordinarily consists of a polarizer and an analyzer. Optical elements such as a compensator and phase modulator are often introduced into the system for precise measurements. In optical polarimetry, the principal optical axes of a crystal are generally taken as the reference axes of the azimuth angle of the polarizer and analyzer. In the case of a hexagonal, trigonal, tetragonal, or orthorhombic crystal, LB is not difficult to measure in this configuration. However, in the case of a monoclinic or triclinic crystal, it becomes quite difficult to measure LB accurately because the optical indicatrix of the crystal readily rotates and the direction of the principal optical axes thereby varies with a change in either wavelength or temperature.

For an isotropic material, OA is obtained from the rotation angle of the linearly polarized light emerging from the material with respect to the polarization direction of incident light. On the other hand, OA of an anisotropic material is in principle obtainable by determining both the ellipticity k of the elliptically polarized eigenwaves passing through the material and the resultant phase retardation Δ between two elliptically polarized eigenwaves. They are in most cases expressed

as

$$k = \frac{G}{2\bar{n}\Delta n^{lin}} = \frac{\Delta n^{cir}}{2\Delta n^{lin}} \quad (16)$$

and

$$\Delta = \frac{2\pi d}{\lambda_0} \Delta n^{lin}, \quad (17)$$

where d is the thickness of the specimen. However, the value of k is generally extremely small, *ca.* 10^{-3} – 10^{-4} , and so it seemed almost impossible to measure OA of anisotropic crystals, particularly monoclinic and triclinic crystals.

The high-accuracy universal polarimeter (HAUP) was invented by Kobayashi 20 years ago. It enabled the simultaneous measurement of LB, OA, and the rotation angle ψ of optical indicatrix [18]. As shown in Figure 1, consider a light beam traveling successively through the linear polarizer P, a crystal S and the linear analyzer A. The light beam propagates along one of the principal axes, say, the X_3 axis in this case. Let the azimuth angle with respect to the X_2 axis be denoted by θ , and Υ be the deflecting angle of the analyzer from the crossed Nicols conditions [18]. The essential process of the HAUP experiment is to measure accurately the intensity of the emergent light from A. Furthermore, HAUP theory has revealed that even the slight systematic errors, which have been entirely disregarded in conventional polarimetric experiments, significantly affect OA because their order of magnitude is comparable to that of k . Firstly, the polarizer and analyzer (which are nominally considered to produce linear polarization) have in fact their own parasitic ellipticities p and q , respectively. Then the relative intensity Γ , which is the ratio of the intensity I of the emergent light to I_0 , the intensity of the incident light, is a biquadratic function of θ and Υ as follows:

$$\Gamma(\theta, \Upsilon, p, q) = A(\theta, p, q) + B(\theta, p, q)\Upsilon + \Upsilon^2, \quad (18)$$

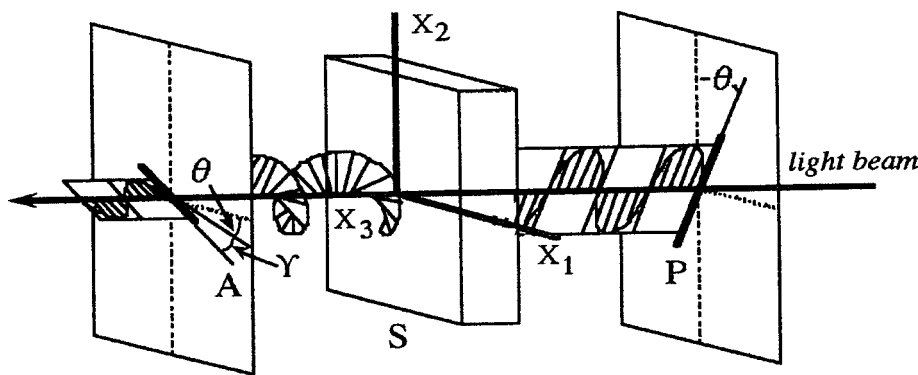


Figure 1 Schematic representation of change of polarization in the HAUP.

where

$$A(\theta, p, q) = A_0 + (p + q)^2 + 4\{k^2 - k(p - q) - pq\} \sin^2(\Delta/2) + 2(p + q) \sin \Delta \theta + 4 \sin^2(\Delta/2) \theta^2 \quad (19)$$

and

$$B(\theta, p, q) = -2(k - p) \sin \Delta + \sin^2(\Delta/2) \theta. \quad (20)$$

The HAUP method thus amounts to two-dimensional intensity measurements in the θ - Υ space [67].

As a crucial example of the serious effects of p and q on polarimetric measurements, let us consider the determination of the position of the principal axes of the indicatrix of a crystal. In the case where a specimen between P and A shows extinction, it was a commonplace belief that the two principal axes of the specimen are strictly parallel to the polarization directions of P and A, as indicated in Figure 2. However, the HAUP theory disclosed that this is not correct when p and q exist. The real extinction takes place in the direction where the polarization axes of P and A are deflected from the principal axes (X_2 and X_3) by θ_0 , called a characteristic angle in the HAUP method, as also depicted in Figure 2.

The recognition of the existence of θ_0 is one of the most important features of HAUP. The value of θ_0 is determined from the condition $(\partial \Gamma / \partial \theta)_{\Upsilon=0} = 0$, i.e.,

$$\theta_0 = -\frac{1}{2}(p + q) \cot(\Delta/2). \quad (21)$$

The relationship between θ_0 and $\Delta/2$ is shown in Figure 3, where $p + q$ is taken as a parameter. Thus, it is worth noting that not only is the occurrence of θ_0 due to systematic errors of the optical systems, but that θ_0 also depends on the phase retardation Δ . Additionally, θ_0 can be determined very accurately through an extinction phenomenon. Thereafter, we can calibrate the azimuth with respect to θ_0 .

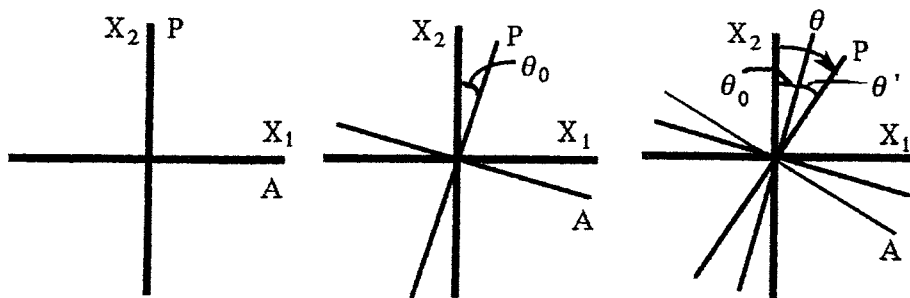


Figure 2 Characteristic angle θ_0 that can be accurately determined by an extinction experiment. X_1 and X_2 are the principal axes of the optical indicatrix. P and A represent the polarization directions of the polarizer and the analyzer, respectively.

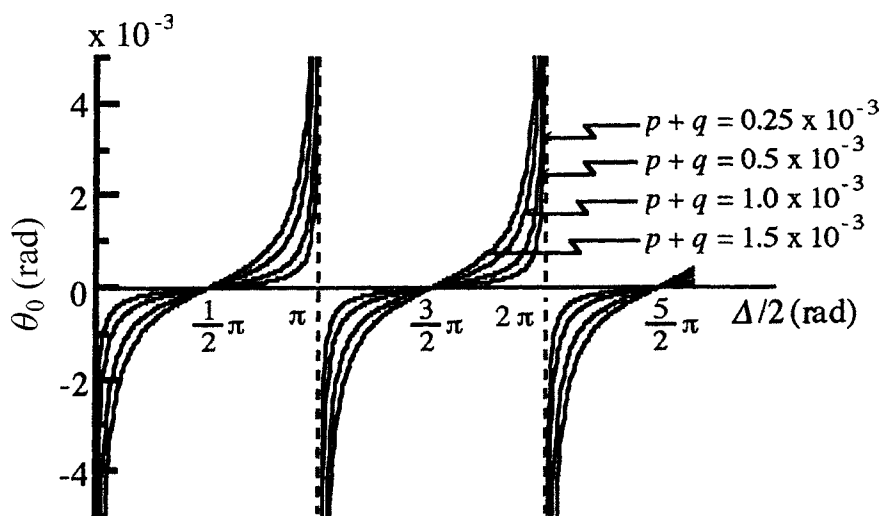


Figure 3 The change of θ_0 with an increase in $\Delta/2$, where parameters of $p+q$ are changed.

In other words, we have only to read the difference θ' between θ and θ_0 , as defined in Figure 2, whence

$$\theta = \theta_0 + \theta'. \quad (22)$$

As θ' can be measured very accurately, the HAUP equations (18)–(20) are rewritten as

$$\Gamma(\theta', \Upsilon, p, q) = A(\theta', p, q) + B(\theta', p, q)\Upsilon + \Upsilon^2, \quad (23)$$

where

$$A(\theta', p, q) = A_0 + (2k - \gamma) \sin^2(\Delta/2) + 4 \sin^2(\Delta/2) \theta'^2 \quad (24)$$

and

$$B(\theta', p, q) = (\gamma - 2k) \sin \Delta + 4 \sin^2(\Delta/2) \theta', \quad (25)$$

with $\gamma = p - q$.

It became evident with further experimentation that another important systematic error left unconsidered in the original HAUP theory also exists [40]. It is associated with the determination of Υ . We can set up the crossed Nicols position accurately without the specimen. However, the preceding crossed Nicols position that is determined accurately can vary when the specimen is inserted between the polarizer and the analyzer, because it is impossible to arrange the surfaces of the specimen to be strictly perpendicular to the light beam. The refraction of the light beam at the surfaces, even though small, cannot be avoided. Thus, we are unable to measure the correct angle of Υ .

Denoting this systematic error as $\delta\Upsilon$ and the measured angle as Υ' , we see that an analogous relation to (22) holds as follows:

$$\Upsilon = \delta\Upsilon + \Upsilon'. \quad (26)$$

On substituting (26) into (18), the modified HAUP equations can be expressed in terms of Υ' as

$$\Gamma(\theta', \Upsilon', p, q) = A(\theta', p, q) + B(\theta', p, q)\Upsilon' + \Upsilon'^2, \quad (27)$$

where

$$\begin{aligned} A(\theta', p, q) = & A_0 + (\gamma - 2k)^2 \sin^2(\Delta/2) + (\gamma - 2k) \sin \Delta \delta\Upsilon \\ & + \delta\Upsilon^2 \cos^2(\Delta/2) + 4 \sin^2(\Delta/2) \theta'^2 \end{aligned} \quad (28)$$

and

$$B(\theta', p, q) = (\gamma - 2k) \sin \Delta + 2\delta\Upsilon \cos^2(\Delta/2) + 4 \sin^2(\Delta/2) \theta', \quad (29)$$

From these equations, θ_0 is obtained again as

$$\theta_0 = -\frac{1}{2}(p + q) \cot(\Delta/2) - \frac{1}{2}\delta\Upsilon. \quad (30)$$

In the HAUP experiment, the transmitted light intensity data measured at various (θ', Υ') positions are fitted to the quadratic function (27). For example, when $\Upsilon' = 0$ (corresponding to the crossed Nicols condition) the relationship between $A(\theta', p, q)$ and θ' in (28) is obtained by changing θ' , and the coefficient $4 \sin^2(\Delta/2)$ [19] can be determined accurately through the least-squared fitting procedure. Similarly, in the measurement at $\theta' = 0$ (when the azimuth angle of the polarizer is in the extinction position), the coefficient $B(0, p, q)$ in (29) can be obtained by changing Υ' .

Information on the ellipticity k is included in $B(0, p, q)$, as seen in (29). Therefore, as the next step to obtain k accurately, the systematic errors γ and $\delta\Upsilon$ in $B(0, p, q)$ must be evaluated and removed from the experimental result. If the specimen possesses an optically inactive phase, the ellipticity is determined relatively easily, because the characteristic values γ and $\delta\Upsilon$ can be predetermined in that phase [39]. Consequently, G and Δn^{lin} of the optically active phase may be deduced from the values of k and Δ .

On the other hand, for a specimen that does not possess an optically inactive phase, the process of determining γ and $\delta\Upsilon$ becomes rather difficult, but the crucial issue has been successfully solved by means of the reference crystal method where an extra HAUP experiment is indispensably performed on the optically inactive crystal prepared as the reference [40]. Another method was also proposed to overcome the difficulty [45]. It needs two different HAUP experiments, where the fast axis is exchanged for the slow one, and *vice versa*, as the reference axis for the azimuth angle.

3.2 General HAUP method

The general HAUP method based on the original version enables one to measure not only LB and CB but also LD and CD in an anisotropic optically active material [32]. In this section, the general HAUP theory is explained by using Jones matrixes [68], which contain dichroism effects.

In the matrix method introduced by Jones, an optical effect (such as LB, LD, CB or CD) and the elliptic vibration of polarization are given by a 2×2 complex-valued matrix and a 2×1 complex-valued matrix or column vector, respectively. Furthermore, the elliptic vibration of polarization passing through a material can be expressed using the matrix representing optical effects of the material $\mathbf{M}$ and the column vector of an incident polarization $\mathbf{V}$, as $\mathbf{M}\mathbf{V}$. By modifying Jones notation, let us denote the N-matrixes that are specific matrixes representing LB, LD, CB and CD for an infinitely thin material as Θ_3 , Θ_4 , Θ_5 , and Θ_6 , respectively; and they are expressed as follows [32], [68]:

$$\Theta_3 = g_0 \begin{bmatrix} \sqrt{-1} & 0 \\ 0 & -\sqrt{-1} \end{bmatrix}, \quad (31)$$

$$\Theta_4 = p_0 \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \quad (32)$$

$$\Theta_5 = \eta \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \quad (33)$$

and

$$\Theta_6 = \delta \begin{bmatrix} 0 & -\sqrt{-1} \\ \sqrt{-1} & 0 \end{bmatrix}, \quad (34)$$

where

$$\eta = \frac{1}{2} \left(\frac{2\pi}{\lambda_0} \right) \Delta n^{cir} \equiv \frac{\pi}{\lambda_0} \frac{G}{\bar{n}}, \quad (35)$$

$$g_0 = \frac{1}{2} \left(\frac{2\pi}{\lambda_0} \right) \Delta n^{lin} \equiv \frac{1}{2} \frac{1}{d} \Delta, \quad (36)$$

$$\delta = \frac{1}{2} \left(\frac{2\pi}{\lambda_0} \right) \Delta m^{cir} \equiv \frac{1}{2} \frac{1}{d} \kappa, \quad (37)$$

and

$$p_0 = \frac{1}{2} \left(\frac{2\pi}{\lambda_0} \right) \Delta m^{lin} \equiv \frac{1}{2} \frac{1}{d} E. \quad (38)$$

The parameters κ and E are total circular and total linear absorptions, respectively. In the general HAUP theory, the N-matrix of an infinitely thin material possessing

all LB, LD, CB and CD, defined as Θ_H , is expressed as

$$\Theta_H = \Theta_3 + \Theta_4 + \Theta_5 + \Theta_6 = \begin{bmatrix} \sqrt{-1}g_0 + p_0 & -(\eta + \sqrt{-1}\delta) \\ \eta + \sqrt{-1}\delta & -(\sqrt{-1}g_0 + p_0) \end{bmatrix}. \quad (39)$$

According to Jones formalism, the Jones matrix $\mathbf{M}_H$ of a material possessing all LB, LD, CB and CD with the thickness d can be obtained using Θ_H . The Jones matrix $\mathbf{M}_H$ of the general HAUP method is written as

$$\mathbf{M}_H = \begin{bmatrix} \cosh Qd + (\sqrt{-1}g_0 + p_0)\frac{\sinh Qd}{Q} & -(\eta + \sqrt{-1}\delta)\frac{\sinh Qd}{Q} \\ (\eta + \sqrt{-1}\delta)\frac{\sinh Qd}{Q} & \cosh Qd - (\sqrt{-1}g_0 + p_0)\frac{\sinh Qd}{Q} \end{bmatrix}, \quad (40)$$

where

$$Q = \sqrt{(p_0 + \sqrt{-1}g_0)^2 - (\eta + \sqrt{-1}\delta)^2} \equiv x + \sqrt{-1}y. \quad (41)$$

The Jones vectors of the polarization produced by the polarizer and the analyzer are expressed respectively as

$$\mathbf{P} = \begin{bmatrix} \cos \theta \cos p + \sqrt{-1} \sin \theta \sin p \\ \sin \theta \cos p - \sqrt{-1} \cos \theta \sin p \end{bmatrix} \quad (42)$$

and

$$\mathbf{A} = \begin{bmatrix} -\sin(\theta + \Upsilon) \cos q + \sqrt{-1} \cos(\theta + \Upsilon) \sin q \\ \cos(\theta + \Upsilon) \cos q + \sqrt{-1} \sin(\theta + \Upsilon) \sin q \end{bmatrix}. \quad (43)$$

Then, the relative intensity ratio of the transmitted light through the P-S-A system is given as

$$\Gamma(\theta, \Upsilon) = I/I_0 = |\mathbf{A}^\dagger \mathbf{M}_H \mathbf{P}|^2, \quad (44)$$

where $\mathbf{A}^\dagger$ is the conjugate transposed vector of $\mathbf{A}$ [67].

When HAUP conditions are applied, p_0 , g_0 , η and δ appear in the HAUP equations up to the second power of θ and Υ . Then,

$$\Gamma(\theta, \Upsilon) = A(\theta) + B(\theta)\Upsilon + C\Upsilon^2, \quad (45)$$

where

$$A(\theta) = H_{11} + H_{12}\theta + H_{13}\theta^2, \quad (46)$$

$$B(\theta) = H_{21} + H_{22}\theta, \quad (47)$$

and

$$C = H_{31}. \quad (48)$$

The coefficients H_{ij} in (45)–(48) are available elsewhere [32]. Furthermore, the relative intensity in the (θ', Υ') space is given by

$$I/I_0 = \Gamma(\theta', \Upsilon') = A''(\theta') + B''(\theta')\Upsilon' + C''\Upsilon'^2, \quad (49)$$

where

$$A''(\theta') = H''_{11} + H''_{12}\theta' + H''_{13}\theta'^2, \quad (50)$$

$$B''(\theta') = H''_{21} + H''_{22}\theta', \quad (51)$$

and

$$C'' = H''_{31}. \quad (52)$$

Each coefficient except H''_{11} is expressed as follows:

$$H''_{12} = 0, \quad (53)$$

$$H''_{13} = \frac{2(p_0^2 + g_0^2)(\cosh 2xd - \cos 2yd)}{x^2 + y^2}, \quad (54)$$

$$\begin{aligned} H''_{21} = & \frac{(p_0\delta - g_0\eta)(g_0x - p_0y) \sinh 2xd}{(x^2 + y^2)(p_0^2 + g_0^2)} + \frac{(p_0\delta - g_0\eta)(p_0x + g_0y) \sin 2yd}{(x^2 + y^2)(p_0^2 + g_0^2)} \\ & + \frac{(p - q)[(g_0x - p_0y) \sinh 2xd + (p_0x + g_0y) \sin 2yd]}{x^2 + y^2} \\ & - \frac{(p - q)[(p_0x + g_0y)^2 - (p_0y - g_0x)^2] \sinh 2xd \sin 2yd}{(x^2 + y^2)(p_0^2 + g_0^2)(\cosh 2xd - \cos 2yd)} \\ & + \frac{(p + q)(p_0x + g_0y)(p_0y - g_0x)(\sinh^2 2xd - \sin^2 2yd)}{(x^2 + y^2)(p_0^2 + g_0^2)(\cosh 2xd - \cos 2yd)} \\ & - \frac{\delta\Upsilon[(p_0x + g_0y)^2 \sinh^2 2xd + (p_0y - g_0x)^2 \sin^2 2yd]}{(x^2 + y^2)(p_0^2 + g_0^2)(\cosh 2xd - \cos 2yd)} \\ & + \frac{\delta\Upsilon[(x^2 + y^2 - \eta^2 - \delta^2) \cosh 2xd + (x^2 + y^2 + \eta^2 + \delta^2) \cos 2yd]}{x^2 + y^2} \\ & + \frac{2\delta\Upsilon(p_0x + g_0y)(p_0y - g_0x) \sinh 2xd \sin 2yd}{(x^2 + y^2)(p_0^2 + g_0^2)(\cosh 2xd - \cos 2yd)}, \end{aligned} \quad (55)$$

$$H''_{22} = \frac{2(p_0^2 + g_0^2)(\cosh 2xd - \cos 2yd)}{x^2 + y^2} + \frac{2[(p_0x + g_0y) \sinh 2xd + (p_0y - g_0x) \sin 2yd]}{x^2 + y^2}, \quad (56)$$

$$H''_{31} = \frac{(p_0^2 + g_0^2 - \eta^2 - \delta^2 + x^2 + y^2) \cosh 2xd}{2(x^2 + y^2)} - \frac{(p_0^2 + g_0^2 - \eta^2 - \delta^2 - x^2 - y^2) \cos 2yd}{2(x^2 + y^2)} + \frac{(p_0x + g_0y) \sinh 2xd + (p_0y - g_0x) \sin 2yd}{x^2 + y^2}. \quad (57)$$

The extinction position θ_0 is expressed as

$$\theta_0 = -\frac{(p + q)[(g_0x - p_0y) \sinh 2xd + (p_0x + g_0y) \sin 2yd]}{2(p_0^2 + g_0^2)(\cosh 2xd - \cos 2yd)} - \frac{\delta \Upsilon [(p_0x + g_0y) \sinh 2xd + (p_0y - g_0x) \sin 2yd]}{2(p_0^2 + g_0^2)(\cosh 2xd - \cos 2yd)}, - \frac{1}{2} \delta \Upsilon + \frac{p_0 \eta + g_0 \delta}{2(p_0^2 + g_0^2)}. \quad (58)$$

These are the general expressions of the HAUP theory.

Now let us apply the approximation conditions to these equations, when CB and CD are extremely small compared with LB and LD; i.e.,

$$g_0, p_0 \gg \eta > \delta \neq 0 \quad (59)$$

and

$$\eta^2 = \delta^2 = \eta \delta = 0. \quad (60)$$

It follows from (59) and (60) that $x = p_0$ and $y = g_0$ in (41). Accordingly, H''_{ij} and θ_0 are rewritten as follows:

$$H''_{13} = e^E + e^{-E} - 2 \cos \Delta, \quad (61)$$

$$H''_{21} = \frac{2\delta \Upsilon \sin^2 \Delta}{e^E + e^{-E} - 2 \cos \Delta} - \frac{2 \sin \Delta [(p - q) \cos \Delta - p e^{-E} + q e^E]}{e^E + e^{-E} - 2 \cos \Delta} - \frac{2k}{K^2 + 1} \sin \Delta, \quad (62)$$

$$H''_{22} = 2(e^E - \cos \Delta), \quad (63)$$

$$H''_{31} = e^E, \quad (64)$$

and

$$\theta_0 = \frac{-(p-q)\sin\Delta}{e^E + e^{-E} - 2\cos\Delta} - \frac{(e^E - \cos\Delta)\delta\Upsilon}{e^E + e^{-E} - 2\cos\Delta} + \frac{Kk + k'}{K^2 + 1}, \quad (65)$$

where

$$k = \frac{\eta}{2g_0}, \quad k' = \frac{\delta}{2g_0}, \quad K = \frac{p_0}{g_0} = \frac{E}{\Delta}.$$

The foregoing equations yield the original HAUP equations (27)–(30) when $p_0 = \delta = 0$. Analogous to the original HAUP theory, H''_{ij} and θ_0 can be determined by the two-dimensional intensity measurement in the (θ', Υ') space. It should be noted that θ_0 contains the CD effect. In other words, the accurate determination of the extinction position is quite important and essential in measuring CD of an anisotropic optically active material. This idea has never been presented in the theory of conventional optical measurements so far.

The final step for obtaining OA and CD in absorbing crystals with the general HAUP is to eliminate the systematic errors p , q and $\delta\Upsilon$ in (62) and (65), which is not presented here because of its complexity [32], [44].

4 Examples of experimental results

4.1 BaMnF₄

The optical indicatrix of a monoclinic or triclinic crystal is readily rotated when external conditions are altered. This phenomenon renders the measurement of even LB using an ordinary polarimeter extremely difficult. Therefore, the HAUP was applied to BaMnF₄, which is monoclinic [28].

The BaMnF₄ crystal has attracted much interest because it exhibits a peculiar incommensurate (IC) phase below 250 K and an antiferromagnetic phase below 26 K, in addition to the pyroelectric phase in the whole temperature range. However, as the average structure of the IC phase was considered monoclinic, accurate measurement of LB and OA in this crystal did not seem possible. Temperature changes of all birefringence and gyration tensor components, and rotation angle of the optical indicatrix were successfully measured with the HAUP. The gyration tensor components of BaMnF₄ are represented in Figure 4 as functions of temperature. These results revealed not only that the components g_{11} and g_{22} appear below the IC transition temperature (T_i) but also that the g_{23} component manifests a sharp peak at the same temperature. The peak of OA has never been reported so far for any crystal, but it deserves further study as a new physical phenomenon.

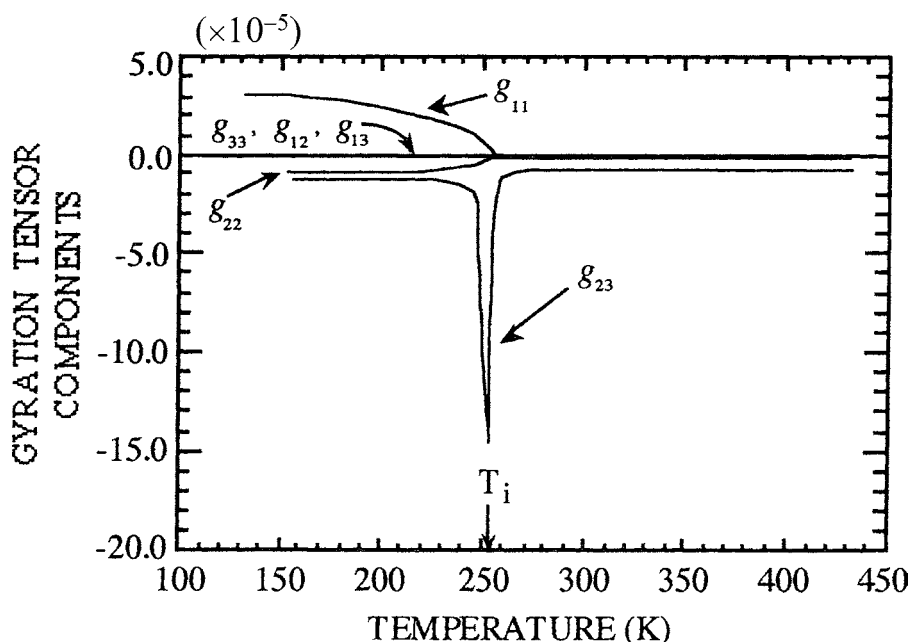


Figure 4 Temperature dependencies of gyration tensor components in BaMnF₄.

4.2 Poly-L-lactic acids

The structure of certain polymers is helical. It would then be natural to expect that such helical conformations will produce a large amount of optical activity when compared with that of ordinary crystals. However, there were no direct methods for testing this interesting idea. The relationship between the helical structure and OA in the solid polymer with preferred orientation was first revealed by applying the HAUP and x-ray diffraction methods to the drawn film of a well-known biopolymer, poly-L-lactic acid (PLLA) [34].

Plenty of PLLA film studies have been reported because its biodegradability makes it useful as a medical material. A precise x-ray study was performed on a drawn-fiber film of PLLA, thereby refining the crystal structure of PLLA as illustrated in Figure 5. The normalized density function $\gamma'_R(\zeta)$ was also determined for the films with various values of R , which is the ratio of the length of a drawn film to that of the original film, and ζ is the Eulerian spiral angle between the axis of drawing and the crystallographic c axis of a crystallite in the film. This function represents the density of crystallites, which are oriented in ζ direction, normalized with respect to the number of crystallites in the specimen drawn by a given value of R [34].

The same specimen used for x-ray experiments were subjected to the HAUP methods. Temperature dependencies of G of PLLA with different R are shown in Figure 6. Gyration was almost independent of temperature, but decreased markedly

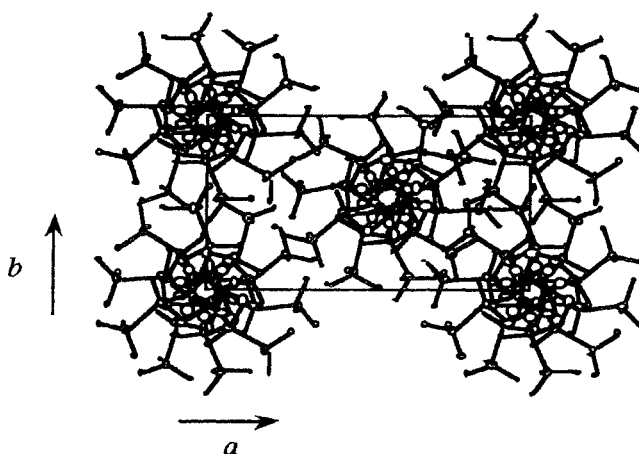


Figure 5 Crystal structure of PLLA in the (001) plane.

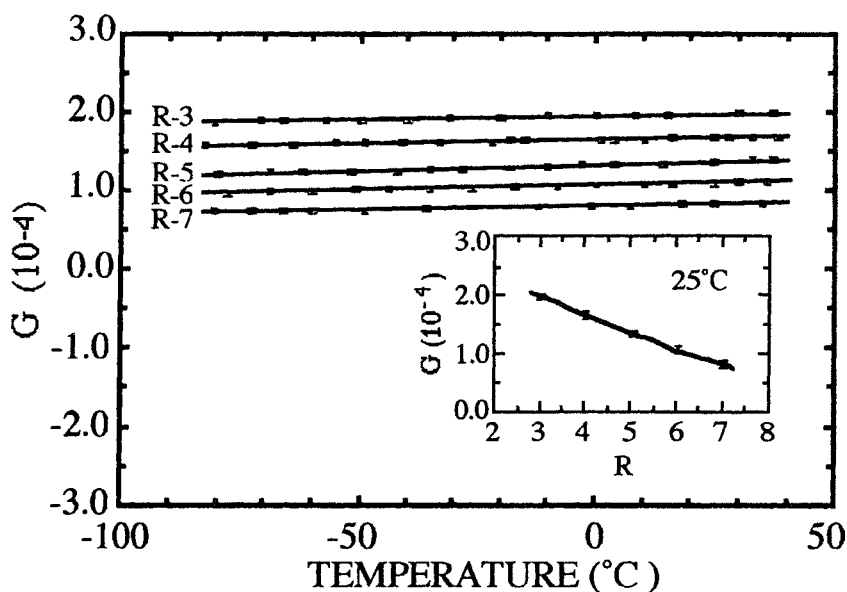


Figure 6 Temperature dependencies of gyration of PLLA along [100] with various drawn ratios R , where the curve of gyration versus R is shown in the inset.

with an increase in R . These results were the first comprehensive measurements of OA in solid polymers.

The gyration of the fiber film drawn by R for $\mathbf{u} (1, 0, 0)$, which denotes the unit wavevector in the direction perpendicular to the film surface, can be calculated as

$$\langle G_R \rangle = \int_0^{2\pi} \gamma'_R(\zeta) \int_0^{\pi/2} \langle G_\zeta \rangle \sin \zeta d\xi d\zeta, \quad (66)$$

where

$$\langle G_{\zeta} \rangle = \frac{1}{2} (g_1 + \cos^2 \zeta g_1 + \sin^2 \zeta g_3), \quad (67)$$

$$g_1 = \frac{1}{2} (g_{11} + g_{22}), \quad (68)$$

and

$$g_3 = g_{33}. \quad (69)$$

In these relations, g_{11} , g_{22} , and g_{33} are gyration components of a crystallite of PLLA. The value of $\langle G_R \rangle$, which equals G in Figure 10, is expressed as

$$\langle G_R \rangle = a_R g_1 + b_R g_3, \quad (70)$$

where

$$a_R = \pi \int_0^{\pi/2} \gamma'_R(\zeta) (1 + \cos^2 \zeta) \sin \zeta d\zeta, \quad (71)$$

and

$$b_R = \pi \int_0^{\pi/2} \gamma'_R(\zeta) \sin^3 \zeta \sin \zeta d\zeta. \quad (72)$$

The values of a_R and b_H were determined graphically. By using (70)–(72), g_1 and g_3 were obtained by using $G (= \langle G_R \rangle)$ of any pair of differently drawn specimens. From G of R -3 (film drawn by three times) and R -7 (film drawn by seven times) it follows that $g_1 = -(6 \pm 3) \times 10^{-5}$ and $g_3 = (3.85 \pm 0.69) \times 10^{-2}$. The most important conclusion is that $g_3 (=g_{33})$ of PLLA is extremely large, about a few orders of magnitude larger than any gyration component of common crystals. The rotatory powers of the direction perpendicular to the helical axis of PLLA molecule and parallel to it are $\rho_1 = -(14 \pm 7)$ deg/mm, and $\rho_3 = (9.2 \pm 1.7) \times 10^3$ deg/mm, respectively. In Table 1, they are compared with rotatory powers of various crystals. Clearly, the helical conformation of molecule in polymers like PLLA produces huge OA. This phenomenon is important for elucidating OA and CD of condensed matter, and will be promising for new optical applications utilizing such large OA.

4.3 Lysozyme crystal

Biological functions of proteins are mainly displayed when they are solutes in liquid solutions. Accordingly, it is necessary to determine the structures of proteins in solution. Protein crystals have been analyzed accurately using x-ray diffraction, while structures of protein molecules in solution have been exclusively analyzed by nuclear magnetic resonance (NMR) [69]. The structures deduced from the two

methods are similar in the main backbone of polypeptide chains. However, there is good evidence that the conformation found by x-ray analysis is in dynamic equilibrium in the solution with a structure slightly different from its solid counterpart [70]. The structural differences had to be determined. Although Moffitt and Yang [11] have given the relationship between OA and the secondary structure of proteins in solution, it had been impossible to discuss the structural difference between two states by using OA, because OA was not successfully measured for any protein crystals. In other words, protein OA studies had not advanced until the HAUP entered the field of protein science.

According to x-ray analysis [71], the structure of lysozyme consisting of 129 residues can be divided into two domains: one domain (I) is composed of residues 1 to 35 and 85 to 129, the other (II) residues 36 to 84. The I domain contains four α -helices and a short 3_{10} helix, while the II domain contains a deformed 3_{10} helix. An α -helix is identified with a nonintegral 18/5 helix presented by Pauling and a 3_{10} helix with a threefold screw axis [72].

Single crystals of hen egg-white lysozyme chloride were prepared by the sitting-drop method of Alderton *et al.* [73]. The x-ray inspection of a specimen

Table 1 Rotatory powers of various crystals.

crystal	λ_0 (Å)	T (K)	ρ (deg cm ⁻¹)
α -SiO ₂	6328	293	$\rho_1 = -1.05 \times 10^{2a}$ $\rho_3 = 2.51 \times 10^2$
Rochelle Salt	6328	298	$\rho_1 = -9.6^b$
BaMnF ₄	6328	160	$\rho_1 = 5.06 \times 10^c$ $\rho_2 = -2.79 \times 10$ $\rho_3 = 1.22 \times 10$
PLLA	5145	298	$\rho_1 = 1.4 \times 10^{2d}$ $\rho_3 = -9.2 \times 10^4$
(NH ₄) ₂ SO ₄	6328	200	$\rho_1 = 3.86 \times 10^e$
Bi ₂ Sr ₂ CaCu ₂ O ₈	4880	20	$\rho_3 = 3.38 \times 10^{2f}$
L-glutamic acid	6328	293	$\rho_1 = -3.0 \times 10^{2g}$ $\rho_2 = -9.78 \times 10^2$ $\rho_3 = 5.4 \times 10$
L-aspartic acid	6328	293	$\rho_1 = 2.23 \times 10^{3h}$ $\rho_2 = -1.86 \times 10^2$ $\rho_3 = -3.0v \times 10^2$
lysozyme	4880	303.4	$\rho_1 = -2.13 \times 10^i$ $\rho_3 = 2.48 \times 10$

^aRef. 41.

^bRef. 27.

^cRef. 28.

^dRef. 34.

^eRef. 31.

^fRef. 32.

^gRef. 36.

^hRef. 37.

ⁱRef. 38.

confirmed that this material belongs to tetragonal D_4 point group [71]. Conoscopic patterns displayed that the a and c axes correspond to the optical X' (vibration direction of the faster polarized light) and Z' (vibration direction of the slower polarized light), respectively, in agreement with a previous report [74].

Protein crystals immediately lose their crystallinity in vacuum. Therefore, measurements of OA in proteins impose an additional difficulty, which can be removed only by using a special humidifier. In order to preserve the crystallinity of thin specimens in the HAUP system, a humidifier was constructed to produce 100% humidity at approximately 305 K temperature [38].

Figure 7 depicts OA of lysozyme as a function of the wavelength in vacuum. The gyration tensor components of lysozyme at $T = 303.4$ K and $\lambda_0 = 4880$ Å are $g_{11} = -0.90 \times 10^{-5}$ and $g_{33} = 1.05 \times 10^{-5}$; and the rotatory powers along the a and c axes are thus calculated as $\rho_1 = -21.3$ deg/cm and $\rho_3 = -24.8$ deg/cm. As compared with the rotatory powers of various crystals in Table 1, those of lysozyme are rather small. Moreover, the anisotropy of lysozyme is very large as can be seen from the fact that signs of ρ_1 and ρ_3 are opposite. This is a remarkable property of the lysozyme crystal.

Let us recall the OA of PLLA, which is composed of two parallel helical strands as schematically shown in Figure 8(a). The structure of lysozyme is shown in Figure 8(b). As described in Section 4.2, PLLA displays a huge rotatory power of nearly 10^5 deg/cm along the helical axis (c axis). Therefore, it is surprising that lysozyme, which contains a considerable amount of helices, manifests low rotatory power values.

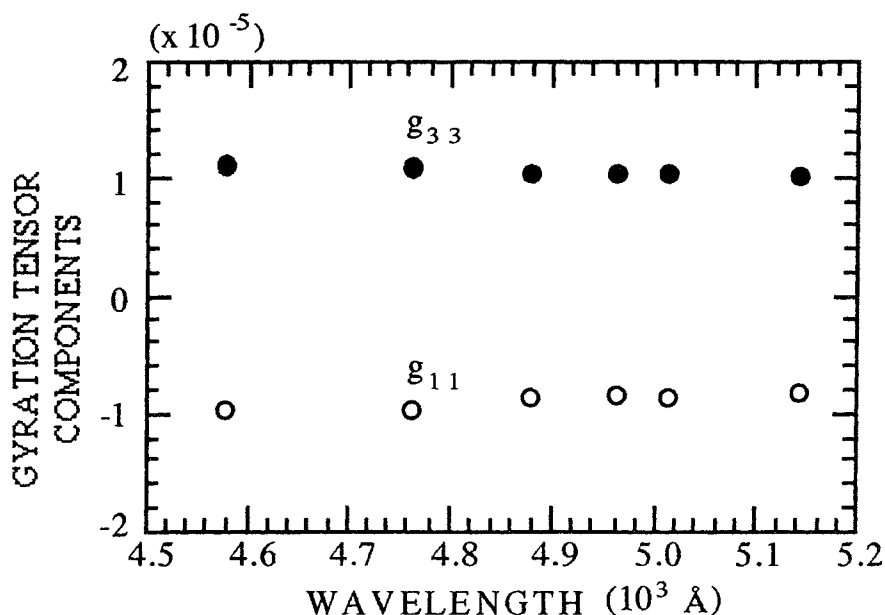


Figure 7 Wavelength dependencies of gyration tensor components of lysozyme.

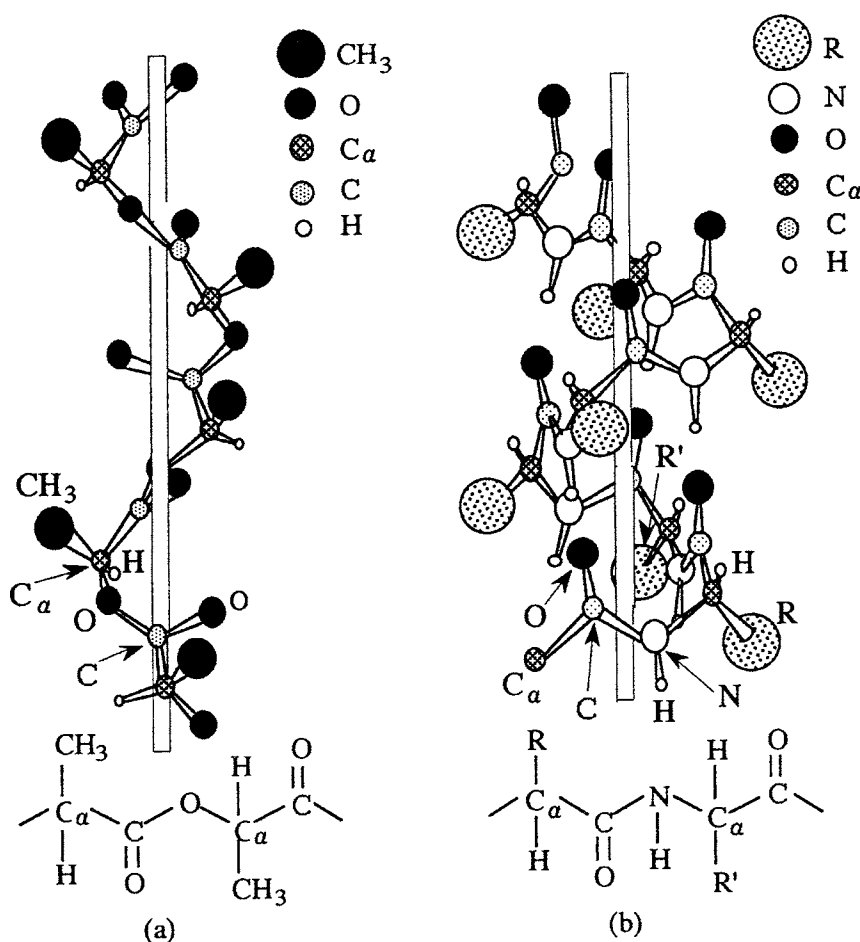


Figure 8 Schematic representation of the helical conformations of (a) PLLA and (b) lysozyme.

With reference to Ramachandran's theory [75], the rotatory power produced by a helix composed of a particular kind of atom can be approximately expressed as

$$\rho = C \frac{l^2 t f F}{L} (\alpha_r^2 - \alpha_t^2). \quad (73)$$

Here, l designates the distance from the atom to the helical axis, L the pitch of the helix, α_r and α_t the polarizability components of the atom in the direction perpendicular and tangential to the radius of a helix, respectively, t the number of turns of the helix contained in the unit thickness (1 mm) of the specimen, f the number of atoms contained in one turn of the helix, F the number of the same helices contained in a unit cell, and C a constant.

On the other hand, it has been found that I domain of lysozyme contains four α -helices and a short 3_{10} helix; and that the main-chain structures in them ob-

tained from NMR studies are almost the same as those determined by x-ray diffraction except for the side-chain structures of exposed residues on the protein surface [70]. On assuming that a B helix (24-36 residues) in I domain that is almost parallel to the c axis only contributes to ρ along the c axis in lysozyme and the polarizability of O atom predominates over those of other atoms, ρ of both crystals can be approximately compared. The parameters in (73) except for the polarizabilities are shown in Table 2. It can be estimated by using the parameters that ρ along the c axes should be approximately equal to or slightly larger than ρ of PLLA. In reality, however, the rotatory power of lysozyme is smaller than that of PLLA by four orders of magnitude. Further detailed investigation is necessary to explain the difference between theoretically predicted and experimental results.

Optical activity increases in essence by increase of the severity of the restriction of the freedom of molecular orientation by forming a crystal lattice [76-77]. In order to represent a specific quantity of this additive crystal-structure contribution to OA in a crystal, a *chirality index* r is defined as [36]

$$r = \frac{\rho_C^0 - \rho_S^0}{\rho_C^0} = 1 - \frac{\rho_S^0}{\rho_C^0}. \quad (74)$$

Here, ρ_C^0 is the rotatory power per molecule in the crystalline state, while ρ_S^0 is that in solution. As a typical example, r of quartz is equal to 1 since it loses OA in the liquid state. On the other hand, $r = 0$ for the materials whose structure is the same in both states. A few amino acids and PLLA manifest r of almost unity, as shown in Table 3. Crystals with $r = 0$ have never been found so far.

Table 2 Parameters in (73) for lysozyme and poly-L-lactic acid.

parameter	lysozyme	poly-L-lactic acid
L	5.4 Å	9.6 Å
l	1.74 Å (O _{carbonyl}) 0.94 Å (O _{carbonyl})	1.24 Å (O _{ester})
t	$9.52 \times 10^5 \text{ mm}^{-1}$	$1.04 \times 10^6 \text{ mm}^{-1}$
f	3.6	3.3
F	8	2

Table 3 Chirality index r of some crystalline amino acids and lysozyme.

crystal	R
PLLA	0.999
glutamic acid	0.998
aspartic acid	0.994
lysozyme	0.16

Let us calculate r for lysozyme. The rotatory power ρ_C of the assemblage of 1 ml is obtained from the experimental results as follows:

$$\rho_C = \frac{180}{\lambda_0 \bar{n}} \bar{g} = \frac{180}{\lambda_0 \bar{n}} \frac{(2g_{11} + g_{33})}{3} = -5.9 \text{ deg cm}^{-1} \text{ ml}^{-1}, \quad (75)$$

where $\bar{g}$ is the average gyration tensor in the crystalline state [33]. By using (75), the value of ρ_C^0 is calculated as

$$\rho_C^0 = \frac{\rho_C}{N_C} = -1.75 \times 10^{-19} \text{ deg cm}^{-1}, \quad (76)$$

where

$$N_C = \frac{Z}{a^2 c} = 3.37 \times 10^{-19} \text{ deg ml}^{-1}, \quad (77)$$

$Z = 8$ being the number of molecules in the unit cell. On the other hand, ρ_S^0 is readily calculated by using the reported OA in solution as [78]

$$\rho_S^0 = \frac{-89.3 \times 10^3}{N_A} = -1.48 \times 10^{-19} \text{ deg cm}^{-1}, \quad (78)$$

where N_A designates Avogadro's number. From (76) and (78), $r = 0.16$. The chirality index of lysozyme exhibits a very remarkable contrast with that of amino acids; see Table 3. Namely, r of lysozyme is less than those of other crystals by one order of magnitude.

Let us say that the ability of producing OA is changed by $x\%$ when lysozyme is transformed from the solution to the crystalline state. Then,

$$x = \frac{\rho_C^0 - \rho_S^0}{\rho_S^0} \times 100 = \frac{r}{1 - r} \times 100 = 19.0. \quad (79)$$

Lysozyme in solution undergoes a structural change that can be estimated as a 19% increase in the ability of producing OA on attaining the crystalline state.

The NMR method has disclosed not only that the calculated structures for the side-chains of exposed residues on the protein surface are considerably disordered in solution, but also that the 3_{10} helix in II domain confirmed by x-ray experiments could not be found in the folding of the solution structures [70]. Therefore, it is very likely that the major part of the enhancement in ρ is brought about by an increase of the constraints of atomic motion in the side-chains of exposed residues and probably by the 3_{10} helix in II domain (which disappears in the solution state). The first confirmation of the structural change between the two states of lysozyme was thus provided by the HAUP experiments, and quantitatively to boot.

4.4 Silver thiogallate

Silver thiogallate (AgGaS_2) belongs to the tetragonal nonenantiomorphic $\bar{4}2m$ class. It is a well-known crystal whose rotatory power was measured by utilizing accidental isotropy at the wavelength of 4974 Å [79]. Besides, the measured value of ρ is as large as 522 deg/mm which is the largest known value in the visible regime among inorganic crystals.

Temperature dependencies of all the coefficients H''_{ij} of (61)–(65) at $\lambda_0 = 6328$ Å were measured using the general HAUP formalism. By applying the least-squared fitting method to the data of H''_{21} and θ_0 , the optical parameters of Δn^{lin} , ρ , Δm^{lin} , and Δm^{cir} were calculated as functions of temperature. The obtained results are shown in Figure 9. The values of Δn^{lin} and ρ almost coincide with those in an earlier report [80], and their temperature dependencies also agree perfectly. However, it was not expected that Δm^{cir} would be larger than Δm^{lin} in silver thiogallate. The present results obtained by using the approximation (59) may need revision, however minor.

5 Chiral physics

OA extracts the presence of electrons whose trajectories are confined to a helix-like structure, such electrons generating antisymmetrical parts of susceptibilities. In this sense, OA can be a new tool for examining crystal structure from a different viewpoint than the x-ray diffraction. The general theoretical clarification of the relationship between OA and crystal structure, however, is an almost unknown field owing to the extreme lack of experimental data for anisotropic crystals.

Furthermore, many phenomenons may have been left unresolved due to a disregard of the effects of axial tensors such as a gyration tensor, since they are generally too small to be observed with an ordinary instrument. These issues can be reasonably resolved by using the HAUP to measure OA. It is exceedingly worthwhile to study the effects of chiral molecular structure on physical properties in anisotropic materials by utilizing information of OA and, furthermore, to develop new methods for measuring other axial tensors with small magnitude. It can be said that the HAUP study contributes greatly to the development of an interdisciplinary field that combines physics, chemistry and biology, which may be called *chiral physics* [81].

Acknowledgment

The authors thank Professor S. C. Abrahams of Southern Oregon University for useful suggestions.

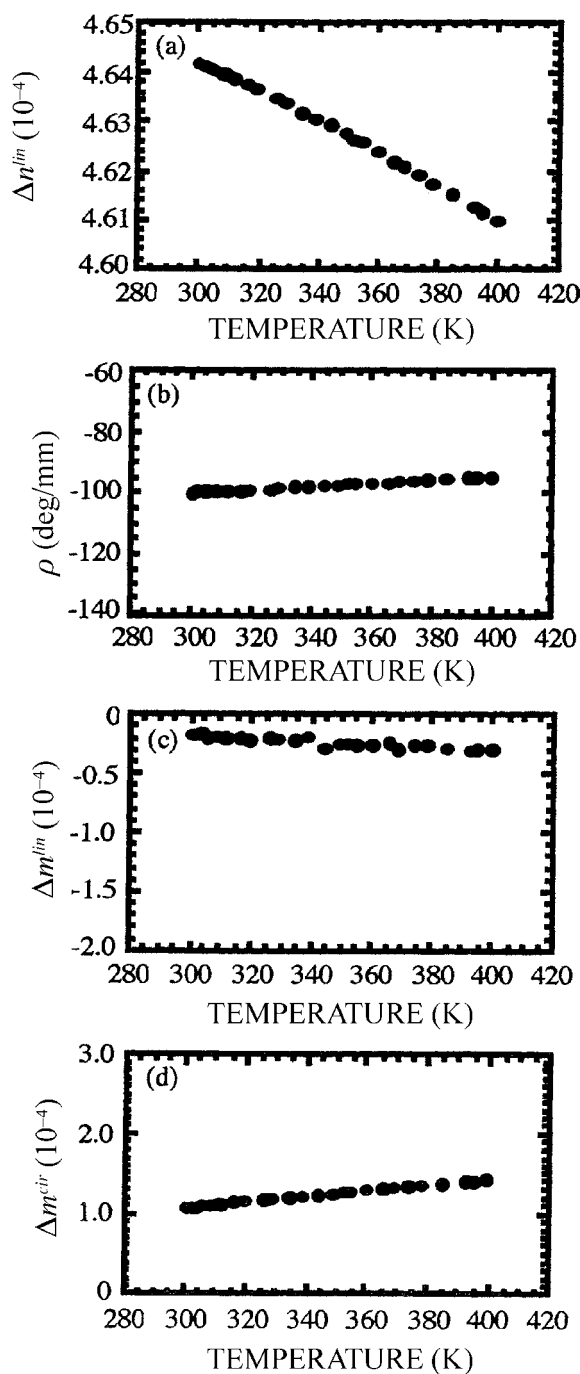


Figure 9 Temperature dependencies of (a) linear birefringence, (b) optical rotatory power, (c) linear dichroism, and (d) circular dichroism of AgGaS_2 .

References

- [1] W.A. Shurcliff, *Polarized light*, Harvard University Press, Cambridge, MA, USA, 1966.
- [2] D.-F. Arago, Mémoire sur une modification remarquable qu'éprouvent les rayons lumineux dans leur passage à travers certains corps diaphanes, et sur quelques autres nouveaux phénomènes d'optique, *Mém Cl Sci Math Phys Inst France* **1** (1811), 93–134.
- [3] J.B. Biot, Mémoire sur les rotations que certains substances impriment aux axes de polarization des rayons lumineux, *Mém Acad Sci (Paris)* **2** (1817), 41–136.
- [4] P. Drude, *Lehrbuch der optik*, S. Hirzel, Leipzig, Germany, 1900.
- [5] A. Fresnel, Double réfraction du verre comprime, *Oeuvres complètes d'Augustin Fresnel* **1**, Johnson Reprint Corp, New York, USA, 1965, 713–751.
- [6] W. Haidinger, Über den Pleochroismus des Amethysts, *Ann Phys Lpz* **70** (1847), 531–544.
- [7] A. Cotton, Absorption inégale des rayons circulaires droit et gauche dans certains corps actifs, *Comptes rendus hebdomadaires des séances de l'Académie des sciences* **120** (1895), 989–991.
- [8] L. Pasteur, Recherches sur les relations qui peuvent exister entre la forme cristalline et la composition chimique, et le sens de la polarisation rotatoire, *Ann Chim Phys* **24** (1848), 442–459.
- [9] J.H. Van't Hoff, Sur les formules de structure dans l'espace, *Archives néerlandaises des sciences exactes et naturelles* **9** (1874), 445–454.
- [10] J.A. Le Bel, Sur les relations qui existent entre les formules atomiques des corps organiques et le pouvoir rotatoire de leurs dissolutions, *Bull Soc Chim* **22** (1874), 337–347.
- [11] W. Moffitt and J.T. Yang, The optical rotatory dispersion of simple polypeptides. I, *Proc Nat Acad Sci USA* **42** (1956), 596–603.
- [12] C. Djerassi, *Optical rotatory dispersion*, McGraw-Hill, New York, NY, USA, 1960.
- [13] L.I. Katzin and E. Gulyas, Optical rotatory dispersion of some amino acids and criteria of protein conformation, *J Am Chem Soc* **80** (1963), 1655–1659.
- [14] E. Shechter and E. Blout, An analysis of the optical rotatory dispersion of polypeptides and proteins, *Proc Nat Acad Sci USA* **51** (1964), 695–702.
- [15] G. Szivessy and C. Münster, Über die Prüfung der Gitteroptik bei aktiven Kristallen, *Ann Phys Lpz* **20** (1934), 703–736.
- [16] A.K. Konstantinova, N.R. Ivanov and B.N. Grechushnikov, Optical activity of crystals in directions different from that of the optic axis. I. Uniaxial crystals, *Sov Phys Crystallogr* **14** (1969), 222–229.
- [17] J. Kobayashi, T. Takahashi, T. Hosokawa and Y. Uesu, A new method for measuring the optical activity of crystals and the optical activity of KH_2PO_4 , *J Appl Phys* **49** (1978), 809–815.

- [18] J. Kobayashi and Y. Uesu, A new optical method and apparatus “HAUP” for measuring simultaneously optical activity and birefringence of crystals. I. Principles and construction, *J Appl Cryst* **19** (1983), 204–211.
- [19] J. Kobayashi, Y. Uesu, J. Ogawa and Y. Nishihara, Optical activity and x-ray studies on incommensurate phase transitions of ferroelectric ammonium fluoroberyllate $(\text{NH}_4)_2\text{BeF}_4$, *Phys Rev B* **31** (1985), 4569–4575.
- [20] K. Saito, H. Sugiya and J. Kobayashi, Optical activity of incommensurate state of $[\text{N}(\text{CH}_3)_4]_2\text{CuCl}_4$, *J Appl Phys* **6** (1990), 732–735.
- [21] J. Kobayashi, Incommensurate phase transitions and optical activity, *Phys Rev B* **4** (1990), 8332–8338.
- [22] K. Saito and J. Kobayashi, Optical study on incommensurate phase transitions of $(\text{C}_3\text{H}_7\text{NH}_3)_2\text{MnCl}_4$, *Phys Rev B* **45** (1992), 10264–10270.
- [23] J. Kobayashi, K. Saito, N. Takahashi and I. Kamiya, Birefringence and optical activity of the incommensurate phase of $[\text{N}(\text{CH}_3)_4]_2\text{ZnCl}_4$, *Phys Rev B* **48** (1993), 10039–10046.
- [24] J. Kobayashi, K. Saito, N. Takahashi and I. Kamiya, Optical activity and birefringence of the incommensurate phase of Rb_2ZnBr_4 , *Phys Rev B* **9** (1994), 6539–6547.
- [25] J. Kobayashi, K. Saito, N. Takahashi, I. Kamiya and H. Utsumi, Optical activity and birefringence of the incommensurate phase of Rb_2ZnCl_4 , *Phys Rev B* **50** (1994), 2766–2774.
- [26] J. Kobayashi, Y. Uesu and H. Takehara, A new optical method and apparatus “HAUP” for measuring simultaneously optical activity and birefringence of crystals. II. Application to triglycine-sulphuric acid $(\text{NH}_2\text{CH}_2\text{CO}_2\text{H})_3 \cdot \text{H}_2\text{SO}_4$, *J Appl Cryst* **16** (1983), 212–219.
- [27] J. Kobayashi, K. Uchino and T. Asahi, Optical properties of Rochelle salt, *Phys Rev B* **43** (1991), 5706–5712.
- [28] T. Asahi, M. Tomizawa and J. Kobayashi, Optical properties of BaMnF_4 , *Phys Rev B* **45** (1992), 1971–1987.
- [29] J. Kobayashi, K. Uchino, H. Matsuyama and K. Saito, Optical study on domain switching of ferroelectric triglycine sulfuric acid and Rochelle salt, *J Appl Phys* **69** (1991), 409–413.
- [30] N. Takahashi, M. Tomizawa and J. Kobayashi, Optical activity and electrogyration effect of $\text{Co}_3\text{B}_7\text{O}_{13}\text{I}$ and $\text{Cu}_3\text{B}_7\text{O}_{13}\text{Cl}$, *Jap J Appl Phys* **31** (1992), 3209–3211.
- [31] T. Asahi, R. Ikeda, M. Nakamura, T. Morikawa, M. Higano, H. Suzuki and J. Kobayashi, Optical and x-ray studies on phase transition of $(\text{NH}_4)_2\text{SO}_4$, *Ferroelectrics Lett* **21** (1996), 47–53.
- [32] J. Kobayashi, T. Asahi, M. Sakurai, M. Takahashi, K. Okubo and Y. Enomoto, Optical properties of superconducting $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_8$, *Phys Rev B* **53** (1996), 11784–11795.
- [33] J. Kobayashi, T. Asahi, M. Ichiki, K. Saito, T. Shimasaki, H. Yoshii, Y. Itagaki and H. Ikawa, Optical study on the phase transition of lead lanthanum zirconate titanate $\text{Pb}_{1-x}\text{La}_x(\text{Zr}_y\text{Ti}_{1-x/4}\text{O}_3$, *Phys Rev B* **51** (1995), 763–778.

- [34] J. Kobayashi, T. Asahi, M. Ichiki, A. Oikawa, M. Suzuki, T. Watanabe, E. Fukada and Y. Shikinami, Structural and optical properties of poly lactic acids, *J Appl Phys* **77** (1995), 2957–2973.
- [35] T. Asahi, M. Nakamura, F. Toda, H. Miyamoto and J. Kobayashi, Optical activity of oxo amide crystal, *J Am Chem Soc* **119** (1997), 3663–3669.
- [36] T. Asahi, H. Utsumi, Y. Itagaki, I. Kagomiya and J. Kobayashi, Optical activity of crystalline glutamic acids, *Acta Cryst A* **52** (1996), 766–769.
- [37] T. Asahi, M. Takahashi and J. Kobayashi, Optical activity of crystalline L-aspartic acids, *Acta Cryst A* **53** (1994), 763–771.
- [38] J. Kobayashi, T. Asahi, M. Sakurai, I. Kagomiya, H. Asai and H. Asami, The optical activity of lysozyme crystals, *Acta Cryst A* **54** (1998), 581–590.
- [39] T. Asahi, T. Osaka, J. Kobayashi, S.C. Abrahams, S. Nanamatsu and M. Kimura, Optical study on a phase transition of fresnoite $\text{Ba}_2\text{Si}_2\text{TiO}_8$, *Phys Rev B* **63** (2001), 94104/1–94104/13.
- [40] J. Kobayashi, H. Kumomi and K. Saito, Improvement of the accuracy of HAUP, high-accuracy universal polarimeter: Application to ferroelectric $[\text{N}(\text{CH}_3)_4]_2\text{ZnCl}_4$, *J Appl Cryst* **19** (1986), 377–381.
- [41] J. Kobayashi, T. Asahi, S. Takahashi and A.M. Glazer, Evaluation of the systematic errors of polarimetric measurements: Application to measurements of tensors of α -quartz by the HAUP, *J Appl Cryst* **21** (1988), 479–484.
- [42] J.R.L. Moxon, A.R. Renshaw and I.J. Tebbutt, The simultaneous measurement of optical activity and circular dichroism in birefringent linearly dichroic crystal sections. II. Description of apparatus and results for quartz, nickel sulphate hexahydrate and benzil, *J Phys D: Appl Phys* **24** (1991), 1187–1192.
- [43] M. Kremers and H. Meekes, Application of the high-accuracy universal polarimeter to magnetic and absorbing crystals, *J Phys D: Appl Phys* **28** (1995), 1212–1224.
- [44] T. Asahi, T. Osaka and J. Kobayashi, Polarimetry of electromagnetic materials, *Complex mediums II: Beyond linear isotropic dielectrics* (A. Lakhtakia, W. Weiglhofer and I. Hodgkinson, eds), Proc SPIE **4467**, Bellingham, WA, USA, 2001, 20–30.
- [45] H. Meekes and A. Janner, Optical activity in the incommensurate structure Rb_2ZnCl_4 , *Phys Rev B* **38** (1988), 8075–8087.
- [46] J.R.L. Moxon and A.R. Renshaw, The simultaneous measurement of optical activity and circular dichroism in birefringent linearly dichroic crystal sections. I. Introduction and description of the method, *J Phys: Condens Matter* **2** (1990), 6807–6836.
- [47] E. Dijkstra, H. Meekes and M. Kremers, The high-accuracy universal polarimeter, *J Phys D: Appl Phys* **24** (1991), 1861–1868.
- [48] J. Ortega, J. Etxebarria, J. Zubillaga, T. Breczewski and M.J. Tello, Lack of optical activity in the incommensurate phases of Rb_2ZnBr_4 and $(\text{N}(\text{CH}_3)_4)_2\text{CuCl}_4$, *Phys Rev B* **45** (1992), 5155–5162.

- [49] J. Simon, J. Weber and H.-G. Unruh, Some new aspects about the elimination of systematical errors in HAUP measurements, *J Phys D: Appl Phys* **30** (1997), 676–682.
- [50] C. Hernandez and P. Gomez, Optical anisotropy of quartz in the presence of temperature-dependent multiple reflections using a high-accuracy universal polarimeter, *J Phys D: Appl Phys* **33** (2000), 2985–2994.
- [51] Y.I. Shopa and M.O. Kravchuk, Study of optical activity in $\text{La}_3\text{Ga}_5\text{SiO}_{14}$ with high-accuracy polarimetric methods, *Phys Stat Sol (a)* **158** (1996), 275–280.
- [52] O. Vlokh, Y. Shopa and M. Kravchuk, Measurements of the optical activity in $\text{Gd}_2(\text{MoO}_4)_3$ using polarimetric technique related to HAUP, *Ferroelectrics* **203** (1997), 107–111.
- [53] W. Kaminsky, Topographies of chiral and associated optical properties in FeBo_3 using a novel polarimeter, the ‘Tilter’, *Ferroelectrics* **204** (1997), 233–246.
- [54] A. Yariv and P. Yeh, *Optical waves in crystals*, Wiley, New York, NY, USA, 1983.
- [55] A. Lakhtakia (ed), *Selected papers on natural optical activity*, SPIE Press, Bellingham, WA, USA, 1990.
- [56] V.M. Agranovich and V.L. Ginzburg, *Crystal optics with spatial dispersion, and excitons*, Springer, Berlin, Germany, 1984.
- [57] L.D. Landau, E.M. Lifshitz and L.P. Pitaevskii, *Electrodynamics of continuous media*, Pergamon Press, New York, NY, USA, 1984.
- [58] T.M. Lowry, *Optical rotatory power*, Dover Press, New York, NY, USA, 1964.
- [59] S.F. Mason (ed), *Optical activity and chiral discrimination*, D. Reidel, Dordrecht, The Netherlands, 1978.
- [60] M. Born, *Optik*, Springer, Berlin, Germany, 1933.
- [61] A.V. Shubnikov, *Principles of optical crystallography*, Consultants Bureau, New York, NY, USA, 1960.
- [62] D.J. Caldwell and H. Eyring, *The theory of optical activity*, Wiley, New York, NY, USA, 1971.
- [63] G.N. Ramachandran and S. Ramaseshan, Crystal optics, *Encyclopedia of physics*, Springer, Berlin, Germany, 1961.
- [64] J.F. Nye, *Physical properties of crystals*, Clarendon Press, Oxford, UK, 1957.
- [65] Y.R. Shen, *The principle of nonlinear optics*, Wiley, New York, NY, USA, 1984.
- [66] L.D. Barron, *Molecular light scattering and optical activity*, Cambridge University Press, Cambridge, UK, 1982.
- [67] R. Azzam and N. Bashara, *Ellipsometry and polarized light*, North-Holland, Amsterdam, The Netherlands, 1977.
- [68] R.C. Jones, A new calculus for the treatment of optical systems. VII. Properties of the N-matrices, *J Opt Soc Am* **38** (1948), 671–685.
- [69] K. Wüthrich, *NMR of proteins and nucleic acids*, Wiley, New York, NY, USA, 1986.

- [70] L.J. Smith, M.J. Sutcliffe, C. Redfield and C.M. Dobson, Structure of hen lysozyme in solution, *J Mol Biol* **229** (1993), 930–944.
- [71] C.C.F. Blake, R.H. Fenn, A.C.T. North, D.C. Phillips and R.J. Poljak, Structure of lysozyme, *Nature* **196** (1962), 1173–1176.
- [72] L. Pauling and R.B. Corey, Atomic coordinates and structure factors for two helical configurations of polypeptide chains, *Proc Nat Acad Sci USA* **37** (1951), 235–240.
- [73] G. Alderton, W.H. Ward and H.L. Fevold, Isolation of lysozyme from egg white, *J Biol Chem* **157** (1945), 43–58.
- [74] F.T. Jones, Optical and crystallographic properties of lysozyme chloride, *J Am Chem Soc* **68** (1946), 854–857.
- [75] G.N. Ramachandran, Theory of optical activity of crystals, *Proc Ind Acad Sci* **33** (1951), 217–308.
- [76] W. Moffitt, Optical rotatory dispersion of helical polymers, *J Chem Phys* **25** (1956), 467–477.
- [77] W. Kauzmann and H. Eyring, The effect of the rotation of groups about bonds on optical rotatory power, *J Chem Phys* **9** (1941), 41–53.
- [78] K. Hamaguchi and H. Sakai, Structure of lysozyme, *J Biochem* **57** (1965), 721–732.
- [79] M.V. Hobden, Optical activity in a non-enantiomorphous crystal silver gallium sulphide, *Nature* **216** (1953), 678.
- [80] J. Etxebarria, C.L. Flocia and J. Ortega, Origin of the optical activity of silver thiogallate, *J Appl Cryst* **33** (2000), 126–129.
- [81] J. Kobayashi, Invitation to chiral physics, *Ferroelectrics* **203** (1997), 27–56.



Toru Asahi is a visiting associate professor of Nano-Science and Nano-Engineering, Waseda University, Japan. He graduated from the Department of Applied Physics of Waseda University in 1988, and received a Doctor Degree of Science in 1992. He was a Research Associate in the Department of Applied Physics of Waseda University (1990–1993) and a Post-Doctoral Fellow of the Japan Society for the Promotion of Science (1993–1995). He was a Research Associate (1995–1996), a visiting associate professor (1996–2002) of Kagami Memo-

rial Laboratory for Materials Science and Technology of Waseda University, and a visiting associate professor for the Advanced Research Institute for Science and Engineering of Waseda University (2002–2003). He is a Planning Committee member of the Magnetic Society of Japan (1999–2003) and a member of the Nano-Technology Forum of Waseda University (1998–). He is a Program Committee member of the annual SPIE conference on complex mediums (2001–). His current research interests are chiral properties, in particular, optical activity and

circular dichroism in materials, provided by the high-accuracy universal polarimeter (HAUP), and nanostructured magnetic thin films for high-density perpendicular magnetic recording.



Jinzo Kobayashi is a Professor Emeritus of Waseda University, Japan. He graduated from the Department of Applied Chemistry of Waseda University in 1951, and received a Doctor Degree of Engineering in 1960. He was an associate professor (1959–1968) and a professor in the Department of Applied Physics of Waseda University (1968–1995). He was a visiting staff member of Centre National de la Recherche Scientifique in Paris (1966), a guest scientist of Battelle Memorial Institute in Geneva (1967, 1971), a visiting professor of Moscow State University (1981) and Oxford University (1987), and a counselor of United Nations' Industrial Development (1988).

He was a lecturer of Fukuoka University (1993) and a counselor of Japan Spectroscopic Company in Tokyo. He has been an Advisory Committee member for the International Meeting on Ferroelectricity since 1977. He established the Japan–Soviet Symposium on Ferroelectricity in 1976 and was Chairman of the Organizing Committee of the symposium (1976, 1980, and 1984). He is an Editorial Board Member of *Ferroelectrics*, *Ferroelectrics Letters*, *Ferroelectrics Review*, *Ferroelectrics Communication*, and *Condensed Matter News*. He was awarded the Prize of Ministry of Science and Technology (1996), and decorated by the Emperor Heisei for the Japanese government (1999). His research field is solid state physics. He invented the High-Accuracy Universal Polarimeter (HAUP) in 1983, and with the HAUP he has been studying the optical activity and circular dichroism in condensed matter.

Generalized Ellipsometry

Mathias Schubert

Abstract

Generalized ellipsometry determines the complex electromagnetic planewave response of general optical systems in terms of Jones or Mueller matrix elements, regardless of sample anisotropy. It also enables characterization of the optical and structural properties of general anisotropic systems, including absorption, and can be applied, for instance, to determine the linear dielectric response tensors including frame-of-reference orientations of birefringent materials in layered samples for wavelengths from the far infrared to the deep ultraviolet. Generalized ellipsometry permits exploration of major indexes of refraction and extinction coefficients of low-symmetry materials as functions of wavelength, providing insight into physical phenomena in layered anisotropic mediums. Precise structural and optical data of novel compound materials can be obtained. A brief presentation of the generalized ellipsometry approach using Jones and Mueller matrixes is given, and data acquisition and analysis are discussed. Examples illustrate a broad range of applicability. Determination of optical constants for orthorhombic absorbing bulk materials is demonstrated. Measurement and assignment of phonon modes in wurtzite-type films on low-symmetry substrates is discussed. Fundamental band-to-band transitions in spontaneously ordered semiconductor compounds, geometry and optical constants of chiral liquid crystal cells, and magneto-optic birefringence in free-carrier plasmas are obtained from analysis of generalized ellipsometry data and subsequent model analysis.

1 Introduction

Thousands of publications have reported on and interpreted the optical constants (refractive index n and extinction coefficient k) that describe the interaction of electromagnetic radiation with optically isotropic (amorphous and cubically symmetric) or uniaxial (hexagonal and tetragonal) materials. There are few, if any, complete and accurate sets of optical constants for biaxial (orthorhombic, monoclinic and triclinic) materials measured over a range of wavelengths. This situation has arisen because (i) most attention has been focused on simple, high-symmetry materials, (ii) existing methods of measuring the larger number of optical constants of low-symmetry materials were much more difficult, requiring several different, precisely oriented samples, and (iii) no accurate rigorous method existed so far for monoclinic or triclinic absorbing materials. The situation is even more difficult for thin films, because (iv) traditional techniques lack sufficient sensitivity to optical constants and their anisotropy. Generalized ellipsometry, based upon the standard ellipsometry approach invented by Drude more than a century ago [1], provides a global approach to fully characterize such general anisotropic samples including situations of multiple-layer systems; and it is presented here as a novel technique. In principle, generalized ellipsometry should be capable of handling anisotropy beyond the simple dielectric response, i.e., for situations when the constitutive relations require more than only the linear dielectric response (permittivity) tensor.

The purpose of this chapter is a comprehensive review of the generalized ellipsometry approach. Theory and experiment are outlined briefly in Sections 3, 4, and 5. Measurements of Jones matrix or Mueller matrix elements [2]–[6], for wavelengths from the infrared to the ultraviolet spectral region [7], provide access to the complex response tensor element spectrum and sample geometry information upon best-fit comparison with model calculations, explained in Section 6. After establishing experimental tools for measurement of the generalized ellipsometry parameters, described in detail previously [3]–[11], the challenge for the experimenter consists of the ellipsometry data model analysis using numerical approaches of high complexity, thus addressing adequately the optical response due to the intricate sample anisotropy. Section 7.1 presents full analysis of the three major complex optical constants of a biaxial absorbing mineral, i.e., a material with full symmetric permittivity tensor properties, which permits us to test the validity of numerical solutions for light propagation in arbitrarily (dielectric) anisotropic materials [12]. The extension of the generalized ellipsometry approach to long wavelengths, and determination of crystal axes and phonon mode parameters within anisotropic substrate/film systems is demonstrated in Section 7.2. Measurement of subtle anisotropy in the near-infrared–visible spectral region due to spontaneous atomic ordering in thin films of zincblende-type semiconductor III–V alloys with reduced lattice symmetry is discussed in Section 7.3. There, valuable information on electronic band structure parameters of ordered III–V compounds is derived from the generalized ellipsometry experiment. The anisotropic refractive

indexes of the uniaxial molecules within chiral liquid crystal cells are determined from analysis of temperature-dependent generalized ellipsometry measurements, which is explained in Section 7.4. Measurement and analysis of the full nonsymmetric magneto-optic permittivity tensor of the free-carrier plasma in GaAs at infrared wavelengths is finally discussed in Section 7.5. There, the generalized ellipsometry approach is demonstrated as an optical scale for the “inertial” mass of free carriers in semiconductor layer structures. Finally, conclusions are given in Section 8.

2 Experimental

Ellipsometer instruments capable of measuring the generalized ellipsometry parameters can be built in-house or are commercially available as spectroscopic equipment operating from far-infrared (free-space wavelength $\lambda_0 \sim 100 \mu\text{m}$) to the deep ultraviolet ($\lambda_0 \sim 0.15 \mu\text{m}$) spectral regions. Details shall not be given here, but may be accessed from online sources [13].

3 Birefringence in stratified mediums

The frequency-domain electromagnetic constitutive relations in Cartesian coordinates can be written as

$$\mathbf{D}(\mathbf{r}, \omega) = \varepsilon_0 [\boldsymbol{\varepsilon}(\mathbf{r}, \omega) \bullet \mathbf{E}(\mathbf{r}, \omega) + \boldsymbol{\xi}(\mathbf{r}, \omega) \bullet \mathbf{H}(\mathbf{r}, \omega)], \quad (1)$$

$$\mathbf{B}(\mathbf{r}, \omega) = \mu_0 [\boldsymbol{\mu}(\mathbf{r}, \omega) \bullet \mathbf{H}(\mathbf{r}, \omega) + \boldsymbol{\zeta}(\mathbf{r}, \omega) \bullet \mathbf{E}(\mathbf{r}, \omega)], \quad (2)$$

where ε_0 and μ_0 are the vacuum permittivity and the vacuum permeability, and $\boldsymbol{\varepsilon}$ and $\boldsymbol{\mu}$ are the material’s relative permittivity and relative permeability tensors, respectively. The cross-coupling tensors $\boldsymbol{\xi}$ and $\boldsymbol{\zeta}$ are the two magnetoelectric tensors. (For definition of the constitutive relations for several anisotropic materials see the textbooks by Kong [15], Lakhtakia [16-17] or Landau and Lifshitz [18].) The optical response of a material is said to be anisotropic, in general, if at least one of the tensors $\boldsymbol{\varepsilon}$ and $\boldsymbol{\mu}$ possesses at least two, possibly complex-valued, diagonal (major) elements, which are different from each other in their real and/or imaginary parts, whereas $\boldsymbol{\xi}$ and $\boldsymbol{\zeta}$ can be zero. A medium wherein all four tensors have complex, nonsymmetric, and nonvanishing elements is referred to as bianisotropic; see the chapter of Weiglhofer [19]. For homogeneous mediums the four tensors do not depend on the spatial position $\mathbf{r}$. A special type of a nonhomogeneous medium is that of the so-called helicoidal bianisotropic medium (HBM), where the four tensors possess rotational nonhomogeneity along the z axis [20]-[24]. Chiral liquid crystals are HBMs with symmetric dielectric properties. In this chapter, materials are considered for which the magnetoelectric tensors vanish, and where the relative permeability tensor is set equal to identity. The remaining tensor is the permittivity (dielectric function) tensor $\boldsymbol{\varepsilon}$, which can describe isotropic, uniaxial, or biaxial di-

Table 1 Properties of the relative permittivity tensor ϵ and optical symmetries of symmetrically dielectric mediums. In the general (triclinic) case, ϵ is composed of three principal (major) refractive indexes n_j , three Euler angles relating principal axes of refraction to crystal axes, three extinction coefficients k_j , and three Euler angles relating principal axes of absorption to crystal axes, for a total of 12 real-valued parameters. Increasing symmetry to monoclinic or orthorhombic cases reduces this to 8 and 6 constants, respectively. Three additional parameters (Euler angles ψ, φ, θ) are required to describe the orientation of the crystal axes relative to the laboratory axes for a sample in general orientation, irrespective of its symmetry. See also Ref. 25, and Figure 1.

Optical symmetry	Crystal system	Permittivity tensor
isotropic	cubic	$\epsilon_x = \epsilon_y = \epsilon_z$
uniaxial	tetragonal	$\epsilon_x = \epsilon_y \neq \epsilon_z$
	hexagonal	
	trigonal	
biaxial	triclinic	$\epsilon_x \neq \epsilon_y \neq \epsilon_z$
	monoclinic	
	orthorhombic	

electric material properties; see Table 1. In fact, the optical response of most solids and liquids can be well understood in terms of the complex-valued diagonal permittivity tensor elements $\epsilon_{x'}$, $\epsilon_{y'}$ and $\epsilon_{z'}$ of ϵ and a rotation $\mathbf{A}$, as shown in Figure 1. The unitary matrix $\mathbf{A}$ relates major tensor axes (x' , y' , z') to the laboratory frame of reference (x , y , z) as

$$\epsilon = \mathbf{A} \bullet \text{diag}(\epsilon_{x'}, \epsilon_{y'}, \epsilon_{z'}) \bullet \mathbf{A}^{-1}. \quad (3)$$

For monoclinic and triclinic systems, $\mathbf{A}$ depends on the photon energy. Magneto-optical effects in solids cause antisymmetric dielectric properties, and ϵ becomes nonsymmetric in general.

A nonsymmetric tensor decomposes into its symmetric and antisymmetric part, and both can have three major complex-valued elements and different coordinate systems. In general, the relative permittivity tensor is a function of the photon energy $\hbar\omega$, which is known as frequency dispersion. For mediums with broken mirror symmetry, the so-called optically active mediums, ϵ depends in addition to $\hbar\omega$ on the photon wavevector $\hbar\mathbf{k}$. This effect, also called spatial dispersion, and measurement of related phenomenons is addressed in detail within the chapter of Asahi and Kobayashi in this volume [26]; see also chapters of Litvin, Dewar, and Mackay [27]–[29].

4 Generalized ellipsometry

Ellipsometry, generally speaking, determines the complex-valued ratio ρ of polarized planewave components. Traditionally this ratio is measured in reflection or transmission for light polarized parallel (p) and perpendicular (s) to the plane of incidence, which is shown in Figure 2 [2]. The result of an ellipsometry measurement is usually presented by real-valued parameters Ψ and Δ , where $\tan\Psi$ is

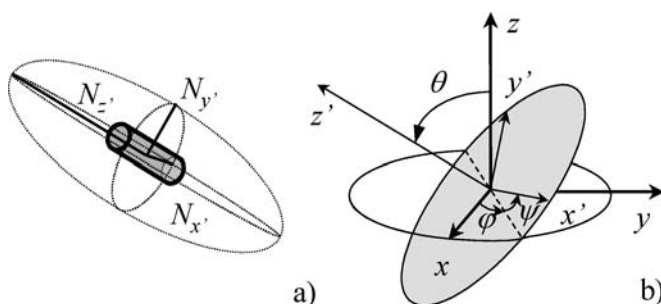


Figure 1 (a) a schematic presentation of the index-of-refraction ellipsoid of a biaxial material. The major complex optical constants N_i are defined as $N_j \equiv \sqrt{\epsilon_j} = n_j + \sqrt{-1}k_j$, $j = "x", "y", "z"$. n_j and k_j address the major indexes of refraction and the extinction coefficients, respectively, within a Cartesian center-of-gravity system. This graphical picture holds only if all extinction coefficients vanish ($k_{x'} = k_{y'} = k_{z'} = 0$). (b) the Euler angles ϕ , ψ and θ , and the orthogonal rotations provided by rotation matrix A . (See Refs. 12 and 14; (x', y', z') and (x, y, z) refer to the material's major axes and the laboratory systems, respectively.)

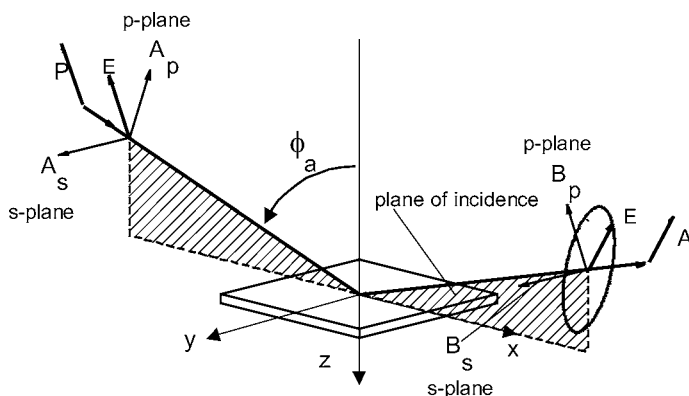


Figure 2 Plane of incidence and angle of incidence ϕ_a . A_p , A_s , B_p and B_s denote the complex amplitudes of the p and s modes before and after reflection, respectively. P and A are the azimuth angles of the linear polarizer used, e.g., in the standard arrangement of rotating-analyzer (polarizer) ellipsometer [4]-[6].

defined as the absolute value of the complex ratio, and Δ denotes the relative phase change of the p and s components of the electric field vector upon interaction with a sample [2].

4.1 Jones matrix presentation

For nondepolarizing samples, the so-called Jones matrix, containing the complex-valued polarized reflection (transmission) coefficients, provides a complete description for ellipsometric analysis [2], [4]-[6], [30]. In Figure 3, a set of multiple beams, characterized by plane-parallel wavefronts, may emerge from a nondepo-

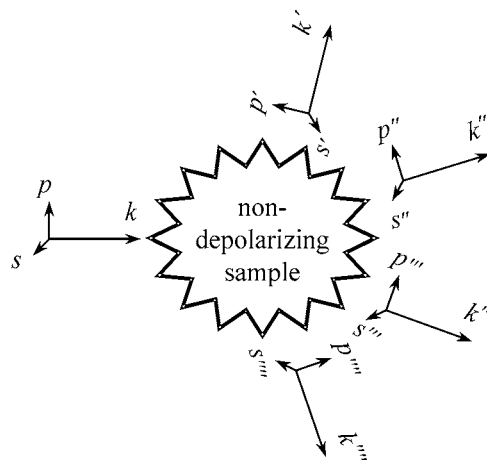


Figure 3 Incident and emergent electromagnetic plane waves at a nondepolarizing optical system. Jones matrixes connect the incident mode (p, s) with emergent plane wave components $(p, s)', (p, s)'', \dots$ with respect to a chosen coordinate system, e.g., the p and s planes in Figure 2.

larizing sample. Such a situation may occur for a grating, for example. If a sample with plane-parallel boundaries is considered, a reflected and a transmitted beam may emerge, characterized by the Jones reflection and transmission matrixes $\mathbf{r}$ and $\mathbf{t}$, respectively:

$$\mathbf{r} = \begin{pmatrix} r_{pp} & r_{sp} \\ r_{ps} & r_{ss} \end{pmatrix}, \quad \mathbf{t} = \begin{pmatrix} t_{pp} & t_{sp} \\ t_{ps} & t_{ss} \end{pmatrix}. \quad (4)$$

The off-diagonal elements are nonzero for optical systems that convert p into s waves, and vice versa. In standard ellipsometry, the off-diagonal elements vanish, and Ψ and Δ do not depend on the polarization state of the incident plane wave, while

$$\rho = \frac{j_p}{j_s} = \tan \Psi \exp(\sqrt{-1} \Delta). \quad (5)$$

Here, j_p and j_s denote the p - and s -polarized complex reflection or transmission coefficients. In generalized ellipsometry, the last equation depends on the polarization state of the incident plane wave [12]. In general, multiple, i.e., spatially separated, beams may emerge from such a system (e.g., a diffraction grating), as indicated in Figure 3. For each beam, the generalized ellipsometry parameters are described by three *complex-valued* ratios of the four elements of $\mathbf{r}$ and/or $\mathbf{t}$ [2-3], [4]-[6], [30]. These ratios define the generalized ellipsometry parameters Ψ_{ij} and

Δ_{ij} in transmission or reflection arrangement:

$$\begin{aligned}\frac{j_{pp}}{j_{ss}} &\equiv J_{pp} = \tan \Psi_{pp} \exp \left(\sqrt{-1} \Delta_{pp} \right), \\ \frac{j_{ps}}{j_{pp}} &\equiv J_{ps} = \tan \Psi_{ps} \exp \left(\sqrt{-1} \Delta_{ps} \right), \\ \frac{j_{sp}}{j_{ss}} &\equiv J_{sp} = \tan \Psi_{sp} \exp \left(\sqrt{-1} \Delta_{sp} \right),\end{aligned}\quad (6)$$

where capitalized symbols are ratios of Jones matrix elements [3], [4]-[6]. This set comprises six *real-valued* quantities out of the eight possible values contained within the Jones matrix—lacking the light beam's absolute intensity and the light beam's absolute phase information. Different techniques exist for acquisition of generalized ellipsometry parameters from anisotropic samples, and the reader is referred to Refs. 2-3, 8-11.

4.2 Mueller matrix presentation

An alternative description of the polarized sample response, which also fully accounts for depolarization (not easily implemented within the Jones matrix formalism), is the formalism of the Mueller matrix and the Stokes vector. In terms of the *p*- and *s*-polarization coordinate system of the electromagnetic plane wave, the four *real-valued* Stokes parameters, having dimensions of *intensities*, can be expressed as

$$\begin{aligned}S_0 &= I_p + I_s, \\ S_1 &= I_p - I_s, \\ S_2 &= I_{45} - I_{-45}, \\ S_3 &= I_{\sigma+} - I_{\sigma-},\end{aligned}\quad (7)$$

where I_p , I_s , I_{45} , I_{-45} , $I_{\sigma+}$ and $I_{\sigma-}$ denote the intensities for the *p*-, *s*-, $+45^\circ$, -45° , right and left circularly polarized (RCP and LCP) light components, respectively [2], [31]. Arranging these elements into a column vector, the Mueller matrix then describes the changes of each quantity upon interaction of the electromagnetic plane wave with any optical system (sample, mirrors, rotators, other optical devices and any combinations thereof). One has

$$\begin{bmatrix} S_0 \\ S_1 \\ S_2 \\ S_3 \end{bmatrix}_{\text{output}} = \begin{pmatrix} M_{11} & M_{12} & M_{13} & M_{14} \\ M_{21} & M_{22} & M_{23} & M_{24} \\ M_{31} & M_{32} & M_{33} & M_{34} \\ M_{41} & M_{42} & M_{43} & M_{44} \end{pmatrix} \begin{bmatrix} S_0 \\ S_1 \\ S_2 \\ S_3 \end{bmatrix}_{\text{input}}. \quad (8)$$

The advantage of this concept is the ability to handle situations with partial depolarization of the electromagnetic plane wave, particularly when caused by device

components' imperfections. Details of this concept, its use in ellipsometry, and its relation to the Jones concept have been outlined previously [2], [31]. For a nondepolarizing system [2], a one-to-one relation exists between the matrixes $\mathbf{r}$ and $\mathbf{M}$, and one can write $\mathbf{M}$ as a function of elements of $\mathbf{r}$ as follows:

$$\begin{bmatrix} \frac{1}{2}(|r_{pp}|^2 + |r_{ss}|^2 + |r_{sp}|^2 + |r_{ps}|^2) & \frac{1}{2}(|r_{pp}|^2 - |r_{ss}|^2 - |r_{sp}|^2 + |r_{ps}|^2) & \operatorname{Re}(r_{pp}\bar{r}_{sp} + \bar{r}_{ss}r_{ps}) & \operatorname{Im}(r_{pp}\bar{r}_{sp} + \bar{r}_{ss}r_{ps}) \\ \frac{1}{2}(|r_{pp}|^2 - |r_{ss}|^2 + |r_{sp}|^2 - |r_{ps}|^2) & \frac{1}{2}(|r_{pp}|^2 + |r_{ss}|^2 - |r_{sp}|^2 - |r_{ps}|^2) & \operatorname{Re}(r_{pp}\bar{r}_{sp} - \bar{r}_{ss}r_{ps}) & \operatorname{Im}(r_{pp}\bar{r}_{sp} - \bar{r}_{ss}r_{ps}) \\ \operatorname{Re}(r_{pp}\bar{r}_{ps} + \bar{r}_{ss}r_{sp}) & \operatorname{Re}(r_{pp}\bar{r}_{ps} - \bar{r}_{ss}r_{sp}) & \operatorname{Re}(r_{pp}\bar{r}_{ss} + \bar{r}_{ps}r_{sp}) & \operatorname{Im}(r_{pp}\bar{r}_{ss} + \bar{r}_{ps}r_{sp}) \\ -\operatorname{Im}(r_{pp}\bar{r}_{ps} + \bar{r}_{ss}r_{sp}) & -\operatorname{Im}(r_{pp}\bar{r}_{ps} - \bar{r}_{ss}r_{sp}) & -\operatorname{Im}(r_{pp}\bar{r}_{ss} + \bar{r}_{ps}r_{sp}) & \operatorname{Re}(r_{pp}\bar{r}_{ss} - \bar{r}_{ps}r_{sp}) \end{bmatrix}, \quad (9)$$

where the vertical bar indicates a complex conjugate. For example, in a rotating-analyzer system the Mueller matrix elements of row 4 and column 4 are not measured, which does not impair the accessibility of the normalized Jones matrix elements except for its absolute phase, and which can only be obtained by including compensator(s). Compensators are needed, in general, to access the additional Mueller matrix elements [31].

Both the Jones and the Mueller matrix elements depend on the photon energy $\hbar\omega$; the layer structure of the sample; the diagonal permittivity (dielectric function) tensor values $\varepsilon_{x'}$, $\varepsilon_{y'}$ and $\varepsilon_{z'}$ of each layer; orientation angles ψ , φ and θ of the permittivity tensor of each layer; each layer's thickness d ; the permittivity tensor of the substrate material that supports the layer stack; the ambient material's permittivity; and the angle of incidence Φ_a .

5 Light propagation in layered anisotropic mediums

5.1 Coherent treatment

Complex-valued matrixes with 4×4 elements can be used to calculate the *coherent* electromagnetic response of stratified anisotropic materials in terms of the Jones or the Mueller matrix elements upon use of (9) [32]. Details of the 4×4 formalism have been discussed and outlined in numerous papers. The 4×4 approach is devoted to details of the *internal* sample polarization-modifying process, whereas the (complex-valued) 2×2 Jones (or the real-valued 4×4 Mueller) matrixes then describe the *external* polarizing sample properties [12], [32]. A 4×4 matrix $\mathbf{T}$ connects wave amplitudes across a "coherent" layer sequence. The coherency of a layer stack is given by its total optical thickness, which shall not exceed the coherence length of the incident light beam. Let A_p , A_s , B_p , B_s , C_p , and C_s , denote the complex-valued amplitudes of incident (A), reflected (B), and transmitted (C) eigenpolarizations outside the layer sequence, e.g., in the p and s presentation. Then, $\mathbf{T}$ is defined by

$$(A_s, B_s, A_p, B_p)^T = \mathbf{T} \bullet (C_s, D_s, C_p, D_p)^T, \quad (10)$$

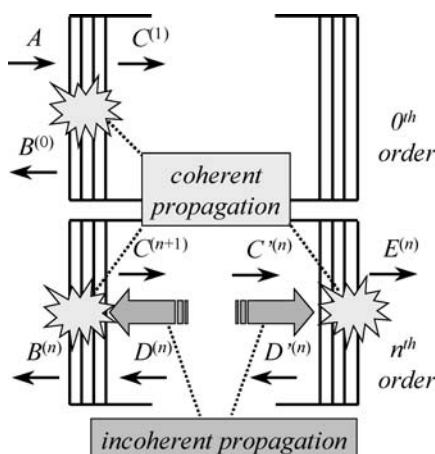


Figure 4 Schematic presentation of incident (A), reflected (B), and transmitted (E) plane waves across a sample with plane-parallel interfaces, and multiple layer stacks at both the front and backside of the supporting substrate. (Indexes for eigenpolarizations, e.g., p , s , are suppressed.) D modes are either reflected off the backside, or caused by modes F (not drawn) incident from the right. If the substrate totally absorbs C , then D and E (F) are meaningless. In general, the wave vectors for (eigen)modes C and D split if the substrate is anisotropic. C and D represent modes just inside the left border of the substrate, while C' and D' represent the modes just inside its right border. Due to finite coherence of real light sources, waves reflected off internally at the substrate's front and backside overlay incoherently when leaving the sample to the right or left. Only $B^{(0)}$ and $E^{(1)}$ emerge completely polarized. $D'^{(n)}$ produce the n -th order front side transmitted modes $B^{(n)}$, as well as the next order of internally reflected beams $C^{(n+1)}$.

where $(\dots)^T$ denotes the transpose of a 4×1 vector. D_s and D_p account for wave amplitudes, which are eventually incident onto the reserve side of the layer sequence (see Figure 4). A characteristic matrix, $\mathbf{T}_p$, accounts for the optical properties of a single homogeneous layer within a stratified sample. Special solutions for $\mathbf{T}_p$ exist for nonmagnetic materials and when

- i) ϵ reduces to a complex-valued scalar, including the case of absorption (isotropic materials) [12];
- ii) ϵ is a complex-valued symmetric anisotropic tensor, including the case of absorption (uniaxial and biaxial materials) [12];
- iii) ϵ is a complex-valued nonsymmetric anisotropic tensor, including the case of absorption (magneto-optic materials) [33];
- iv) ϵ is a complex-valued symmetric anisotropic tensor with rotational twist perpendicular to the layer interfaces, including the case of absorption [8], [14].

Matrixes for ambient ($\mathbf{L}_a$) and exit ($\mathbf{L}_f$) mediums embed the layer stack according to the surrounding materials. Details for matrixes $\mathbf{L}_a$ and $\mathbf{L}_f$ were discussed in Ref. 12. The incident matrix depends on Φ_a and the index of refraction n_a of the ambient material, which is commonly isotropic. If the substrate is

anisotropic, the exit matrix follows from the eigenvectors $\mathcal{E}_i$ of a characteristic coefficient matrix. Explicit expressions of $\mathcal{E}_i$ for nonsymmetric dielectric mediums are given in Ref. 34. If the major axes of a symmetrically dielectric substrate coincide with the laboratory axes (x, y, z), $\mathbf{L}_f$ can be found in Refs. 4-6. The overall transfer matrix $\mathbf{T}$ follows from the ordered product of all n layers' matrixes $\mathbf{T}_p$, pre- and postmultiplied by the incident and exit matrixes, as follows:

$$\mathbf{T} = \mathbf{L}_a^{-1} \mathbf{T}_{p1}^{-1} \dots \mathbf{T}_{pn}^{-1} \mathbf{L}_f. \quad (11)$$

The Jones matrixes $\mathbf{r}$ and $\mathbf{t}$ follow then from $\mathbf{T}$, as described previously. For example, the reflection mode generalized ellipsometry parameters read (“ J_{ij} ” = “ R_{ij} ”, “ ij ” = “ pp ”, “ ps ”, “ sp ”) [35]

$$R_{pp} = \frac{T_{11}T_{43} - T_{41}T_{13}}{T_{21}T_{33} - T_{23}T_{31}}, \quad R_{ps} = \frac{T_{11}T_{12} - T_{21}T_{13}}{T_{11}T_{43} - T_{41}T_{13}}, \quad R_{sp} = \frac{T_{41}T_{33} - T_{43}T_{31}}{T_{21}T_{33} - T_{23}T_{31}}. \quad (12)$$

5.2 Incoherent treatment

Conveniently, accessible light sources have limited coherence lengths. Hence, the situation of *incoherent* wave propagation in anisotropic materials must be considered for the generalized ellipsometry approach. The eigenstates, maintaining internal coherence with negligible spatial separation, must be treated incoherently with respect to other beams on different bounces at the front- or backside of substrates when the optical path length exceeds the coherency length of the incident wave. Such a situation is sketched in Figure 4. Cases where copropagating eigenmodes within an anisotropic substrate separate spatially due to strong anisotropy may require lateral displacement of the polarization state detection unit away from the zero-order reflected or transmitted beam in order to capture the higher-order beams into the detector. Incoherent beam overlay causes depolarization of reflected and transmitted beams. The only valid approach for appropriate measurement is the Mueller matrix formalism. However, Jones matrixes can still be used for mathematical modeling of Mueller matrix elements observed experimentally.

When $\mathbf{T}_+$ and $\mathbf{T}_-$ denote the general transfer matrixes for the “coherent” layer stacks at the front and backside of a sample, respectively, the matrixes $\mathbf{r}_+(\mathbf{t}_+)$ and $\mathbf{r}_-(\mathbf{t}_-)$ are the corresponding Jones reflection (transmission) matrixes, respectively, given by

$$\begin{aligned} \begin{pmatrix} B_p \\ B_s \end{pmatrix} &= \mathbf{r}_+ \begin{pmatrix} A_p \\ A_s \end{pmatrix}, \\ \begin{pmatrix} C_p \\ C_s \end{pmatrix} &= \mathbf{t}_+ \begin{pmatrix} A_p \\ A_s \end{pmatrix}, \\ D_p &= D_s = 0, \end{aligned} \quad (13)$$

$$\begin{aligned}
\begin{pmatrix} D'_p \\ D'_s \end{pmatrix} &= \mathbf{r}_- \begin{pmatrix} C'_p \\ C'_s \end{pmatrix}, \\
\begin{pmatrix} E_p \\ E_s \end{pmatrix} &= \mathbf{t}_- \begin{pmatrix} C'_p \\ C'_s \end{pmatrix}, \\
F_p &= F_s = 0.
\end{aligned} \tag{14}$$

Note that in order to obtain elements of $\mathbf{r}_+(t_+)$ and $\mathbf{r}_-(t_-)$ consider (10) as systems of four linear relations between the modes A , B , C , D , and E , F . Modes $D(A)$ and $(C')F$ are zero when one asks for transmission and reflection coefficients for light incident from left (right) to right (left) in Figure 4. To include the beams that are internally reflected off the backside of the sample, we need the reflection and transmission coefficients for light incident from right to left onto the backside of the top layer stack in Figure 4. These reflection and transmission matrixes are denoted by $\mathbf{r}_{+i}$ and $\mathbf{t}_{+i}$, respectively, and follow from $\mathbf{T}_+$:

$$\begin{aligned}
\begin{pmatrix} B_p \\ B_s \end{pmatrix} &= \mathbf{r}_{+i} \begin{pmatrix} D_p \\ D_s \end{pmatrix}, \\
\begin{pmatrix} C_p \\ C_s \end{pmatrix} &= \mathbf{t}_{+i} \begin{pmatrix} D_p \\ D_s \end{pmatrix}, \\
A_p &= A_s = 0.
\end{aligned} \tag{15}$$

The modes within the substrate change upon traveling over a distance d as per

$$\begin{aligned}
\begin{pmatrix} D_p \\ D_s \end{pmatrix} &= \mathbf{m}_- \begin{pmatrix} D'_p \\ D'_s \end{pmatrix}, \\
\begin{pmatrix} C'_p \\ C'_s \end{pmatrix} &= \mathbf{m}_+ \begin{pmatrix} C_p \\ C_s \end{pmatrix}, \\
\mathbf{m}_{\pm} &= \begin{pmatrix} \exp[\sqrt{-1}k_0 d q_1^{\pm}] & 0 \\ 0 & \exp[\sqrt{-1}k_0 d q_2^{\pm}] \end{pmatrix},
\end{aligned} \tag{16}$$

where k_0 is the free-space wavenumber and $q_i^{\pm}$ are the eigenvalues of the matrix $\mathbf{\Delta}$, defined and discussed in Refs. 4-6 and 12. Matrix $\mathbf{\Delta}$ is characteristic for the medium, its response tensors' orientation with respect to the laboratory system, and the ambient incident wave properties. The zero-order reflected beam amplitudes follow from the front reflection matrix $\mathbf{r}_+$ (modes A are set by the ellipsometer) as

$$B^{(0)} = \mathbf{r}_+ A. \tag{17}$$

The first-order transmitted beam is obtained from

$$E^{(1)} = t_- m_+ t_+ A, \quad (18)$$

and the first-order reflected beam reads

$$B^{(1)} = t_{+i} m_- r_- m_+ t_+ A. \quad (19)$$

Higher-order amplitudes are computed accordingly. For analysis of transmission-type data, it is often sufficient to consider the first-order beam only, because the higher-order reflected beams undergo lossy internal reflections twice per round trip. For complete model description of the experimental data provided by a given ellipsometer system (e.g., a polarizer, sample, rotating-compensator, analyzer system), specific equations must relate measured intensities with the above discussed model system. Such treatment is given in Ref. 4 for the polarizer, rotating-compensator, sample and rotating-analyzer system.

6 Generalized ellipsometry data analysis

In principle, the thickness and the relative permittivity tensor of any layer, regardless of the layer structure, can be determined from a generalized ellipsometry experiment by comparing calculated data to measured data [2], [36]-[38]. Traditionally, point-by-point fits are performed where the dielectric function values of interest are extracted from the experimental data for each wavelength, and independent of all other spectral data points. For this procedure, thickness and dielectric functions of all other sample constituents have to be known. The dielectric function obtained from the point-by-point fit then needs further comparison with model assumptions in order to obtain values of physically relevant parameters such as phonon mode frequencies or band-to-band transition energies. Instead, the fitting of parameterized model dielectric functions to experimental data (simultaneously for all spectral data points) provides a direct connection between measured data and physical parameters of interest. Parametric models further prevent wavelength-by-wavelength measurement noise from becoming part of the extracted dielectric functions. Parametric model assumptions greatly reduce the number of free parameters. The use of parametric models involves a certain risk for subtle spectral features to be subsumed by the lineshape of the model function. Nevertheless, parameterization of ϵ based on a physical model is the best choice for ellipsometry data analysis, especially when the point-by-point fit method is inapplicable.

The standard model for analyzing spectroscopic ellipsometry data is based on a sequence of homogeneous layers with smooth and parallel interfaces. For anisotropic materials, the ellipsometric measurements depend on the principal-axis dielectric functions ϵ_x , ϵ_y and ϵ_z ; the orientation of the crystal principal axes with respect to the plane of incidence; and the polarization state of the incident light

beam [2]-[6], [12]. For a general situation of optical anisotropy and sample orientation, the simple definition of ρ in (5) is no longer a complete description, and the generalized ellipsometry approach is needed.

In addition to Ψ and Δ (the standard ellipsometry situation), or Ψ_{pp} , Δ_{pp} , Ψ_{ps} , Δ_{ps} , and Ψ_{sp} , Δ_{sp} (the generalized ellipsometry situation), their standard deviations σ^Ψ and σ^Δ are measured [39-40]. The Levenberg-Marquardt algorithm is used to fit the model parameters by minimizing the following weighted test function (this is called the maximum likelihood approach) [41]

$$\xi^2 = \frac{1}{2S - K} \sum_{i=1}^S \left[\left(\frac{\Psi_i - \Psi_i^c}{\sigma_i^\Psi} \right)^2 + \left(\frac{\Delta_i - \Delta_i^c}{\sigma_i^\Delta} \right)^2 \right], \quad (20)$$

where S denotes the number of measured data pairs (Ψ_i, Δ_i) , K is the number of real-valued fit parameters, and Ψ_i^c and Δ_i^c are the calculated ellipsometric parameters at photon energy $\hbar\omega_i$ [40], [42]. For the generalized ellipsometry situation, the test function is set up accordingly [3]

$$\begin{aligned} \xi^2 = & \frac{1}{2S - K} \sum_{i=1}^S \left[\left(\frac{\Psi_{pp,i} - \Psi_{pp,i}^c}{\sigma_{pp,i}^\Psi} \right)^2 + \left(\frac{\Psi_{ps,i} - \Psi_{ps,i}^c}{\sigma_{ps,i}^\Psi} \right)^2 \right. \\ & \left. + \left(\frac{\Psi_{sp,i} - \Psi_{sp,i}^c}{\sigma_{sp,i}^\Psi} \right)^2 \right] \\ & + \frac{1}{2S - K} \sum_{i=1}^S \left[\left(\frac{\Delta_{pp,i} - \Delta_{pp,i}^c}{\sigma_{pp,i}^\Delta} \right)^2 + \left(\frac{\Delta_{ps,i} - \Delta_{ps,i}^c}{\sigma_{ps,i}^\Delta} \right)^2 \right. \\ & \left. + \left(\frac{\Delta_{sp,i} - \Delta_{sp,i}^c}{\sigma_{sp,i}^\Delta} \right)^2 \right]. \end{aligned} \quad (21)$$

Issues of parameter accessibility from spectroscopic ellipsometry data have been thoroughly discussed in the literature [39]. The random experimental errors $(\sigma_i^\Psi, \sigma_i^\Delta)$ are propagated appropriately into the error bars on the fit parameters. These error bars represent existing and finite correlation values between the fit parameters. The light propagation within the sample is calculated using standard matrix formalism for anisotropic multilayered systems with plane parallel interfaces [12], [39]. Strategies for best data acquisition and analysis for general anisotropic materials using generalized ellipsometry was presented and discussed in detail in Refs. 3-8, 33, and 43-46, and references therein. Substantial new information on ellipsometry matters will become available in a forthcoming handbook publication [47].

7 A survey of birefringent material applications

7.1 Orthorhombic bulk minerals

The optical constants n_j and k_j are, in general, incompletely or poorly known for low-symmetry materials, even for most of the natural-occurring minerals. Use of polarized intensity measurements to extract n_j with incident polarization directed along a high-symmetry axis j , on surfaces cut appropriately from bulk crystals, potentially yields uncertainties due to misalignment between polarization and crystal axes. The intensity approach fails when the crystal is absorbing. Measurement of the polarization state after Drude [1] (ellipsometry) is valid for absorption, but fails for arbitrary orientations, and also when the symmetry is monoclinic or triclinic. Furthermore, surfaces may undergo reactions, e.g., oxidation, soon after preparation, which can drastically affect the data derived from optical experiments. The generalized ellipsometry technique, as demonstrated for the orthorhombic case here, is able to relatively easily determine all optical constants, including crystal orientation, even from a single sample, if cut appropriately [48]. Numerical removal of surface over-layer effects during the data analysis allows further investigation of possible differences between surface and bulk optical properties. The intensity approach is even more cumbersome when the crystal is absorbing. All difficulties are dispensed with by the generalized ellipsometry technique.

In Figure 5 data are presented at $\lambda_0 = 589$ nm for an orthorhombic mineral, specifically stibnite (Sb_2S_3) acquired from sections cut from a single crystal parallel to (100) (a plane), (010) (b plane) and (001) (c plane). None of the surface, however, is cut perfectly parallel to either plane. The generalized ellipsometry data contain all necessary information to determine the actual surface orientation. The three Euler angle coordinates of the crystal major axes for each sample with respect to the ellipsometer laboratory frame can be obtained, together with the major permittivity tensor elements ε_a , ε_b , and ε_c from the data shown in Figure 5. The coordinate system used here is that defined previously, and shown in Figures 1 and 2. Spectrums Δ_{pp} , Δ_{ps} , and Δ_{sp} are omitted for brevity. The variations in Ψ_{pp} , Ψ_{ps} , and Ψ_{sp} versus φ_0 and Φ_a reflect the biaxial anisotropy of this orthorhombic crystal. The refraction and extinction coefficients for light polarization parallel to the a , b , and c axis, and the crystal surface orientation of each individual sample produce the individual anisotropic sample response, which is mapped out here in terms of the Jones matrix elements by the generalized ellipsometry approach. Note the excellent match between experiment and model data. Analysis of this experiment also provides the actual crystallographic sample surface orientation.

The Euler angles (φ , ψ , θ) in Figure 1 provide rotation between the crystal and the laboratory frames, as defined in Refs. 4-6 and 12. Because the samples were cut from the same crystal, the assumption holds that the major dielectric function values of all three samples are the same. The number of available data largely exceeds the number of values required. The sets of biaxial refraction and extinction coefficients given in Table 2 are obtained from simultaneous analysis of the

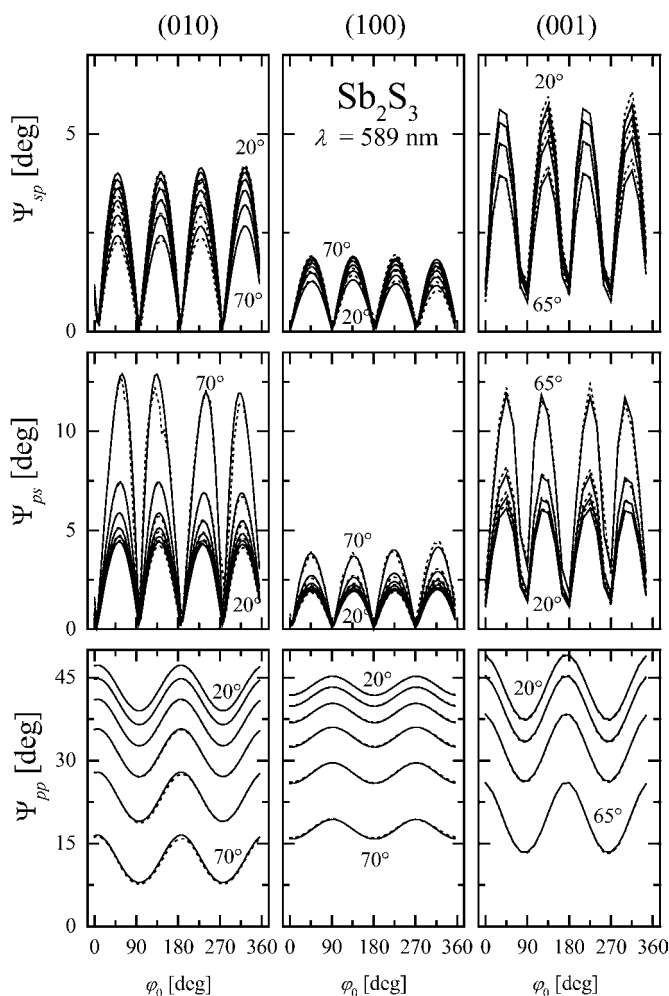


Figure 5 Experimental (broken lines) and best-fit calculated (solid lines) generalized ellipsometry data for stibnite at $\lambda_0 = 589$ nm versus in-plane sample rotation φ_0 and angle of incidence Φ_a .

multiple-azimuth, multiple-angle-of-incidence, and multiple-sample experimental data. Note that for the orthorhombic situation, the Euler angle sets (φ, ψ, θ) are common to both the index of refraction (n_a, n_b, n_c) and the extinction coefficient (k_a, k_b, k_c) ellipsoid for each sample, according to the expected optical properties of orthorhombic crystals. Angles φ , θ and ψ were found as $(89^\circ, 89^\circ, 88^\circ)$ for the (100), $(84^\circ, 93^\circ, 7^\circ)$ for (010), $(78^\circ, -1^\circ, 163^\circ)$ for (001), and $(59^\circ, 51^\circ, 42^\circ)$ for (313), all within $\pm 1^\circ$. The fitted Euler angle sets were found in good overall agreement with x-ray results [48].

It is of interest to compare the optical constants results with historical intensity measurements: For absorbing orthorhombic materials the fraction of incident light intensity, with polarization directed along axis j (this axis must exactly lie in the surface), reflected from a smooth surface at normal incidence is given by

Table 2 Stibnite (Sb_2S_3) index of refraction and extinction coefficient data from generalized ellipsometry measurements at $\lambda_0 = 589$ nm. Values in parentheses represent the 90% confidence interval for the last digits. Comparison of reflectivity calculated from the n and k values obtained here between those tabulated from polarized near-normal incidence intensity measurement after Criddle and Stanley. Previously incomplete data are given in [50].

axis	n_j	k_j	n_j (Ref. [50])	k_j (Ref. [50])	R_{calc}	R_{tab}
a	3.379(9)	0.090(40)	–	–	0.2977	0.3010
b	5.075(11)	0.100(100)	5.03	0.13	0.4506	0.4515
c	4.417(17)	0.270(85)	4.58	0.19	0.3992	0.4045

$R_j = [(n_j - n_a)^2 + k_j^2] / [(n_j + n_a)^2 + k_j^2]$. The two unknowns, n_j and k_j , involved in this expression can be recovered from measurements made in two different immersion mediums, for example, air and an immersion oil. This method has been used for over a century to obtain (n, k) -estimations especially of strongly absorbing samples which transmit insufficiently the light to use an index-matching technique. Sb_2S_3 reflectivity at normal incidence, for the electric field vector $\mathbf{E}$ parallel to the a , b and c axes, in air and in immersion oil, were tabulated at 20-nm intervals for wavelengths from 400 nm to 700 nm by Criddle and Stanley [49]. Table 2 compares reflectivity data calculated using the n_j and k_j values with the historic intensity measurement data, and the agreement is excellent considering the large uncertainty by which the previous intensity experiment was potentially affected, such as by surface overlayers, roughness, misalignment of polarization direction and/or crystallographic misorientation. A detailed discussion of these results is given in Ref. 48. The future task is to explore the n_j and k_j data over the largest possible spectral region to explore phonon modes, fundamental band-to-band transitions and the spectral dependence of the birefringence for this interesting semiconducting compound.

7.2 Phonons in wurtzite-type films on sapphire substrates

Generalized ellipsometry is applicable over the entire electromagnetic spectrum, in principle, provided ellipsometers capable of measuring the Jones or Mueller matrix elements can be designed. Phonon mode and free-carrier properties affect the infrared response of uniaxial wurtzite-type group-III nitride semiconductor film structures from $\lambda_0 \sim 10 \mu\text{m}$ to $\lambda_0 \sim 35 \mu\text{m}$ [51]–[58]. Infrared active phonon modes cause singularities (transverse optic modes: TO) and zeros (longitudinal optical modes: LO) within the materials' dielectric function spectrums, and can be obtained from lineshape functions matched to the ellipsometry data upon model calculations [51]–[58]. Noncubic solids are anisotropic at wavelengths λ_0 near infrared active phonon modes, which differ for the polarizations along the major crystal axes. Generalized ellipsometry can be used to differentiate crystal axis orientation and major dielectric function spectra in anisotropic noncubic film structures.

Figure 6 shows Ψ_{pp} , Ψ_{ps} and Ψ_{sp} spectra from a c plane oriented AlN film grown on a -plane sapphire. The surface lattice plane orientation for the wurtzite and corundum structures are r -plane: ($1\bar{1}02$), a -plane: ($11\bar{2}0$), and c -plane: (0001). Details about phonon modes and infrared properties of wide-band-gap AlN films and sapphire are given in Refs. 59-61. Besides the AlN TO mode perpendicular to the c axis (E_1), the generalized ellipsometry data reveal distinct features at all phonon mode frequencies of the sapphire substrate, which can be determined from this measurement as indicated by brackets in Figure 6. For the AlN film, according to selection rules, only the $E_1(\text{TO})$ mode is allowed. For general mode assessment, off-axis orientations are required as shown in Figure 7.

Several high-symmetry orientations of substrates were studied [59] in order to determine phonon frequency and broadening parameters for sapphire parallel (E_u) and perpendicular (A_{2u}) to the c axis using standard ellipsometry. For the geometry in Figure 6, the generalized ellipsometry data Ψ_{ps} and Ψ_{sp} crucially depend on the sample in-plane orientation. Here, the sapphire c axis was found from the model analysis to form an angle of -45.97° with the plane of incidence, and was aligned parallel to the sample surface in agreement with the supplier information; the substrate was an a -plane cut. The TO mode of the AlN film parallel to the AlN c axis is forbidden in this geometry, and only the TO mode perpendicular (E_1) to the lattice c axis is resonantly excited. Figure 7 depicts Ψ_{pp} , Ψ_{ps} and Ψ_{sp} spectra for an a -plane wurtzite-type ZnO film grown on r -plane sapphire.

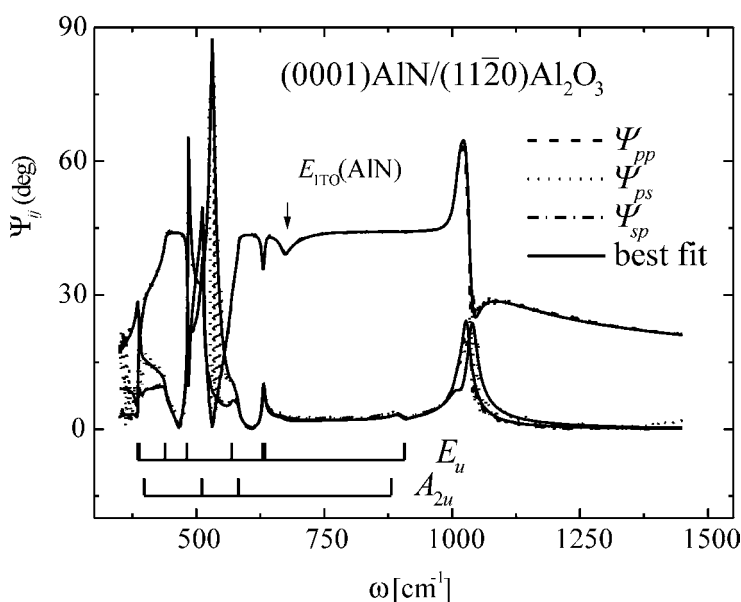


Figure 6 Experimental (broken lines) and best-fit model calculations (solid lines) of infrared generalized ellipsometry data Ψ_{pp} , Ψ_{sp} , and Ψ_{ps} for a (0001)-oriented AlN film (~ 30 nm) deposited on a -plane ($11\bar{2}0$) sapphire [61]. Model analysis reveals the in-plane crystallographic orientation and phonon bands (A_u , E_{2u} , Ref. 59) of the sapphire substrate.

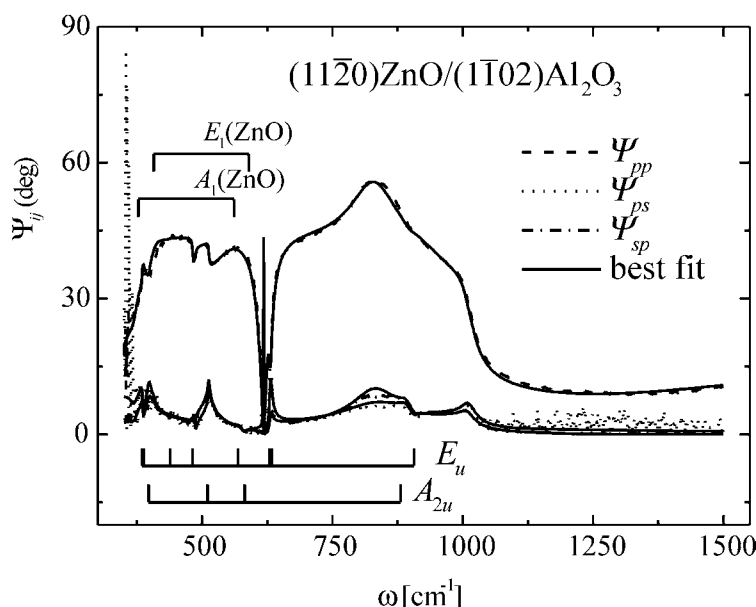


Figure 7 Experimental (broken lines) and best-fit model calculations (solid lines) of infrared spectroscopic generalized ellipsometry data Ψ_{pp} , Ψ_{sp} and Ψ_{ps} for an a -plane ZnO film deposited on r -plane sapphire. The ZnO film adopts an a -plane surface orientation during growth. Data shown here belong to a sample azimuth orientation φ , where φ equals the angle between the plane of incidence and the projection of the optical axis onto the sample surface. See also Figure 1. Here, both c axes and the normal sample share one common plane, and $\varphi = 50.46^\circ$.

ZnO and related compounds have recently gained new interest as possible candidates for blue, violet and ultraviolet emitter and detector structures; see [62] and references therein for recent studies of fundamental optical properties of novel wurtzite-type ZnMgO compounds. Due to its off-axis surface orientation, the ZnO film in Figure 7 reveals all phonon modes with polarization parallel (A_1) and perpendicular (E_1) to the c axis. These are indicated by brackets in Figure 7. The lower-frequency vertical bars indicate ω_{TO} ; the upper-frequency vertical bars denote ω_{LO} . Furthermore, the epitaxial relationship, i.e., the lattice orientation between the sapphire and the heteroepitaxial film, is determined because the generalized ellipsometry data reveal the c axes orientations of both materials with respect to each other, and with respect to the laboratory coordinate system. The sapphire is r -plane, and its c axis is found to form an angle of 54.5° with the growth direction (normal sample). For the geometry in Figure 7 the in-plane orientation of the sapphire c axis is found 50.46° away from the plane of incidence. For the ZnO film, the model analysis reveals that the c axis is tilted by 89.4° from the normal sample, i.e., parallel to the film surface. Its angle towards the plane of incidence also amounts to 50.46° indicating that both substrate and film c axes and the sample normal share one common plane. Because of the p - s polarization mode coupling for off-axis-cut layers, the generalized ellipsometry approach is mandatory, and the

standard ellipsometry technique fails during data acquisition. For unknown materials in layered structures, measurements at multiple in-plane sample azimuths and subsequent simultaneous analysis of all data sets is highly recommended. Such data were discussed previously for an AlN/GaN heterostructure grown on *a* plane sapphire [63]. The data shown here for the ZnO film depict only one set out of sets measured at multiple azimuth angles φ and multiple angles of incidence Φ_a .

7.3 Partially CuPt-type ordered (Al,Ga)InP₂

Long-range chemical ordering (LRO) is widely observed in ternary or quaternary III-V semiconductor alloys grown by metal-organic vapor phase or molecular beam epitaxy [64]. The spontaneous arrangement of group-III elements during the growth process in alternating $\{111\}_B$ planes constitutes ordering of the CuPt_(B) type in, e.g., quaternary $(A_yB_{1-y})^{\text{III}}C^{\text{III}}D_2^{\text{V}}$ zincblende compounds such as $\text{Al}_y\text{Ga}_{1-y}\text{InP}_2$, or $\text{Al}_y\text{Ga}_{1-y}\text{InAs}_2$. The stacking direction of the superlattice planes is usually found according to ordering within either one of the $\{111\}_B$ planes, i.e., parallel to the $[\bar{1}11]$ or the $[1\bar{1}1]$ direction shown in Figure 8, hereafter referred to as CuPt-type. The ordered phase consists of a 1,1-superlattice of monolayer planes $(A_yB_{1-y})_{x+\eta/2}C_{x-\eta/2}$ and $(A_yB_{1-y})_{x-\eta/2}C_{x+\eta/2}$, where x and y are the composition parameters of the random substitutional $(A_yB_{1-y})_xC_{1-x}D$ alloy, and η is the difference between the compositions x of two subsequent sublattice planes within the CuPt superlattice. A perfect alternate stacking of the superlattice planes was not reported yet. Only partially ordered epilayers are formed during the growth process [64]. For partially ordered material the parameter η can be used to quantify the degree of spontaneous CuPt-type LRO. As a result of CuPt-type LRO, the symmetry of the ordered structure is lowered from cubic (T_d) to trigonal (C_{3v}) by the superperiodicity along the ordering direction. This symmetry change affects the physical properties of the ordered material (the trigonal symmetry causes ϵ to

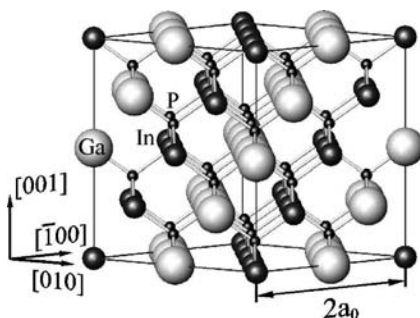


Figure 8 Crystal structure of perfectly CuPt_B-ordered $\text{Al}_y\text{Ga}_{1-y}\text{InP}_2$ for ordering along the $[111]$ direction. Ga(Al) and In arrange at common sublattice sites within subsequent $\{111\}$ planes, spontaneously during nonequilibrium growth processes (metal-organic vapor epitaxy, or molecular beam epitaxy), and depending on growth parameters, which rule the surface growth kinetics. Perfect ordering has not been reported yet, but already partial CuPt-type ordering causes substantial deviation from cubic symmetry and related properties [64].

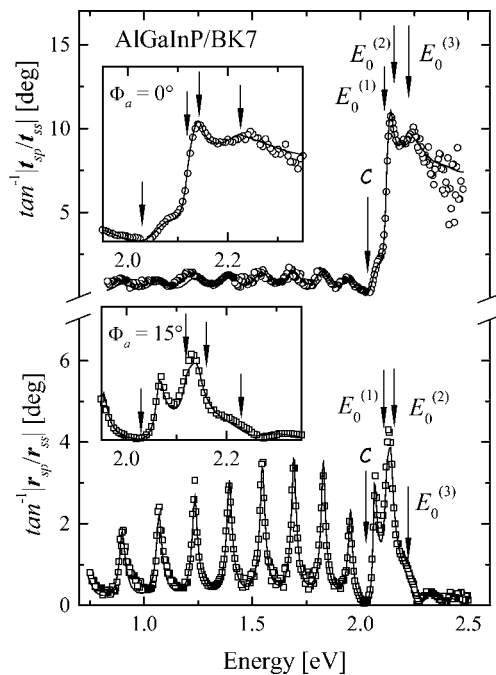


Figure 9 Transmission (○) and reflection (□) data ψ_{sp} (—best fit) from a CuPt-ordered $\text{Al}_{0.48}\text{Ga}_{0.52}\text{InP}_2$ placed on glass. The insets enlarge the zone-center transition region.

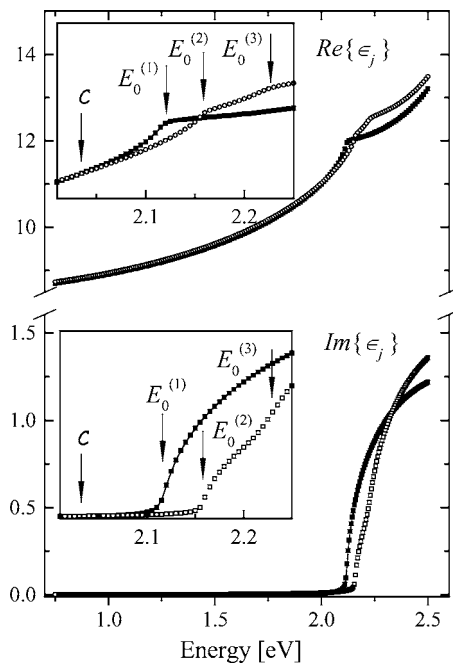


Figure 10 $\varepsilon_{\perp}$ (solid squares) and $\varepsilon_{\parallel}$ (open squares) of CuPt-ordered $\text{Al}_{0.48}\text{Ga}_{0.52}\text{InP}_2$. The order birefringence allows simultaneous identification of the zone-center transitions of the CuPt-ordered alloy.

be uniaxial), and has caused strong research interest within the last decade (see Ref. 64, or references in Refs. 45 and 65).

CuPt LRO reduces the fundamental band gap by ΔE_g . This gap reduction can be as large as 430 meV or 270 meV for $\eta = 1$ in GaInP₂ or AlInP₂, respectively, which are the working-horse alloying materials for today's high-power light-emitting diodes used for traffic and automobile illumination [66]. Hence, CuPt LRO-induced physical properties are still of great importance. CuPt LRO also splits the Brillouin-zone Γ -point valence band states of the disordered compound into states $|\Gamma_{4,5v} \rangle$, $|\Gamma_{6(1)v} \rangle$, and $|\Gamma_{6(2)v} \rangle$, and leads to polarized interband transition matrix elements. Sometimes treated as "chemical" strain, the effects of the CuPt LRO are similar to biaxial strain along the LRO direction. The CuPt-LRO-induced birefringence (order-birefringence) is very small.

The spectral dependencies of $\varepsilon_a = \varepsilon_b = \varepsilon_{\perp}$, $\varepsilon_c = \varepsilon_{\parallel}$ reveal the transitions energies from the splitted valence band states to the conduction band states (Γ_{6c}) at the center of the Brillouin zone ($E_0^{(1)} : \Gamma_{4,5v} - \Gamma_{6c}$, $E_0^{(2)} : \Gamma_{6(1)v} - \Gamma_{6c}$, $E_0^{(3)} : \Gamma_{6(2)v} - \Gamma_{6c}$). These transitions can further be used to derive the order parameter η from optical measurements, such as those demonstrated here [45]. The coordinate system of the CuPt-type ordered structure does not coincide with the laboratory coordinate system (x , y , z) because the optical axis of the uniaxial material is tilted from the sample normal (z axis) by $\cos \theta = 1/\sqrt{3} \sim \cos 54.7^\circ$.

A thin film of Al_{0.48}Ga_{0.52}InP on a glass slide is shown here as example. The degree of CuPt LRO is $\eta = 0.47$, and the band gap is redshifted by 84 meV, with $E_0^{(1)} = 2.116$ eV, $E_0^{(2)} = 2.141$ eV, and $E_0^{(3)} = 2.22$ eV. The film thickness is 1210 nm. Figure 9 depicts Ψ_{sp} data measured in transmission and reflection, where the optical axis (the ordering direction, which is perpendicular to the ordered sublattice planes in Figure 8), is equally oriented between the (xz) and (yz) planes; see Figure 2, i.e., $\varphi = 45^\circ$. The data sets are fingerprints of the near-band-gap CuPt order birefringence, which is depicted in Figure 10. There, the ordinary and extraordinary dielectric function spectrums, obtained upon model analysis of the generalized ellipsometry data from Figure 9, are presented. These spectrums reveal the near-band-gap CuPt order birefringence, and the zone-center transition energies, which are indicated by vertical arrows; see Ref. 45 for a full account of the results reviewed here. The long-wavelength response, rendered by anisotropic phonon modes within the same sample, is discussed elsewhere [65]. Note finally that if the cations within the quaternary alloy were randomly distributed within the respective sublattices, all data points in Figure 9 would be at zero; i.e., no p - s mode polarization conversion would occur.

7.4 Anisotropic refractive indexes and geometry of chiral liquid crystals

Figure 11 presents a summary of a generalized ellipsometry study on cells filled with different chiral liquid crystals [8]. Depicted are data taken in transmission (the cells are bound by glass slides), from where the thickness of the cells, the

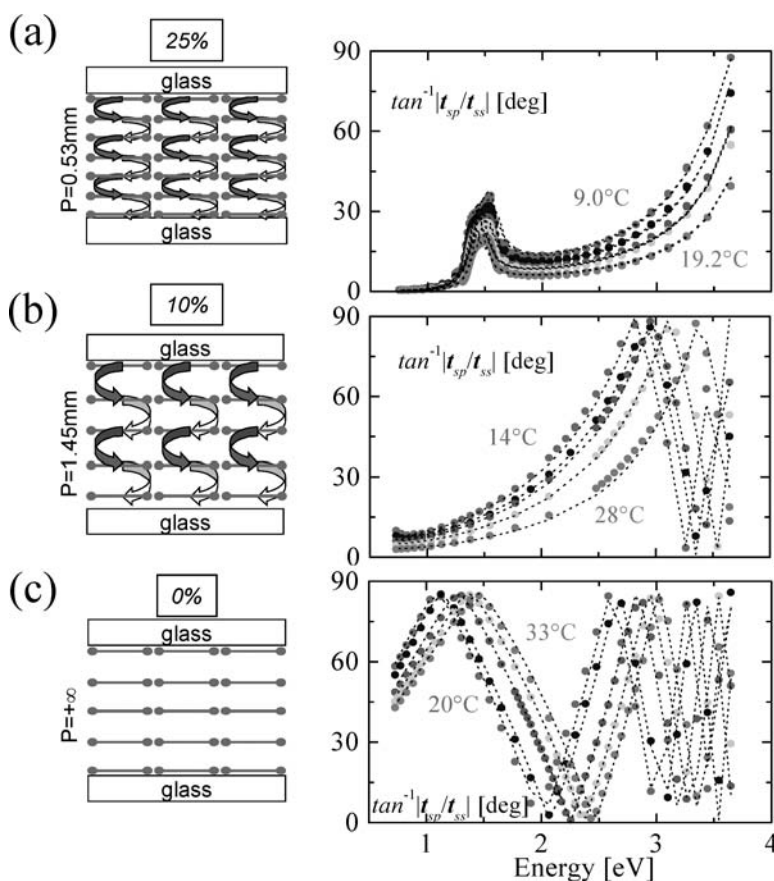


Figure 11 Summary of generalized (transmission at normal incidence) ellipsometry data Ψ_{sp} as example, and schematic sample structures for cells of chiral liquid crystal mixtures 4-cyano-4'-pentylbiphenyl (5CB) and 4-cyano-4'-(2methyl)-butylbiphenyl (CB15), and three different contents of CB15; 25% in (a), 10% in (b) and 0% in (c). The pitch P of the chiral liquid crystal is the distance between one full turn of molecule direction n . The sign of P refers to a right (positive)- or left (negative)-handed helical arrangement of the molecules.

handedness, the length of the pitches, and the spectral dependencies of n_o and n_e were obtained upon best-fit model calculations.

Chiral liquid crystals are fundamental for flat-panel display technologies. Optical phenomena such as polarization-selective reflection and polarization rotation are still interesting research topics. Due to the helical periodicity, thin layers of chiral liquid crystals provide Bragg conditions for LCP or RCP light within spectral regions of the so-called selective reflection. Depending on the structural handedness, the wave with the circular polarization that matches the arrangement of the helix is reflected, whereas the other polarization is transmitted. At normal incidence, and for chiral liquid crystals made from uniaxial rod-like molecules, this band of high reflectivity for either LCP or RCP waves is

centered at $\lambda_0 = |P(n_e + 2n_o)|/3$, and the width of this band is approximately $\Delta\lambda_0 \sim |P(n_e - n_o)|$, where $|P|$ is the helix pitch.

Different values of P can be used to build devices that will pass or reflect light of different wavelengths. Therefore, P , n_e and n_o are important for device designs. The generalized ellipsometry approach is the only technique known so far that allows simultaneous measurement of P , n_e , and n_o within the twisted nematic phase in thin cells. Such a measurement needs to be performed as a function of temperature T and at very stable sample temperatures, because LRO in chiral liquid crystals depends crucially on the sample temperature. This dependence can be clearly seen in Figure 11. The sensitivity to temperature-induced changes of n_e and n_o is revealed by distinct changes in Ψ_{ps} and Ψ_{sp} versus T . The strong change of the Jones matrix elements when the structure (P) changes in Figure 11 reflects the uniqueness of the tunable optical properties of chiral liquid crystals cells. The temperature- and wavelength-dependent refractive indexes of the chiral liquid crystal from the cells studied in Figure 11 are shown in Figure 12. These values were obtained from a best-fit model analysis of the generalized ellipsometry parameters [8].

7.5 Sculptured thin films

A new class of designed matter is sculpturing solid-state materials in thin-film form [67]. Such designs involve physical deposition techniques in three-dimensional growth regimes, where, depending on growth parameters and appropriate substrate rotation, “zig-zag” patterns, “S”-shapes or helixes can be deposited [68]–[71]. Helical arrangements are the solid-state analog of the chiral liquid crystal classes. Design dimensions can be well within the nanometer region, providing interesting grounds for new applications such as micro-filters, micro-antenna arrays, micro-springboards, or “frozen” chiral liquid crystals for optical

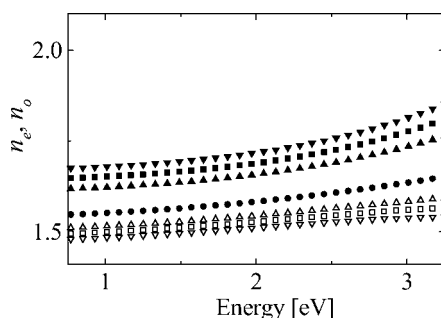


Figure 12 Extraordinary (solid symbols) and ordinary (open symbols) refractive indexes of chiral liquid crystal mixtures 4-cyano-4'-pentylbiphenyl (5CB) and 4-cyano-4'-(2methyl)-butylbiphenyl (CB15). Except for different clearing temperatures T_c (307.75 K for 0% CB15, 301.7 K for 10% CB15, 293.05 K for 25% CB15), all mixtures possess the same optical constants at equally reduced temperatures $t = T/T_c$. At $t = 1$ the liquid crystals are isotropic (circles). Shown here are n_e, n_o for $t = 0.998$ (up-triangles), 0.98 (squares), and 0.9 (down-triangles).

filters. Crucial for design is the choice of the right dimensions, because electrical and optical properties will depend on confinement effects leading to form birefringence. Due to the complexity of such films, optical characterization is a challenge.

The generalized ellipsometry approach was recently extended for nondestructive characterization of chiral sculptured thin films [63]. The normal-incidence setup was not applicable there because the substrate (silicon) is absorbing for the wavelength range of interest. Analysis of data gained under off-axis light propagation has become available some years ago [23], and its feasibility in generalized ellipsometry has also been demonstrated [14]. Perfect agreement between model and experiment was not achieved [63] because inhomogeneous growth and depolarization due to light scattering impaired proper modeling of the real sample structure and increased the uncertainty limits of the experimental generalized ellipsometry data, respectively, and affected the correctness of the resulting optical constants. The results thus far are encouraging for application of generalized ellipsometry in order to determine structure and optical properties of sculptured thin films, and further work in this area is anticipated.

7.6 Far-infrared magneto-optic birefringence in *n*-type GaAs

A long-term goal has been the accurate measurement of the “inertial” effective mass m of free carriers with charge q in matter by noncontact optical means. The effective mass concept descends from the similarity with the Newton force equation (acceleration of a body with mass m) and the acceleration experienced by a Bloch electron due to an external force. The inverse tensor obtained thereby depends on the curvature of the electron’s energetic states versus its momentum, and which is diagonal by a suitable choice of axes. Different experiments require different concepts, resulting in a definition of the effective conductivity mass, the density-of-states effective mass, the Hall effective mass, or the cyclotron effective mass, all of which are not discussed here. For the material used here as an example (GaAs), the response of the zincblende Γ -point conduction-band (single species, i.e., single-valley) Bloch electrons studied at infrared wavelengths is on a time scale much smaller than the average time between scattering events of the free electrons. One may also refer to the effective mass here as an (infrared) optical effective mass.

According to the Drude theory, the relative permittivity ε provides access to the square of the free-charge-carrier (unscreened) plasma frequency ω_p^{*2} and broadening γ_p . Under simplified assumptions about the free-carrier momentum distribution function—a single-species carrier plasma, for example— ω_p^{*2} and γ_p translate into the coupled quantities N/m and $m\tilde{\mu}$. To determine the effective mass m and the free-carrier mobility $\tilde{\mu}$, the free-carrier concentration N must be known from a different experiment. Usually, the electrical Hall effect is measured to access N , which requires ohmic contacts to the sample. No distinction can be made between N , m , and $\tilde{\mu}$ from measurement of ε alone, nor can the sign of q be differentiated because ω_p^{*2} is proportional to q^2 . In a magnetic field, however, the Lorentz force,

an axial vector, diverts the movement of the free charge carriers, which causes birefringence, and which is proportional to m^{-1} and $\text{sgn}(q)$. The permittivity tensor at optical wavelengths is then a complex nonsymmetric tensor, predicted by Drude theory.

Generalized ellipsometry at far-infrared wavelengths (wavenumbers from 150 cm^{-1} to 600 cm^{-1}) can be applied for measurement of the anisotropic dielectric response of doped polar semiconductors in layered structures within an external magnetic field $\mu_0 \mathbf{H}$. Upon determination of normalized Mueller matrix elements, and subsequent derivation of the normalized complex Jones reflection matrix $\mathbf{r}$ of a n -type doped GaAs substrate covered by a highly resistive GaAs layer, the spectral dependence of the magneto-optic permittivity (dielectric function) tensor of n -type GaAs with a free-electron concentration of $1.6 \times 10^{18} \text{ cm}^{-3}$ is obtained on a wavelength-by-wavelength basis, where the magnetic field of 2.3 Tesla was applied at room temperature. These data are in excellent agreement with values predicted by the Drude model. From the magneto-optic generalized ellipsometry measurements of the layered structure, the free-carrier concentration N , their optical mobility $\tilde{\mu}$, the effective mass parameters m , and the sign of the charge carriers can be determined independently in a nondestructive manner, i.e., *without* performing electrical measurements that require contacts to the sample and are thus invasive. A detailed description of the complex measurement procedure is beyond the scope of this review. While a thorough discussion and data presentation is available elsewhere [72-73], a brief survey of the results is given here.

A GaAs homostructure, consisting of a Te-doped n -type GaAs substrate and an undoped (intrinsic, i -type; thickness d) GaAs epilayer, is chosen as an example. The classical Drude theory of free carriers within a polar dielectric material at long wavelengths, subject to an external magnetic field, suffices to describe the optical response of such a material. In this picture, ϵ becomes a nonsymmetric tensor, of which the diagonal elements are proportional to the plasma frequency ω_p , and the off-diagonal elements are proportional to the cyclotron frequency ω_c . The imaginary parts of both diagonal and off-diagonal elements reveal information on the plasma-broadening parameter γ_p . The parameters ω_p , ω_c and γ_p transform into N , $\tilde{\mu}$, and m^* , provided that certain assumptions about the carrier momentum distribution function hold. The sign distribution on the antisymmetric part of ϵ reveals the sign of the charge carriers (holes have positive, and electrons have negative signs).

Effects related to this birefringence are the well-known Faraday rotation of long-wavelength radiation that passes through a thick layer of such a medium. The novelty of the present approach is that the full tensor ϵ can be mapped on a wavelength-by-wavelength basis, and in reflection across spectral regions where the sample is opaque. This is particularly important for polar semiconductors, where lattice resonance absorption prohibits light penetration through bulklike samples. Furthermore, the ellipsometry approach allows to detect such information from thin films, and can handle complex layered heterostructures as well.

Figure 13 depicts Ψ_{pp} , Ψ_{ps} and Ψ_{sp} from the i -GaAs/ n -GaAs homostructure, with the magnetic field $\mathbf{H}$ turned off and on. The upper set of graphs depict the

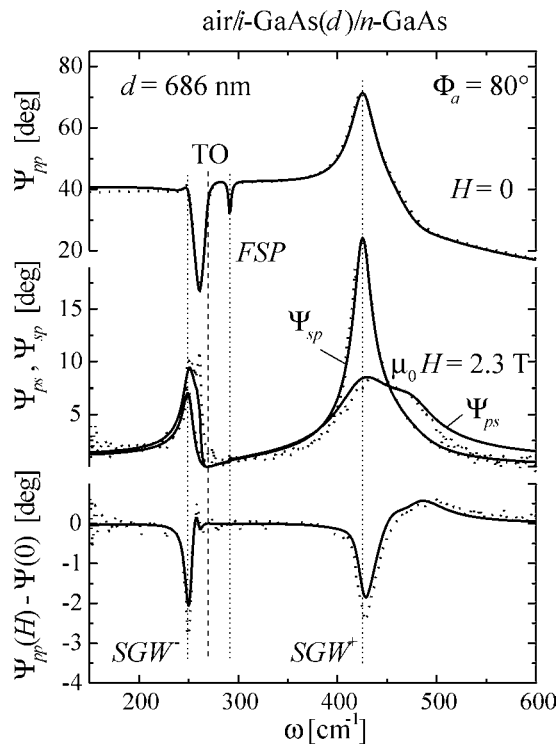


Figure 13 Experimental (dotted lines) and best-fit (solid lines) far-infrared magneto-optic generalized ellipsometry (Jones matrix) spectra for an *i*-GaAs(*d*)/*n*-GaAs homostructure. Upper panel: Ψ_{pp} at $|\mathbf{H}| = 0$. Dotted vertical lines indicate frequencies of interface modes (FSP , $SGW^{+,-}$) discussed in Ref. 72. The dashed vertical line denotes the wavenumber of the GaAs TO frequency. Middle panel: Ψ_{ps} , Ψ_{sp} at $\mu_0|\mathbf{H}| = 2.3$ T. Lower panel: difference spectra $\Psi_{pp}(\mu_0|\mathbf{H}| = 2.3 \text{ T}) - \Psi_{pp}(0)$.

Ψ_{pp} spectra for $\mu_0|\mathbf{H}| = 0$. The middle and lower set of graphs depict Ψ_{ps} , Ψ_{sp} and Ψ_{pp} , respectively, as differences between spectra for $\mu_0|\mathbf{H}| = 2.3$ Tesla and spectra for $|\mathbf{H}| = 0$; note that $\Psi_{sp} = \Psi_{ps} = 0$ for $|\mathbf{H}| = 0$. The magnetic field was applied in the direction perpendicular to the sample surface (“polar magneto optic” setup; PMO). The dotted lines represent a wavelength-by-wavelength model inversion of the experimental data, which were actually taken in the Mueller matrix presentation. The solid lines are best-fit model calculations based on the Drude theory for the *n*-type GaAs. Figure 14 presents the complex-valued spectra of ϵ , following upon the wavelength-by-wavelength as well as the Drude model approach. The inset displays the best-fit Drude model parameters. As can be seen, an almost perfect match between both approaches is obtained, providing direct experimental evidence for the validity of classical Drude theory for doped GaAs. Note that the complex-valued spectra of ϵ can be obtained only by the generalized ellipsometry procedure, which was not anticipated by traditional reflection or transmission intensity techniques. Magneto-optic generalized ellipsometry is a novel approach for exploration of free-carrier parameters in complex organic or

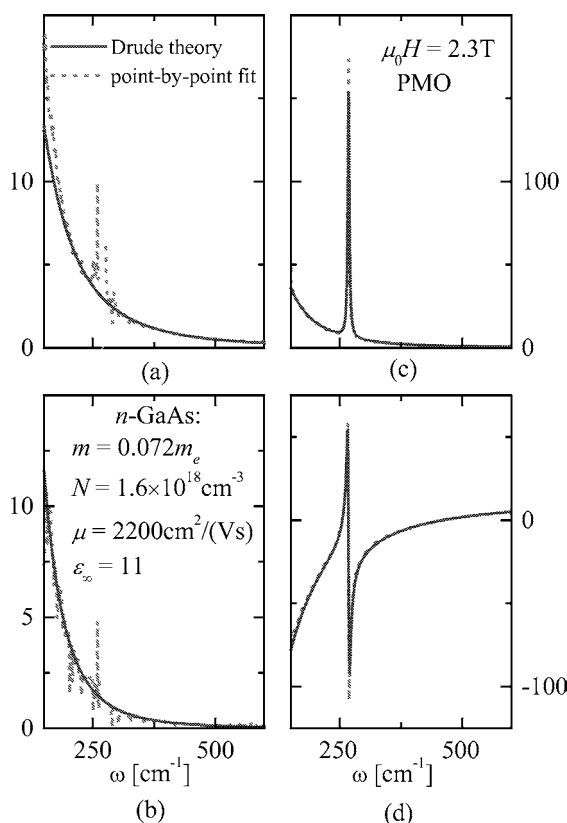


Figure 14 Wavelength-by-wavelength inverted (dotted lines) and Drude-model best-fit (solid lines) far-infrared magneto-optic spectra ε_{xx} , and ε_{xy} for n -GaAs obtained through model analysis of the experimental data shown in Figure 13. The wavelength-by-wavelength inverted spectra follow closely the Drude model. (a) $\text{Im}\{\varepsilon_{xy}\}$, (b) $\text{Re}\{\varepsilon_{xy}\}$, (c) $\text{Im}\{\varepsilon_{xx}\}$, (d) $\text{Re}\{\varepsilon_{xx}\}$. The ε_{zz} spectra are virtually identical to ε_{xx} , and are therefore omitted here.

inorganic semiconducting materials heterostructures, regardless of the anisotropic properties of the individual constituents, and shall become useful in the near future.

8 Conclusions

The present status of the theory and application of generalized ellipsometry capable for measuring the permittivity tensor at optical wavelengths was reviewed. Generalized ellipsometry was portrayed as a global approach to complex layered samples consisting of mediums with arbitrary anisotropy. Subtle and strong birefringence, caused by low-symmetry materials, or complex geometries can be treated and resolved on a model-supported basis. Two fundamental methods for description of electromagnetic light propagation in layered samples—the Jones and Mueller matrix approach—and their relation to the generalized ellipsometry framework were provided. It was demonstrated how material-related properties such as optical constants, phonon modes, structure and atomic ordering can be derived from layered

anisotropic materials of high complexity. A great potential for application is expected in academic and industrial environments, including low-symmetry materials for which practically no information on ϵ exist, or for optically anisotropic organic materials. Extension of the approach to mediums with nontrivial permeability and magnetoelectric tensors shall be an interesting challenge to the experiment.

Acknowledgments

Contributions to this work, provided by C.M. Herzinger, T.E. Tiwald (JAW Co.), J.A. Woollam, D.W. Thompson (UNL), W. Dollase (UCLA), I.J. Hodgkinson, M. Arnold (Univ. Otago, New Zealand), A. Kasic, T. Hofmann, G. Leibiger, V. Gottschalch, C. Bundesmann, N. Ashkenov, B. Rheinländer, M. Lorenz, E.M. Kaidashev, M. Grundmann, V. Riede, W. Grill, M. Ziese, P. Esquinazi (Univ. Leipzig), E. Schubert, H. Neumann, G. Wagner (IOM Leipzig), F. Scholz (Univ. Stuttgart), I. Pietzonka (OSRAM Opto Semiconductors), V. Darakchieva, H. Arwin, B. Monemar (Univ. Linköping, Sweden), are acknowledged. The author acknowledges generous support provided by the Deutsche Forschungsgemeinschaft (Germany), grants Rh 28/1-1,2 and Rh 28/3-1,2; by the National Science Foundation, USA, grant DMI-9901510, the Bundesministerium für Bildung und Forschung (Germany), grant 03WKI09, and the JOANNEUM Research Forschungsgesellschaft (Austria). Fellowships from the Studienstiftung des Deutschen Volkes (Germany), the Center for Microelectronic and Optical Materials Research at the University of Nebraska-Lincoln, USA, the Swedish Foundation for International Cooperation in Research and Higher Education (Sweden), and the Linköpings Universitet and Linköping Institute of Technology (Sweden), are acknowledged. Permission to reproduce the contents of Figure 5 and Table 2 from Ref. 48, Figures 9 and 10 from Ref. 45, Figures 11 and 12 from Ref. 8, and Figures 14 and 13 from Ref. 72 is also acknowledged. I sincerely thank Werner S. Weiglhofer, who left sadly and untimely, for careful reading of the manuscript and many valuable comments.

References

- [1] P. Drude, Über die Gesetze der Reflexion und Brechung des Lichtes an der Grenze absorbierender Kristalle, *Ann Phys Lpz* **32** (1887), 584–652.
- [2] R.M.A. Azzam and N.M. Bashara, *Ellipsometry and polarized light*, North-Holland, Amsterdam, 1984.
- [3] M. Schubert, B. Rheinländer, B. Johs, C.M. Herzinger and J.A. Woollam, Extension of rotating analyzer ellipsometry to generalized ellipsometry: determination of the dielectric function tensor from uniaxial TiO₂, *J Opt Soc Am A* **13** (1996), 875–883.
- [4] M. Schubert, Theory and application of generalized ellipsometry, *Handbook of Ellipsometry* (G.E. Irene and H.W. Tompkins, eds), Noyes Publications, Park Ridge, NJ, USA, to be published.

- [5] M. Schubert, B. Rheinländer, B. Johs and J.A. Woollam, Application of generalized ellipsometry to complex optical systems, *Polarimetry and ellipsometry* (M. Pluta and T.R. Wolinski, eds), Proc SPIE **3094**, Bellingham, WA, USA, 1997, 255–265.
- [6] M. Schubert, Generalized ellipsometry and complex optical systems, *Thin Solid Films* **313–314** (1998), 323–332.
- [7] T.E. Tiwald and M. Schubert, Measurement of rutile TiO₂ from 0.148 to 33 μm using generalized ellipsometry, *Optical diagnostics methods for inorganic materials II* (L.M. Hanssen, ed), Proc SPIE **4103**, Bellingham, WA, USA, 2000, 19–29.
- [8] M. Schubert, B. Rheinländer, C. Cramer, H. Schmiedel, B. Johs, C.M. Herzinger and J.A. Woollam, Generalized transmission ellipsometry for twisted biaxial dielectric media: Application to chiral liquid crystals, *J Opt Soc Am A* **13** (1996), 1930–1940.
- [9] A. En Naciri, L. Johann, R. Kleim, M. Sieskind and M. Amann, Spectroscopic generalized ellipsometry based on Fourier analysis, *Appl Opt* **38** (1999), 4802–4811.
- [10] G.E. Jellison, Jr. and F.A. Modine, Two-modulator generalized ellipsometry: experiment and calibration, *Appl Opt* **36** (1998), 8184–8189.
- [11] G.E. Jellison, Jr. and F.A. Modine, Two-modulator generalized ellipsometry: theory, *Appl Opt* **36** (1998), 8190–8198.
- [12] M. Schubert, Polarization-dependent optical parameters of arbitrarily anisotropic homogeneous layered systems, *Phys Rev B* **53** (1996), 4265–4274.
- [13] <http://www.jawoollam.com/> (02-01-2003), or <http://www.uni-leipzig.de/~hlp/ellipsometrie/> (02-01-2003).
- [14] M. Schubert and C.M. Herzinger, Ellipsometry on anisotropic materials: Bragg conditions and phonons in dielectric helical thin films, *Phys Stat Sol (a)* **188** (2001), 1563–1575.
- [15] J.A. Kong, *Electromagnetic wave theory*, EMW Publishing, Cambridge, MA, USA, 2000.
- [16] A. Lakhtakia, *Beltrami fields in chiral media*, World Scientific, Singapore, 1994.
- [17] A. Lakhtakia, *Selected papers on natural optical activity*, SPIE Press, Bellingham, WA, USA, 1990.
- [18] L.D. Landau and E.M. Lifschitz, *Electrodynamics of continuous media*, Pergamon Press, New York, NY, USA, 1960.
- [19] W.S. Weiglhofer, Constitutive characterization of simple and complex mediums, *Introduction to complex mediums for optics and electromagnetics* (W.S. Weiglhofer and A. Lakhtakia, eds), SPIE Press, Bellingham, WA, USA, 2003.
- [20] A. Lakhtakia and W.S. Weiglhofer, On light propagation in helicoidal bianisotropic mediums, *Proc Roy Soc Lond A* **448** (1995), 419–437; erratum: **454** (1998), 3275.

- [21] A. Lakhtakia and W.S. Weiglhofer, Further results on light propagation in helicoidal bianisotropic mediums: oblique propagation, *Proc Roy Soc Lond A* **453** (1997), 93–105; erratum: **454** (1998), 3275.
- [22] W.S. Weiglhofer and A. Lakhtakia, Wave propagation in a continuously twisted biaxial dielectric medium parallel to the helical axis, *Optik* **96** (1994), 179–183.
- [23] A. Lakhtakia and W.S. Weiglhofer, Simple and exact solution for oblique propagation in a cholesteric liquid crystal, *Microw Opt Techn Lett* **12** (1996), 245–248.
- [24] V.C. Venugopal and A. Lakhtakia, Electromagnetic plane-wave response of non-axially excited slabs of dielectric thin-film helicoidal bianisotropic mediums, *Proc Roy Soc Lond A* **456** (2000), 125–161.
- [25] M. Born and E. Wolf, *Principles of optics*, 7th ed., Cambridge University Press, MA, USA, 1999.
- [26] T. Asahi and J. Kobayashi, Polarimeter for anisotropic optically active media, *Introduction to complex mediums for optics and electromagnetics* (W.S. Weiglhofer and A. Lakhtakia, eds), SPIE Press, Bellingham, WA, USA, 2003.
- [27] G.B. Smith, Nanostructured thin films, *Introduction to complex mediums for optics and electromagnetics* (W.S. Weiglhofer and A. Lakhtakia, eds), SPIE Press, Bellingham, WA, USA, 2003.
- [28] T.G. Mackay, Homogenization of linear and nonlinear complex composite materials, *Introduction to complex mediums for optics and electromagnetics* (W.S. Weiglhofer and A. Lakhtakia, eds), SPIE Press, Bellingham, WA, USA, 2003.
- [29] D.B. Litvin, Point group symmetries, *Introduction to complex mediums for optics and electromagnetics* (W.S. Weiglhofer and A. Lakhtakia, eds), SPIE Press, Bellingham, WA, USA, 2003.
- [30] R.M.A. Azzam and N.M. Bashara, Generalized ellipsometry for surfaces with directional preference: Application to diffraction gratings, *J Opt Soc Am* **62** (1972), 1521–1523.
- [31] A. Röseler, *Infrared spectroscopic ellipsometry*, Akademie, Berlin, Germany, 1992.
- [32] D.W. Berreman, Optics in stratified and anisotropic media, *J Opt Soc Am* **62** (1972), 502–510.
- [33] M. Schubert, T.E. Tiwald and J.A. Woollam, Explicit solutions for the optical properties of arbitrary magneto-optic materials in generalized ellipsometry, *Appl Opt* **38** (1999), 177–187.
- [34] W. Xu, L.T. Wood and T.D. Golding, Optical degeneracy in anisotropic layered media: Treatment of singularities in a 4×4 matrix formalism, *Phys Rev B* **61** (2000), 1740–1743.
- [35] P. Yeh, *Optical waves in layered media*, Wiley, New York, NY, USA, 1988.
- [36] D.E. Aspnes, Spectroscopic ellipsometry, *Handbook of optical constants of solids* (E.D. Palik, ed), Academic Press, New York, NY, USA, 1998.

- [37] J.A. Woollam, B. Johs, C.M. Herzinger, J. Hilfiker, R. Synowicky and C.L. Bungay, Overview of variable spectroscopic ellipsometry (VASE), part I: Basic theory and typical applications, *Optical metrology* (G.A. Al-Jumaily, ed), SPIE Press, Bellingham, WA, USA, 1999, 3–28.
- [38] B. Johs, J.A. Woollam, C.M. Herzinger, J. Hilfiker, R. Synowicky and C.L. Bungay, Overview of variable spectroscopic ellipsometry (VASE), part II: Advanced applications, *Optical metrology* (G.A. Al-Jumaily, ed), SPIE Press, Bellingham, WA, USA, 1999, 29–56.
- [39] G.E. Jellison, Spectroscopic ellipsometry data analysis: Measured versus calculated quantities, *Thin Solid Films* **313–314** (1998), 33–39.
- [40] C.M. Herzinger, P.G. Snyder, B. Johs and J.A. Woollam, InP optical constants between 0.75 and 5.0 eV determined by variable-angle spectroscopic ellipsometry, *J Appl Phys* **77** (1995), 1715–1724.
- [41] W.H. Press, B.P. Flannery, S.A. Teukolsky and W.T. Vetterling, *Numerical recipes in C*, Cambridge University Press, Cambridge, UK, 1988.
- [42] C.M. Herzinger, H. Yao, P.G. Snyder, F.G. Celii, Y.-C. Kao, B. Johs and J.A. Woollam, Determination of AlAs optical constants by variable angle spectroscopic ellipsometry and a multisample analysis, *J Appl Phys* **77** (1995), 4677–4687.
- [43] J.-D. Hecht, A. Eifler, V. Riede, M. Schubert, G. Krauß and V. Krämer, Birefringence and reflectivity of single crystal CdAl₂Se₄ by generalized ellipsometry, *Phys Rev B* **57** (1998), 7037–7042.
- [44] M. Schubert, B. Rheinländer, E. Franke, H. Neumann, J. Hahn, M. Röder and F. Richter, Anisotropy of boron nitride thin-film reflectivity spectra by generalized ellipsometry, *Appl Phys Lett* **70** (1997), 1819–1821.
- [45] M. Schubert, T. Hofmann, B. Rheinländer, I. Pietzonka, T. Saß, V. Gottschalch and J.A. Woollam, Near-band-gap CuPt order birefringence in Al_{0.48}Ga_{0.52}InP, *Phys Rev B* **60** (1999), 16618–16634.
- [46] A. Kreuter, K. Otte, G. Lippold, G. Wagner, A. Schindler and M. Schubert, Anisotropic dielectric function spectra from single crystal CuInSe₂ with orientation domains, *Appl Phys Lett* **78** (2001), 195–197.
- [47] H.W. Tompkins and G.E. Irene, eds, *Handbook of ellipsometry*, Noyes Publications, Park Ridge, NJ, USA, 2003.
- [48] M. Schubert and W. Dollase, Generalized ellipsometry for biaxial absorbing materials: Determination of crystal orientation and optical constants of Sb₂S₃, *Opt Lett* **27** (2002), 2073–2075.
- [49] A.J. Criddle and C.J. Stanley, *The quantitative data file for ore minerals*, London, British Museum of Natural History, London, UK, 1986, 420.
- [50] E.P.T. Tyndall, Optical properties of some metallic sulfides, *Phys Rev* **21** (1923), 162–180.
- [51] M. Schubert, A. Kasic, S. Figge, M. Diesselberg, S. Einfeldt, D. Hommel, U. Köhler, J.D. As, B. Kuhn, J. Off, F. Scholz, J.A. Woollam and C.M. Herzinger, Infrared spectroscopic ellipsometry for nondestructive characterization of free-carrier and crystal-structure properties of group- III nitride semi-

- conductor device heterostructures, *Optical metrology roadmap for the semiconductor, optical and data storage industries II* (A. Duparré and B. Singh, eds), Proc SPIE **4449**, Bellingham, WA, USA, 2001, 48–58.
- [52] M. Schubert, J.A. Woollam, A. Kasic, B. Rheinländer, J. Off, B. Kuhn and F. Scholz, Phonon and free-carrier properties of group III-nitride heterostructures measured by infrared ellipsometry, *MRS Internet J Nitride Semicond Res* **4** (1999), 11.
- [53] M. Schubert, A. Kasic, T.E. Tiwald, J.A. Woollam, V. Härle and F. Scholz, Phonons and free-carriers in a strained hexagonal GaN-AlN superlattice measured by infrared ellipsometry and Raman spectroscopy, *MRS Internet J Nitride Semicond Res* **5** (2000), W11.39.
- [54] A. Kasic, M. Schubert, S. Einfeldt, D. Hommel and T. E. Tiwald, Free-carrier and phonon properties of *p*- and *n*-type *a* GaN films measured by infrared ellipsometry, *Phys Rev B* **62** (2000), 7365–7377.
- [55] A. Kasic, M. Schubert, B. Kuhn, F. Scholz, S. Einfeldt and D. Hommel, Disorder-activated infrared modes and surface depletion layer in highly Si-doped hexagonal GaN, *J Appl Phys* **89** (2001), 3720–3734.
- [56] A. Kasic, M. Schubert, J. Off and F. Scholz, Strain and composition dependence of the $E_1(\text{TO})$ mode in hexagonal $\text{Al}_{1-x}\text{In}_x\text{N}$ thin films, *Appl Phys Lett* **78** (2001), 1526–1528.
- [57] A. Kasic, M. Schubert, Y. Saito, Y. Nanishi and G. Wagner, Effective carrier mass and phonon mode behavior in *n*-type hexagonal InN, *Phys Rev B* **65** (2002), 115206.
- [58] A. Kasic, M. Schubert, T. Frey, U. Köhler, D.J. As and C.M. Herzinger, Optical phonon modes and interband transitions in cubic AlGaIn, *Phys Rev B* **65** (2002), 184302.
- [59] M. Schubert, T.E. Tiwald and C.M. Herzinger, Infrared dielectric anisotropy and phonon modes of sapphire, *Phys Rev B* **61** (2000), 8187–8201.
- [60] See, e.g., *GaN and Related alloys 2000* (U. Mishra, M.S. Shur, C.M. Wetzel, B. Gil and K. Kishino, eds), *Mater Res Soc Proc* **639** (2001).
- [61] V. Darakchieva, P.P. Paskov, T. Paskova, J. Birch, S. Tungasmita and B. Monemar, Deformation potentials of the $E_1(\text{TO})$ mode in AlN, *Appl Phys Lett* **80** (2002), 2302–2304.
- [62] C. Bundesmann, M. Schubert, D. Spemann, T. Butz, M. Lorenz, E.M. Kaidashchev, M. Grundmann, N. Ashkenov, H. Neumann, G. Wagner, Infrared dielectric functions and phonon modes of wurtzite $\text{Mg}_x\text{Zn}_{1-x}\text{O}$ ($x \leq 0.2$), *Appl Phys Lett* **81** (2002), 2376–2378.
- [63] M. Schubert, A. Kasic, T. Hofmann, V. Gottschalch, J. Off, F. Scholz, E. Schubert, H. Neumann, I.J. Hodgkinson, M. Arnold, W. Dollase and C.M. Herzinger, Generalized ellipsometry of complex mediums in layered systems, *Complex mediums III: Beyond linear isotropic dielectrics* (A. Lakhtakia, G. Dewar and M.W. McCall, eds), Proc SPIE **4806**, Bellingham, WA, USA, 2002, 264–276.

- [64] A. Zunger, Spontaneous atomic ordering in semiconductor alloys, *MRS Bull* **22** (1997), 27–32.
- [65] T. Hofmann, V. Gottschalch and M. Schubert, Far-infrared dielectric anisotropy and phonon modes in spontaneously CuPt ordered $\text{Ga}_{0.52}\text{In}_{0.48}\text{P}$, *Phys Rev B* **66** (2002), 19204 1–10.
- [66] S.-H. Wei and A. Zunger, Fingerprints of CuPt ordering in III-V semiconductor alloys: valence-band splitting, band-gap reduction, and x-ray structure factors, *Phys Rev B* **57** (1998), 8983–8988.
- [67] A. Lakhtakia and R. Messier, The past, the present, and the future of sculptured thin films, *Introduction to complex mediums for optics and electromagnetics* (W.S. Weiglhofer and A. Lakhtakia, eds), SPIE Press, Bellingham, WA, USA, 2003.
- [68] R. Messier and A. Lakhtakia, Sculptured thin films—II. Experiments and applications, *Mater Res Innovat* **2** (1999), 217–222.
- [69] Q. Wu, I.J. Hodgkinson and A. Lakhtakia, Circular polarization filters made of chiral sculptured thin films: experimental and simulation results, *Opt Eng* **39** (2000), 1863–1868.
- [70] I.J. Hodgkinson and Q.h. Wu, Serial bideposition of anisotropic thin films with enhanced linear birefringence, *Appl Opt* **38** (1999), 3621–3625.
- [71] I.J. Hodgkinson, Q.h. Wu, B. Knight, A. Lakhtakia and K. Robbie, Vacuum deposition of chiral sculptured thin films with high optical activity, *Appl Opt* **39** (2000), 642–649.
- [72] M. Schubert, T. Hofmann and C.M. Herzinger, Generalized far-infrared magneto-optic ellipsometry for semiconductor layer structures: Determination of free-carrier effective mass, mobility and concentration parameters in *n*-type GaAs, *J Opt Soc Am A* **20** (2003), 347–356.
- [73] T. Hofmann, M. Schubert, C.M. Herzinger and J.A. Woollam, Far-infrared magneto-optic generalized ellipsometry determination of free-carrier parameters in semiconductor thin film structures, *Advanced characterization techniques for optical, semiconductor, and data storage components* (A. Duparré and B. Singh, eds), Proc SPIE **4779**, Bellingham, WA, USA, 2003, 90–97.
- [74] M. Schubert and T. Hofmann, Phonons and polaritons in semiconductor layer structures, *Complex mediums IV: Beyond linear isotropic dielectrics* (M.W. McCall and G. Dewar, eds), Proc SPIE **5218**, Bellingham, WA, USA, 2003, 210–222.



Mathias Schubert is scientific assistant at the Assistant Professor level of the Faculty of Physics and Geoscience at the University Leipzig in Germany. He received his Ph.D. in physics in 1997, and the Habilitation in 2003, both from the University of Leipzig. He has published more than 80 journal articles, contributed chapters to three books, and reviewed for 11 journals. He received research fellowships from the Studienstiftung des Deutschen Volkes (Germany), the University of Nebraska-Lincoln, and the Swedish Foundation for International Cooperation in Research and Higher Education (Sweden). He is a member of the German Physical Society and the Materials Research Society. He participated in the program committee for the 3rd International Conference on Spectroscopic Ellipsometry in Vienna, 2003. He is member of the extended executive board of the German Arbeitskreis Ellipsometrie, and has participated in the organization of the biennial German Ellipsometry Workshop. His current research interests are development and application of polarization-sensitive optical techniques operating at far-infrared to deep-ultraviolet wavelengths for determination of polariton properties in complexly layered structures involving anisotropic inorganic and organic semiconductors, ferroelectrics and polymers.

Professor Werner S. Weiglhofer (1962–2003)

David R. Fearn

Professor Werner Weiglhofer of the Department of Mathematics, University of Glasgow, died on January 12, 2003, aged 40. He was struck down by an avalanche while snowshoeing alone on the slopes of Bispen, in the Trollstigen area of Norway, about 350 kilometers (220 miles) northwest of Oslo. The alarm was sounded later that day when he failed to return as expected. A search and rescue team found his body the next day, the day he had been scheduled to fly back to Glasgow. He had been on a week's holiday to an area he knew very well.

The news of his untimely death came as a tremendous shock to all of his friends and colleagues, and tributes came in from all over the world. His is a tremendous loss to the University of Glasgow, and he will be sadly missed. He is survived by his parents Erich and Heide in Austria; he was their only child.

Since we heard the tragic news, I have talked to many people about Werner. We all have our personal memories and stories to tell, and I have been struck by the warmth and affection with which he is remembered by so many. It has been comforting to hear the stories that might not otherwise have been related. I have learned much I hadn't previously known about Werner. There is so much to tell.

Born on 25 August 1962 in Bruck an der Mur, Steiermark, Austria, Werner obtained a doctorate in technical sciences from the Technical University of Graz in 1986. A prestigious Australian–European Fellowship then took him to the University of Adelaide. My own memories go back to May 1988, when Werner arrived in Glasgow to take up a three-year Research Assistant position, to work with me on a project investigating the stability of the Earth's magnetic field. A car journey to a conference in Cambridge three days after he arrived gave me an early opportunity to hear about his enthusiasms. He had just returned from a year in Australia via a conference in China. We heard about river-rafting in Tasmania and the Australian grand prix in Adelaide.

At the end of his postdoc, Werner was successful in being appointed to a Lectureship in Applied Mathematics in Glasgow. Freed from the constraints of a research project, he quickly developed as an independent researcher, returning to the field he had learned as a postgraduate student: electromagnetics of complex materials. He developed a collaboration and very close friendship with Akhlesh Lakhtakia

of The Pennsylvania State University. The collaboration was extraordinarily productive, and Lakhtakia & Weiglhofer papers became a very prominent feature of the Department of Mathematics preprint list. Werner's growing status was recognised through promotion to Senior Lecturer, to Reader, and finally to Professor of Applied Mathematics in August 2002, in recognition of his internationally leading research role in the field of electromagnetic theory of complex materials. Werner was energetic in developing the mathematical methods and theoretical apparatus that is necessary to analyze, understand and eventually exploit novel electromagnetic effects in complex materials.

Figure 1 shows a very proud Werner on the day he was inducted into the University Senate. Werner received the letter announcing his promotion in March last year. He did not open that letter immediately. A friend, Willie Coupar, takes up the story later in the day: "I espied an unusually serious Werner making his way towards the South Front flagpole, clutching an unopened envelope. He informed me that he had applied for a professorship and that the university's decision lay within that envelope which he had been reluctant to open all afternoon. I crossed



Figure 1 Werner on the occasion of his induction as Professor into the Senate of the University of Glasgow.

my fingers for him and watched as he made his way to the flagpole and began to open the envelope. To my delight I got an e-mail from him the next morning to say that it had contained good news.”

With his new status came a larger office, newly decorated and carpeted. Werner was very pleased with this. So woe betide friend and colleague Dr. Alec Mason, who had the temerity to enter the office immediately after a lecture, leaving chalky footprints on the brand-new carpet. Another close friend and colleague Dr. Ted Spence told me that he subsequently always removed his shoes before entering Werner’s office.

Werner had a very caring side to his character. Another of the many stories that have been related in recent weeks concerns Ted Spence’s last lecture preceding his retirement. Such an occasion deserves to be marked, but I certainly had not registered the significance of that day. Werner did though, and arrived in the lecture with a celebratory bottle of champagne.

Werner was extremely active in pursuing his research as well as promoting his field, and was meticulously organised in all he did. During his short career, he authored or co-authored more than 125 research publications in peer-reviewed international journals. His work was marked by an elegance that was spartan in style and extensive in scope. He was actively involved with conference organisation, including the very successful *Bianisotropics*’97, held in Glasgow. He raised money from several sources to fund the meeting and was keen that the participants gained a good impression of Glasgow and Scotland. He served on the editorial boards of two journals and was a prolific reviewer for more than 30 different international journals. He won several awards to support his research, including an SOEID/RSE Support Research Fellowship that allowed him to focus on his research for the year 2000.

I had the sad task of visiting Werner’s office to look for lecture notes to pass on to those who were to take on his duties. The notes were clearly arranged, with corresponding files easily found on his computer. An important work in progress was this book. Again, all the paperwork was neatly filed and electronic versions completely up-to-date.

While dedicated to his research, Werner was also an enthusiastic teacher. He co-authored an undergraduate textbook on ordinary differential equations, and was leading a review of the place of this important topic in the mathematics syllabus. He spent many hours devising numerous projects in applied mathematics for his students and planned to publish a collection of these to show that mathematics is not only a thing of beauty but eminently useful too. In his memory we have set up a fund that will award an annual prize, in his name, for the best project.

His enthusiasm for publication was not limited to his scientific work. His post-doctoral travels motivated several articles, including one on Glasgow, that were published by local newspapers in Bruck an der Mur and Graz. More recently, he was motivated to write to local and national newspapers on many issues.

The pre-eminent loves of Werner’s life were mountains. He hiked, he skied, he snowshoed, he photographed, read and wrote about them. The close proximity of



Figure 2 Werner where he loved to be.

the Highlands to Glasgow was one factor in attracting him here in 1988; and it took only four years for him to bag all the Munros (Scottish mountains over 3000 ft in altitude), many of which he since conquered several times. Latterly, he was set on adding the Munro Tops and the Corbetts (Scottish mountains between 2500 ft and 3000 ft) to his remarkable list of achievements. He kept very detailed records and diaries of all his hikes. Most of these were made alone, but there are several friends for whom he shifted down a gear or so to accompany up a Munro.

Vacation time took him to Norway, and in 1991 he discovered the peaks of Romsdal, Norway. He described this chance encounter in his guide to the peaks: “It was in the spring of 1991 that I was planning my third visit to Norway. Having spent some time in the National Parks of Jotunheimen and Rondane in the two previous years, I studied my maps for a new and interesting mountain area to explore. While doing so, the National Park of Dovrefjell caught my eye. Rather than travel there via Oslo, I decided to fly to Molde and continue to Dovrefjell with an overnight stop at a Youth Hostel in a small town named Åndalsnes. And so I found myself at the Vestnes ferry terminal in Molde on 3 July 1991 under a completely blue summer sky. Across the Romsdals-fjord, a landscape of mountains dominated the skyline that appeared to me as if it had been created in a world of fantasy. This first impression was emphasised later that day, when I walked across Setnesmoen just behind the Youth Hostel in Åndalsnes, surrounded by some of the most exciting peaks I had ever set my eyes upon. Since 1991, I have returned to Åndalsnes

on many occasions, to climb and hike in the marvellous mountains of Romsdal and for other, more significant, private reasons.”

He was never happier than when surrounded by snow above clouds, which is where he left his parents, his colleagues, numerous friends, students and admirers.

February 2003

Personal Memories of Werner S. Weiglhofer

Tom G. Mackay

My first encounter with Werner S. Weiglhofer took place in October 1996, when I attended his undergraduate lecture course on *Mathematical Methods*. At that time I was a mature part-time student in the University of Glasgow, with an interest in applied mathematics. I recall being greatly impressed by Werner's meticulous lecturing style. His notes were precise and highly structured.¹ Werner had an unconventional—and rather telling, I believe—strategy for ensuring the accuracy of his printed notes (as well as the attentiveness of his class). Each student who brought a typographical error to their teacher's attention was rewarded with an Austrian chocolate. Needless to add, very few chocolates were ever handed out.

These initial impressions encouraged me to sign up as Werner's first (and only, as it transpired) doctoral student in October 1998. As I had no experience of electromagnetic theory, our first few months together involved Werner patiently taking me step by step through the basics. His explanations were characteristically methodical, concise and honest. He was an approachable and unpretentious supervisor who always had time to help me out, even at the most antisocial of hours. As our relationship developed, our meetings would focus less on electromagnetics and more on Werner's latest mountain trip. Typically, we would discuss work issues at length in the office and then adjourn to a nearby café to chat leisurely about hiking, skiing, snow conditions, etc. It was clear to me that while mathematics satisfied Werner's intellectual appetite, the mountains were his abiding passion.

I completed my postgraduate studies under Werner's supervision in the summer of 2001, and took up a lecturing position at the University of Edinburgh. After my arrival in Edinburgh, Werner and I collaborated on research projects and we remained good friends. In our collaborations, I was often grateful for his acute eye for detail and his uncanny ability to identify the subtlest of inconsistencies. Werner also took a keen interest in my career and was ever helpful in providing support and advice.

¹Subsequently, these notes formed the backbone of the textbook *Ordinary Differential Equations & Applications* by W.S. Weiglhofer and K.A. Lindsay.

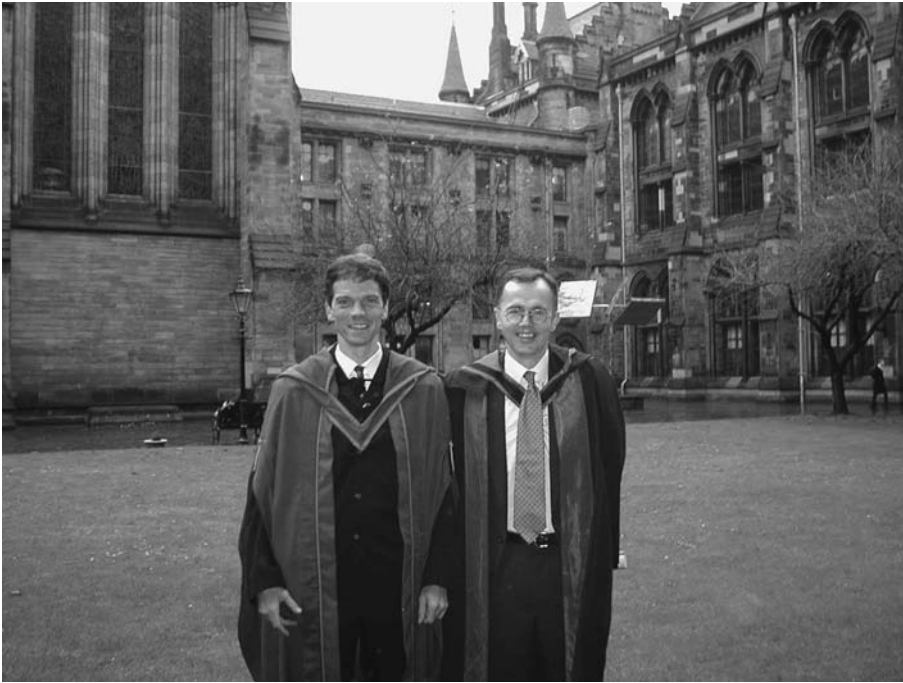


Figure 1 Werner Weiglhofer (left) and Tom Mackay, on November 30, 2001, after the latter received a doctorate from the University of Glasgow under the former's supervision. [Photograph © Estate of Werner S. Weiglhofer.]

Werner and I met for the last time three weeks before his death, when my wife and I were delighted to receive him at our wedding. He quickly endeared himself to my family, and it was a very happy occasion for all concerned. He left saying he was heading up to the North of Scotland to do some hiking over the Christmas period, and we wished each other well.

I shall remember Werner most for his integrity and humanity: he was a true friend to me.

Werner S. Weiglhofer—A Personal Tribute

Edward Spence

Werner came to Glasgow as a Research Assistant in 1988 at the age of 26; but it was not mathematics that brought us together initially, it was running. He was determined to keep himself fit, but unfortunately he was prone to injuries incurred while running, and this leisure activity gradually became secondary to his main love of the outdoors, which was, as all his friends knew, getting to the tops of mountains. Within a short time (four years to be exact), he had ascended all 284 peaks in Scotland over 3000 ft, the so-called Munros. Each trip made and each mountain climbed were meticulously recorded, initially in a notebook, but latterly on his computer. Moreover, with the advent of digital cameras, Werner was able to record his trips in photograph form, and these too were cataloged along with the written reports. Ultimately he built up a gallery containing more than 10,000 photographs, all of which were easily accessible. If he were to be asked, for example, “When was the last time you were up Sgurr nan Gilleann (a mountain in Skye)?” he would be able in a moment to dig out the information along with a photographic record of his trip.

However, he was not content to sit back on his laurels. He was constantly looking for pastures anew, and in 1991 he discovered a range of mountains in Norway that were to become, in his mind, heaven on earth. These are the Romsdal Alps in the Norwegian province of Møre and Romsdal. Naturally, he set his sights on ascending all the summits in the area. I cannot say exactly how many summits there are, but I have seen a map with each summit conquered colored in red and it looks as if the map has measles! He returned to this particular area of Norway many times in the intervening years, and many of these mountains were climbed several times.

It was only in the last few years that I myself became interested in hiking, and consequently he and I, from being colleagues in the same department, became close friends. Discovering my interest, Werner gave me a present of a book on the Scottish mountains with an inscription telling me to have fun in the mountains, as he always did. He was the expert and I the novice, but he was very patient. He took me along the Aonach Eogach ridge in Glencoe (one of the most challenging

ridge walks in the Scottish mainland), and this was certainly something I would not have done alone. By and large, Werner was happiest with the solitude that came with the mountains. He regarded them as a challenge and was not afraid of the physical effort that some mountains required. There is a group of mountains near Ben Nevis called the Mamores, which today incorporate 10 Munros with a total ascent of 3810 meters and which Werner conquered in 17 hours.

Werner was a man whose word you could trust. In all the time I knew him, I never heard him say a bad word about anyone. He was a man of integrity, and a perfectionist. Although English was not his first language, he spoke it and knew its grammar better than many native speakers, a fact with which some of the contributors to this volume might agree. In his role as editor, everything had to be just right. As another indication of his character and his thoughtfulness, I relate the following story. In December 2002, I was due to give my final lecture to the students as a member of staff of Glasgow University, and Werner knew this. Imagine my surprise, and delight, when Werner appeared at the end of my lecture with a bottle of champagne! I am sure that many of Werner's friends will have similar tales to tell.

The two of us had a hike together just before Christmas 2002 and would have been together again before his fateful trip to Norway, but for unforeseen circumstances. Werner perished in the mountains that he loved so much and about which he had written a book, *The Summits of the Romsdal Alps*. I am sure that he would not mind that I quote from his book: “. . . as we move through our lives, it is not the final destinations that matter most but the encounters which happen on the way.” Werner encountered many people in his journey through life and left his mark wherever he went. He was a true friend to me, and I shall miss him.

Memories of Werner S. Weiglhofer

Martin W. McCall

Tiredness overcame me. It was about midnight. I had to indulge in one last e-mail fix before turning in. A couple of insignificant items had come; and a posting to the ComplexMediums yahoogle, subject “Werner,” from Akhlesh Lakhtakia. Odd! I was expecting a communication from Werner anytime with a little push for me to send the revised version of my chapter for this book. But why via Akhlesh? The horror of what hit me next: *I have just received terrible news. My best friend, Werner Weiglhofer died in an avalanche yesterday in Molde, Norway. His body was recovered today by a search party. Sincerely, Akhlesh.* Sleep no longer possible, I telephoned Akhlesh. It seemed the only thing to do.

I first met Werner at *Complex Mediums II* held at San Diego in 2000, at which he gave an invited talk on homogenization of composite materials. Afterwards, I quickly became aware of his longstanding collaboration covering many topics with Akhlesh. In particular, their work during the 90’s on thin-film helicoidal bianisotropic mediums has been an inspiration and a springboard for me to carry out theoretical research on the same theme. As a relative newcomer to the *complex mediums community*, I’ve sometimes felt like a gatecrasher to a long-established party. But Werner and Akhlesh (for me the two names go together) always made me feel welcome and tolerated kindly the naiveties of a “new kid on the block.” As a trio, we were beginning to find a unique synergy in our endeavors, fashioning papers of distinctive style: Akhlesh’s energy, my physicist’s eye, focused by Werner’s acute precision and somewhat irritating rigor with, of all things for a non-native, the English language. I believe this would have blossomed over the coming years. We worked together primarily on just one topic, entering into the current controversy on *negative phase velocity mediums* on which we published a few papers and wrote a chapter for this book.

Strangely for an Austrian, an Indian–American and an Englishman, the three of us shared an interest in cricket for which I often found myself having to make up some nonsense in defense of the game’s vagaries. Such were the discussions we would have away from science, and were as rich and fulfilling as befitted an atheist, an agnostic and a believer. The attraction of opposites also worked at the mundane level: at SPIE conferences, Werner and Akhlesh, both teetotallers, always donated to my benefit their beverage tickets!

Few of us work alone, and we take for granted, perhaps too much, the ease with which we can consult each other over research problems. To have part of that research connectivity suddenly wrenched out like this is bad enough at a professional level. To have lost a friend is terrible. No longer to hear that baritone “Werner here” when he answers the phone. His idiosyncratic conclusion of most sentences with “yerp,” with a rising inflection. His occasionally prickly temperament. All were the trademarks of his tireless and burgeoning intellect, and I will miss them all.

A note from a friend of mine summed it all up rather poignantly: Accidental deaths just seem so mind-blowingly pointless and frustrating, except to remind us to live each minute like we mean it. Let us live our moments in Werner’s memory like we mean them. Werner Weiglhofer, comrade and colleague, rest in peace.

My Friend Werner

Akhlesh Lakhtakia

Werner must have sent me a reprint request around September 1987, because I have a letter dated Dec. 30, 1987, from him acknowledging receipt of that reprint. He was then at the Department of Physics, University of Adelaide, South Australia, as the beneficiary of an European–Australian Fellowship.

In the same letter, he also pointed out that the derivation of a Green function I had published a few months earlier “seems to be a bit unsatisfactory,” although the result was correct. A few months thereafter, his own derivation appeared in print [1], along with a criticism of my derivation. Immediately, I fired off a comment, to which he replied on October 28, 1988, from the Department of Mathematics, University of Glasgow, UK. He acknowledged in his letter that my mathematical “manipulation is indeed permissible [sic],” and that is where the matter rested.

Hardly an auspicious beginning for a deep friendship, one would think. Yet, when it terminated on January 13, 2003, on my learning of his death in an avalanche, I was able to say that Werner Siegfried Weiglhofer was the closest of my friends ever; indeed, he had become my brother. Many relationships endure because of kinship and marriage, but only a few do because of a convergence in views that is broad enough to celebrate diversity too. Ours was like that. We met for dinner one evening in Stockholm, during the 1989 URSI International Symposium on Electromagnetic Theory. There was some talk of Werner spending a year at Penn State. It came to nought, but we became fast friends. He did visit Penn State often thereafter, most recently in the summer of 2000; and I visited him in Glasgow, the last time in early spring, 1999. I also spent a sabbatical semester in 1995 in Glasgow. We met every year at one conference, at least, and we corresponded several times every day by e-mail.

I am not going to recount here the various research topics that we tackled jointly over a dozen years. It suffices to state here that all of those topics fall under the rubric of time–harmonic electromagnetic fields in complex mediums. Instead, let me reminisce about my friend’s personality.

Werner was passionately devoted to research in electromagnetics. But it is not perhaps widely known, even in the electromagnetics community, that his passion for mountaineering stood a notch higher. During a short visit to Glasgow in March

1995, I was asked by him to buy a lottery ticket for him at Safeway (a supermarket). He hoped that I would bring him better luck than he had; and I joked that, should he win lottery, he would found an *Institute of Mountaineering Electromagnetics*. Werner was not wise in his choice of lucky friends. He lost a round pound.

Anyhow, he went to Romsdal Alps in Norway so often during the 1990s, that the local newspapers, *Åndalsnes Avis* and *Romsdals Budstikke*, chronicled his comings and goings with his pictures. He hiked, he skied, he snow-shoed, and he climbed steep inclines. He was never happier than when in snow, with clouds floating far below. Mt. Bispen was perhaps the most challenging of the mountains he climbed; and, so perhaps, it is fitting that he left us all on the slope of that mountain. He taught himself the vernacular, and published probably the first guidebook in the English language to the Romsdal Alps [2].

Although the Norwegian mountainous terrain was his clear favorite, he often visited the Cascades in Washington State, U.S.A., as well. He had a special fondness for Mt. Rainier. In early summer 1995, he went to Alaska to explore the prospects of ascending Mt. Denali. However, he fell ill in Seattle, and his condition worsened enough for him to be evacuated to Glasgow from Anchorage. During his early years in Scotland, he climbed almost every Munro, and there are more than 300 of these relatively low peaks. He enjoyed happier times in the Alps of central Europe, where he climbed almost all of the major peaks, including Mont Blanc and Großglockner. He had entertained some notions for a while to attempt an Andean peak too, but resolutely kept away from the Himalayas.

Werner's guidebook on the Romsdal Alps was a serious manifestation of what I used to call his listmania. He maintained meticulous records of every hike, every route, and every ascent. He had a fancy watch to measure altitudes, which were duly recorded within 24 hours of his return to Glasgow from every trip. Not only that, he could tell you if he took a flight on a certain day, the flight number, the type of aircraft, the seating arrangement, his seat number, and the arrival and departure times. His diaries are sure to be attractive to some graduate student in history, a hundred years hence, looking for a topic to prepare a doctoral dissertation on. Perhaps, that dissertation would be entitled: *Incidence of aircraft-experienced turbulence during the lifetime of a university professor from the late 20th century*. Had he lived longer, Werner would have shown the door many times to Samuel Pepys.

The managerial qualities of Werner greatly helped him in organizing *Bianisotropics*'97, a specialized three-day conference convened in Glasgow on complex materials. It was a great success, as can be gathered from the conference proceedings [3]. A group photograph of the attendees is reproduced in Figure 1. Werner also enthusiastically contributed to all the following conferences in the *Bianisotropics* series. A picture of him with several other attendees at *Bianisotropics 2000* in Lisbon, Portugal, is shown in Figure 2.

His flair for organization significantly benefitted two SPIE conferences—*Complex Mediums* in 2000 [4] and *Complex Mediums II* in 2001 [5]—as well as the two SPIE conferences bracketing those two.



Figure 1 Group photograph of the attendees of *Bianisotropics'97*, taken on June 5, 1997. Werner is in the top row, sixth from the left. He organized this conference at the University of Glasgow. [Photograph © Estate of Werner S. Weiglhofer.]



Figure 2 At *Bianisotropics 2000* in Lisbon on September 26, 2000 (from left to right): Akhlesh Lakhtakia (Pennsylvania State University), Werner Weiglhofer (University of Glasgow), George Borzdov (Belarus State University), Stephen Sathiaraj (University of Botswana), Elif Ertekin (University of California, Berkeley), Bernhard Michel (Scientific Consulting GmbH), Alexander Borzdov (Belarus State University), Tom Mackay (University of Edinburgh), Gregory Slepyan and Sergey Maksimenko (Belarus State University). [Photograph © Akhlesh Lakhtakia.]

From about 1994, Werner rarely attended a conference alone. A stuffed toy rabbit named Svein Oskar Hansen accompanied him on most trips. Comfortably ensconced in Werner's backpack, often peeking out, Svein would listen to speaker after speaker. Svein rarely said a word; at least, I never heard him. But Werner talked to him all the time. Occasionally, either my family or I would receive e-mails from sveinhansen@hotmail.com, with complaints of indifference from Werner. Svein appears in the 1997 picture reproduced in Figure 3.

Remarkably, Svein co-authored a research paper with Werner on Faraday chiral mediums. That paper appeared in an IEEE journal [6], which also carries authors' biographies. Svein described himself as a Norwegian researcher of independent means, with research interests in dentistry, antennas, and underground prospecting. Werner had more than a chuckle on seeing the published version. I wonder what the editors of the journal would exclaim on finding out that they admitted a toy rabbit to the distinguished coterie of IEEE authors!

Werner was a fiercely political person, though he was so shy that he rarely held forth on nonscientific topics in public. He was devoted to Steiermark, Austria, and the European Union—in that order. He preferred Scotland to England, and

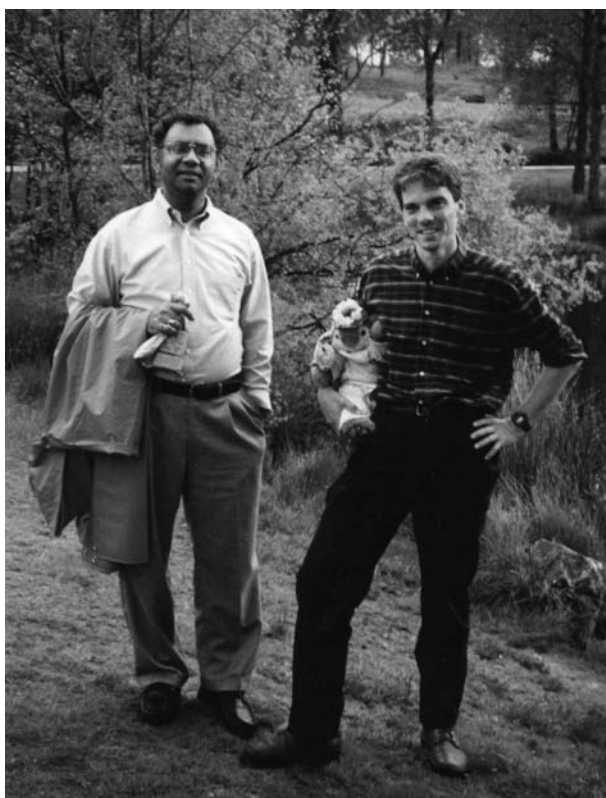


Figure 3 Akhlesh and Werner, with Svein in Werner's arm, in the Queen Elizabeth Forest Park, north of Glasgow, a day after *Bianisotropics'97* ended. [Photograph © Akhlesh Lakhtakia.]

particularly detested the city of London. I could easily understand the causes for his dislike of London, for they are just about the same as for my dislike of New York City, where I have not set foot in 15 years. London charms me, however.

Werner was a pacifist. His family had borne the brunts of two world wars; perhaps, that is the reason for his opposition to war. With the passage of time, his antiwar convictions became more pronounced. In December 2002, he told me that he would not visit the United States ever again, if it bombed Iraq without United Nations sanction.

I am reminded of an argument that occurred between him and me in 1999. He was then guest-editing a special issue of *Electromagnetics* [7], to which I had contributed a paper. My paper was dedicated to *a better future for the Iraqi children, squeezed between a tyrannical oppressor at home and unfeeling wretches abroad*. One of the two referees objected, and Werner ruled that I should not mix politics and research. Eventually, we compromised, and the words *squeezed ... abroad* were eliminated. Late in 2002, one day Werner called to tell me that he had been wrong in making me delete those words. Mixing politics with research was wrong, but mixing research with love for others could not be more right.

Werner was unabashedly an European. He found many American (i.e., U.S.) attitudes inexplicable. I think I alternate between the left and the right on most social and political issues. Werner was always to my left. He did not appreciate excessive American cultural presence in Europe, and would not eat at McDonald's even when very hungry, although he was very fond of Hollywood offerings. He was even more firmly a West European; for instance, after the Cold War ended, he could not stomach the dethroning (in West European magazines) of Mont Blanc as Europe's highest peak by Mt. Elbrus at the border of Georgia and Kazakhstan. But he was always willing to learn. When I suggested in November 2001 that Mt. Elbrus had always (in historical times) been Europe's highest peak but that West Europeans had remained misinformed for a long time, he relented.

So, Werner was a work in progress. He had splendid foundations, personally as well as professionally. Perfectionist that he was, he would only have become better. But that was not to be. A mountain that he loved claimed him at the young age of 40. Surely it must have been someone like Werner who inspired William Wordsworth to compose the following lines:¹

—It is the generous Spirit, who, when brought
Among the tasks of real life, hath wrought
Upon the plan that pleased his boyish thought:
Whose high endeavours are an inward light
That makes the path before him always bright:
Who, with a natural instinct to discern
What knowledge can perform, is diligent to learn;
Abides by this resolve, and stops not there,
But makes his moral being his prime care;

¹W. Wordsworth (1770–1850), *The Happy Warrior*.

References

- [1] W.S. Weiglhofer, Isotropic chiral media and scalar Hertz potentials, *J Phys A: Math Gen* **21** (1988), 2249–2251.
- [2] W.S. Weiglhofer, *The summits of the Romsdal Alps—Toppene av Romsdal-alpene*.²
- [3] W.S. Weiglhofer (ed), *Proceedings of Bianisotropics'97, International conference and workshop on electromagnetics of complex media*, Glasgow, UK, 1997.
- [4] A. Lakhtakia, W.S. Weiglhofer and R.F. Messier (eds), *Complex mediums*, Proc SPIE **4097**, Bellingham, WA, USA, 2000.
- [5] A. Lakhtakia, W.S. Weiglhofer and I.J. Hodgkinson (eds), *Complex mediums II: Beyond linear isotropic dielectrics*, Proc SPIE **4467**, Bellingham, WA, USA, 2001.
- [6] W.S. Weiglhofer and S.O. Hansen, Faraday chiral media revisited—I. Fields and sources, *IEEE Trans Antennas Propagat* **47** (1999), 807–814.
- [7] W.S. Weiglhofer, guest editor, *Electromagnetics* **19** (6), 1999.

²The first edition appeared in 1999, the second two years later. The guidebook could be bought in Åndalsnes, Molde, and nearby towns in Norway, as well as directly from Werner himself.

Published Scientific Works of Werner S. Weiglhofer

Tom G. Mackay

During the period 1985 to 2003, Prof. Werner S. Weiglhofer contributed many works to the scientific literature. Various aspects of electromagnetic theory are described in his publications, including scalar and vector potentials, magnetohydrodynamic instabilities, dyadic Green functions, depolarization dyadics, wave propagation, constitutive relations and homogenization of particulate materials. Complex mediums provide the setting for many of these studies. There follows a comprehensive bibliography of Prof. Weiglhofer's scientific publications.¹ This compilation begins with books and book chapters, followed by journal articles, and finishes with conference publications.

1 Books

1. W.S. Weiglhofer (ed), *Proceedings of Bianisotropics'97, International conference and workshop on electromagnetics of complex media*, Glasgow, UK, 1997.
2. W.S. Weiglhofer and K.A. Lindsay, *Ordinary differential equations & applications: Mathematical methods for applied mathematicians, physicists, engineers, bioscientists*, Horwood Publishing Limited, Chichester, UK, 1999.
3. A. Lakhtakia, W.S. Weiglhofer and R. Messier (eds), *Complex mediums*, Proc SPIE **4097**, Bellingham, WA, USA, 2000.
4. A. Lakhtakia, W.S. Weiglhofer and I.J. Hodgkinson (eds), *Complex mediums II: Beyond linear isotropic dielectrics*, Proc SPIE **4467**, Bellingham, WA, USA, 2001.

¹In addition to the scientific publications listed here, Prof. Weiglhofer also authored numerous nonscientific publications, mostly relating to his passion for mountain pursuits. These are published in various Austrian, British and Norwegian newspapers and magazines. He also wrote and published a monograph entitled *The summits of the Romsdal Alps*.

5. W.S. Weiglhofer and A. Lakhtakia (eds), *Introduction to complex mediums for optics and electromagnetics*, SPIE Press, Bellingham, WA, USA, 2003.

2 Book chapters

1. W.S. Weiglhofer, Frequency-dependent dyadic Green functions for bianisotropic media, *Advanced electromagnetism: Foundations, theory, applications* (T.W. Barrett and D.M. Grimes, eds), World Scientific, Singapore, 1995.
2. W.S. Weiglhofer, Scalar Hertz potentials for linear bianisotropic mediums, *Electromagnetic fields in unconventional materials and structures* (O.N. Singh and A. Lakhtakia, eds), Wiley, New York, NY, USA, 2000.
3. W.S. Weiglhofer, A flavour of constitutive relations: The linear regime, *Advances in electromagnetics of complex media and metamaterials* (S. Zouhdi, A. Sihvola and M. Arsalane, eds), Kluwer, Dordrecht, The Netherlands, 2003.
4. T.G. Mackay and W.S. Weiglhofer, A review of homogenization studies for biaxial bianisotropic materials, *Advances in electromagnetics of complex media and metamaterials* (S. Zouhdi, A. Sihvola and M. Arsalane, eds), Kluwer, Dordrecht, The Netherlands, 2003.
5. W.S. Weiglhofer, Constitutive characterization of simple and complex mediums, *Introduction to complex mediums for optics and electromagnetics* (W.S. Weiglhofer and A. Lakhtakia, eds), SPIE Press, Bellingham, WA, USA, 2003.

3 Guest editorials for special issues of journals

1. W.S. Weiglhofer, A perspective of bianisotropy and Bianisotropics'97, *Int J Appl Electromag Mech* **9** (1998), 93–101.
2. W.S. Weiglhofer, Introduction to special issue: Mathematical techniques in electromagnetics of complex mediums, *Electromagnetics* **19** (1999), 475–478.
3. W.S. Weiglhofer, A.M. Barbosa and A.L. Topa, Special issue of Bianisotropics 2000: Guest editorial, *Arch Elektron Übertrag* **55** (2001), 209–210.

4 Refereed journal articles

1. W. Weiglhofer and W. Papousek, Skalare Hertz'sche Potentiale für anisotrope Medien, *Arch Elektron Übertrag* **39** (1985), 343–346.
2. W. Weiglhofer and W. Papousek, Scalar Hertz potentials for transversally oriented current density distributions in gyrotropic media, *Arch Elektron Übertrag* **41** (1987), 41–45.
3. W. Weiglhofer, Reduction of dyadic Green's functions to scalar Hertz potentials for gyrotropic media, *Radio Sci* **22** (1987), 209–215.

4. W. Weiglhofer, Scalarization of Maxwell's equations in general inhomogeneous bianisotropic media, *IEE Proc, Part H* **134** (1987), 357–360.
5. W. Weiglhofer, On scalar superpotentials for inhomogeneous gyrotropic media, *Arch Elektron Übertrag* **41** (1987), 371–372.
6. W. Weiglhofer, Field representation in gyrotropic media by one scalar superpotential, *IEEE Trans Antennas Propagat* **35** (1987), 1301–1302.
7. W. Weiglhofer, A formal time-domain approach to cold magnetized plasmas, *Austral J Phys* **41** (1988), 55–62.
8. W.S. Weiglhofer, Isotropic chiral media and scalar Hertz potentials, *J Phys A: Math Gen* **21** (1988), 2249–2251.
9. W. Weiglhofer, Symbolic derivation of the electrodynamic Green's tensor in an anisotropic medium, *Am J Phys* **56** (1988), 1095–1097.
10. W.S. Weiglhofer, A simple and straightforward derivation of the dyadic Green's function of an isotropic chiral medium, *Arch Elektron Übertrag* **43** (1989), 51–52.²
11. W. Weiglhofer, Delta-function identities and electromagnetic field singularities, *Am J Phys* **57** (1989), 455–456.
12. W.S. Weiglhofer, Dyadic Green's functions for general uniaxial media, *IEE Proc Microw Antennas Propagat* **137** (1990), 5–10.
13. W.S. Weiglhofer, Electromagnetic field representation in (inhomogeneous) isotropic chiral media, *Electromagnetics* **10** (1990), 271–278.
14. D.R. Fearn and W.S. Weiglhofer, Magnetic instabilities in rapidly rotating spherical geometries I. From cylinders to spheres, *Geophys Astrophys Fluid Dyn* **56** (1991), 159–181.
15. W.S. Weiglhofer, A note on the inversion of the dyadic operator $\nabla \times \nabla \times \mathbf{I} - a^2 \mathbf{I} - b \nabla \times \mathbf{I}$, *J Electromag Waves Appl* **5** (1991), 953–958.
16. D.R. Fearn and W.S. Weiglhofer, Magnetic instabilities in rapidly rotating spherical geometries II. More realistic fields and resistive instabilities, *Geophys Astrophys Fluid Dyn* **60** (1991), 275–294.
17. D.R. Fearn and W.S. Weiglhofer, Resistive instabilities and the magnetostrophic approximation, *Geophys Astrophys Fluid Dyn* **63** (1992), 111–138.
18. A. Lakhtakia and W.S. Weiglhofer, Scattering by an electrically small bianisotropic sphere in a gyroelectromagnetic uniaxial medium, *IEE Proc Microw Antennas Propagat* **139** (1992), 217–220.
19. W.S. Weiglhofer, Green's functions and magnetized ferrites, *Int J Electron* **73** (1992), 763–771.
20. D.R. Fearn and W.S. Weiglhofer, Magnetic instabilities in rapidly rotating spherical geometries III. The effect of differential rotation, *Geophys Astrophys Fluid Dyn* **67** (1992), 163–184.

²Reprinted in *Selected papers on natural optical activity* (A. Lakhtakia, ed), SPIE Press, Bellingham, WA, USA, 1990.

21. A. Lakhtakia and W.S. Weiglhofer, Maxwell–Garnett estimates of the effective properties of a general class of discrete random composites, *Acta Cryst A* **49** (1993), 266–269.³
22. W.S. Weiglhofer, Dipole radiation in uniaxial bianisotropic medium, *Electron Lett* **29** (1993), 844–845.
23. W.S. Weiglhofer, A dyadic Green’s functions representation for electrically gyrotropic media, *Arch Elektron Übertrag* **47** (1993), 125–129.
24. W.S. Weiglhofer, A. Lakhtakia and J.C. Monzon, Maxwell Garnett model for composites of electrically small uniaxial objects, *Microw Opt Technol Lett* **6** (1993), 681–684.
25. A. Lakhtakia and W.S. Weiglhofer, Axial propagation in general helicoidal bianisotropic media, *Microw Opt Technol Lett* **6** (1993), 804–806.
26. W.S. Weiglhofer, Analytic methods and free-space dyadic Green’s functions, *Radio Sci* **28** (1993), 847–857.
27. A. Lakhtakia and W.S. Weiglhofer, Axial propagation in a magnetic-dielectric cholesteric medium, *Liq Cryst* **15** (1993), 659–667.
28. W.S. Weiglhofer and A. Lakhtakia, Electromagnetic wave propagation in super-cholesteric materials parallel to the helical axis, *J Phys D: Appl Phys* **26** (1993), 2117–2122.⁴
29. W.S. Weiglhofer and I.V. Lindell, Analytic solution for the dyadic Green function of a nonreciprocal uniaxial bianisotropic medium, *Arch Elektron Übertrag* **48** (1994), 116–119.
30. W.S. Weiglhofer and I.V. Lindell, Fields and potentials in general uniaxial bianisotropic media I. Axial sources, *Int J Appl Electromag Mat* **4** (1994), 211–220.
31. W.S. Weiglhofer and A. Lakhtakia, Time-dependent scalar Beltrami–Hertz potentials in free space, *Int J Infrared Millim Waves* **15** (1994), 1015–1026.
32. W.S. Weiglhofer and A. Lakhtakia, Wave propagation in a continuously twisted biaxial dielectric medium parallel to the helical axis, *Optik* **96** (1994), 179–183.
33. W.S. Weiglhofer and A. Lakhtakia, Time-dependent Beltrami fields in free space: Dyadic Green functions and radiation potentials, *Phys Rev E* **49** (1994), 5722–5725.
34. I.V. Lindell and W.S. Weiglhofer, Green dyadic and dipole fields for a medium with anisotropic chirality, *IEE Proc Microw Antennas Propagat* **141** (1994), 211–215.
35. I.V. Lindell and W.S. Weiglhofer, Green dyadic for a uniaxial bianisotropic medium, *IEEE Trans Antennas Propagat* **42** (1994), 1013–1016.
36. W.S. Weiglhofer, Dyadic Green functions for unbounded general uniaxial bianisotropic media, *Int J Electron* **77** (1994), 105–115.

³Reprinted in *Selected papers on linear optical composite materials* (A. Lakhtakia, ed), SPIE Press, Bellingham, WA, USA, 1996.

⁴Reprinted in *Selected papers on linear optical composite materials* (A. Lakhtakia, ed), SPIE Press, Bellingham, WA, USA, 1996.

37. A. Lakhtakia and W.S. Weiglhofer, Are linear, nonreciprocal, bi-isotropic media forbidden? *IEEE Trans Microw Theory Tech* **42** (1994), 1715–1716.
38. A. Lakhtakia and W.S. Weiglhofer, Time-harmonic electromagnetic fields in source regions in a simple uniaxial bianisotropic medium, *Int J Appl Electromag Mat* **5** (1994), 101–108.
39. A. Lakhtakia and W.S. Weiglhofer, Influence of pitch on attenuation and handedness of axial propagation modes in helicoidal bianisotropic mediums, *Opt Commun* **111** (1994), 199–202; erratum: **113** (1995), 570.
40. W.S. Weiglhofer and A. Lakhtakia, Uniformity constraint on recently conceptualized linear uniaxial bianisotropic media, *Electron Lett* **30** (1994), 1656–1657.
41. W.S. Weiglhofer, Dielectric polarizability of a sphere with arbitrary anisotropy: comment, *Opt Lett* **19** (1994), 1663.
42. W.S. Weiglhofer and A. Lakhtakia, Towards the generalized supercholesteric material I. Construction and axial wave propagation, *Arch Elektron Übertrag* **48** (1994), 339–343.
43. W.S. Weiglhofer, On a medium constraint arising directly from Maxwell's equations, *J Phys A: Math Gen* **27** (1994), L871–L874.
44. A. Lakhtakia and W.S. Weiglhofer, Constraint on linear, homogeneous constitutive relations, *Phys Rev E* **50** (1994), 5017–5019.
45. A. Lakhtakia and W.S. Weiglhofer, On light propagation in helicoidal bianisotropic mediums, *Proc Roy Soc Lond A* **448** (1995), 419–437; erratum: **454** (1998), 3275.
46. W.S. Weiglhofer and A. Lakhtakia, On singularities of dyadic Green functions and long-wavelength scattering, *Electromagnetics* **15** (1995), 209–221.
47. A. Lakhtakia and W.S. Weiglhofer, Covariances and invariances of the Beltrami–Maxwell postulates, *Proc IEE Science Measure Tech* **142** (1995), 262–266.
48. W.S. Weiglhofer and A. Lakhtakia, A brief review of a new development for constitutive relations of linear bi-anisotropic media, *IEEE Antennas Propagat Mag* **37** (1995), 32–35.
49. A. Lakhtakia and W.S. Weiglhofer, On a constraint on the electromagnetic constitutive relations of nonhomogeneous linear media, *IMA J Appl Math* **54** (1995), 301–306.
50. W.S. Weiglhofer, One-dimensional dyadic Green functions for general linear bianisotropic media, *J Electromag Waves Appl* **9** (1995), 1421–1432.
51. W.S. Weiglhofer and A. Lakhtakia, Dyadic Green functions for axially varying fields in helicoidal bianisotropic media, *Int J Appl Electromag Mech* **6** (1995), 221–234.
52. A. Lakhtakia and W.S. Weiglhofer, Comment on 'Are nonreciprocal bi-isotropic media forbidden indeed?' *IEEE Trans Microw Theory Tech* **43** (1995), 2722–2723.

53. W.S. Weiglhofer and A. Lakhtakia, Causality and natural optical activity (chirality), *J Opt Soc Am A* **13** (1996), 385–386.
54. W.S. Weiglhofer and A. Lakhtakia, Perturbational solution for propagation in periodically bent nematic liquid crystals and thin films, *Microw Opt Technol Lett* **11** (1996), 320–323.
55. A. Lakhtakia and W.S. Weiglhofer, Dilute nonlinear particulate composites comprised by linear bianisotropic inclusions in nonlinear dielectric host media, *Opt Eng* **35** (1996), 1112–1118.
56. A. Lakhtakia and W.S. Weiglhofer, Lorentz covariance, Occam's razor, and a constraint on linear constitutive relations, *Phys Lett A* **213** (1996), 107–111; erratum: **222** (1996), 459.
57. W.S. Weiglhofer and A. Lakhtakia, Oblique propagation in a cholesteric liquid crystal: 4×4 matrix perturbational solution, *Optik* **101** (1996), 111–114.
58. A. Lakhtakia and W.S. Weiglhofer, Maxwell Garnett approach for nonlinear dilute particulate composites with bi-isotropic host media, *Int J Electron* **80** (1996), 665–676.
59. W.S. Weiglhofer, Fields and potentials in general uniaxial bianisotropic media II. General sources and inhomogeneities, *Int J Appl Electromag Mech* **7** (1996), 1–9.
60. A. Lakhtakia and W.S. Weiglhofer, Simple and exact analytic solution for oblique propagation in a cholesteric liquid crystal, *Microw Opt Technol Lett* **12** (1996), 245–248.
61. W.S. Weiglhofer and A. Lakhtakia, New expressions for depolarization dyadics in uniaxial dielectric-magnetic media, *Int J Infrared Millim Waves* **17** (1996), 1365–1376.
62. W.S. Weiglhofer and A. Lakhtakia, A nonlinear bianisotropic composite made with a weakly quadratic nonlinear dielectric host medium, *Arch Elektron Übertrag* **50** (1996), 285–288.
63. W.S. Weiglhofer and A. Lakhtakia, On causality requirements for material media, *Arch Elektron Übertrag* **50** (1996), 389–391.
64. W.S. Weiglhofer and A. Lakhtakia, Mediation of nonlinear polarization by the magnetic field in a composite medium with a chiral component, *Microw Opt Technol Lett* **13** (1996), 285–287.
65. A. Lakhtakia and W.S. Weiglhofer, Constraint on linear, spatiotemporally nonlocal, spatiotemporally nonhomogeneous constitutive relations, *Int J Infrared Millim Waves* **17** (1996), 1867–1878.
66. A. Lakhtakia and W.S. Weiglhofer, Further results on light propagation in helicoidal bianisotropic mediums: Oblique propagation, *Proc Roy Soc Lond A* **453** (1997), 93–105; erratum: **454** (1998), 3275.
67. A. Lakhtakia, B. Michel and W.S. Weiglhofer, The role of anisotropy in the Maxwell Garnett and Bruggeman formalisms for uniaxial particulate composite media, *J Phys D: Appl Phys* **30** (1997), 230–240.
68. A. Lakhtakia and W.S. Weiglhofer, On the application of duality to Telle-gen media, *Electromagnetics* **17** (1997), 199–204.

69. A. Lakhtakia and W.S. Weiglhofer, Green function for radiation and propagation in helicoidal bianisotropic medium, *IEE Proc Microw Antennas Propagat* **144** (1997), 57–59.
70. A. Lakhtakia, B. Michel and W.S. Weiglhofer, Bruggeman formalism for two models of uniaxial composite media: Dielectric properties, *Compos Sci Technol* **57** (1997), 185–196.
71. W.S. Weiglhofer and A. Lakhtakia, On the non-existence of linear non-reciprocal bi-isotropic (NRBI) media, *J Phys A: Math Gen* **30** (1997), 2597–2600.
72. W.S. Weiglhofer, Comment on Scattering by a biisotropic body by J. Cesar Monzon, *IEEE Trans Antennas Propagat* **45** (1997), 748–749.
73. A. Lakhtakia, B. Michel and W.S. Weiglhofer, Field induced in an electrically small dielectric particle embedded in an uniaxial dielectric host medium, *Arch Elektron Übertrag* **51** (1997), 231–232.
74. A. Lakhtakia and W.S. Weiglhofer, The Fikioris approach for the source-region electromagnetic field in a simple symmetric bianisotropic medium, *Microw Opt Technol Lett* **15** (1997), 84–86.
75. A. Lakhtakia and W.S. Weiglhofer, Time-harmonic electromagnetic field in a source region in an uniaxial dielectric-magnetic medium, *Int J Appl Electromag Mech* **8** (1997), 167–177.
76. A. Lakhtakia and W.S. Weiglhofer, On electromagnetic fields in a linear medium with gyrotropic-like magnetoelectric properties, *Microw Opt Technol Lett* **15** (1997), 168–170.
77. W.S. Weiglhofer, A. Lakhtakia and B. Michel, Maxwell Garnett and Bruggeman formalisms for particulate composite with bianisotropic host medium, *Microw Opt Technol Lett* **15** (1997), 263–266; erratum: **22** (1999), 221.
78. B. Michel and W.S. Weiglhofer, Pointwise singularity of dyadic Green function in a general bianisotropic medium, *Arch Elektron Übertrag* **51** (1997), 219–223; erratum **52** (1998), 31.
79. W.S. Weiglhofer and A. Lakhtakia, Source-region electric and magnetic fields in an uniaxial bianisotropic medium, *Electromagnetics* **17** (1997), 387–401.
80. A. Lakhtakia and W.S. Weiglhofer, Comments on Space-time and momentum representations of electromagnetic signals by P. Hillion, *Int J Appl Electromag Mech* **8** (1997), 285–287.
81. W.S. Weiglhofer and A. Lakhtakia, Further remarks on the nonexistence of linear nonreciprocal bi-isotropic (NRBI) media, *J Phys A: Math Gen* **31** (1998), 1113–1114.
82. W.S. Weiglhofer and A. Lakhtakia, Remarks on ‘Optical pulses in an optically active moving Tellegen medium,’ *J Opt* **29** (1998), 49–50.
83. A. Lakhtakia and W.S. Weiglhofer, Source-region electromagnetic field in an affinely transformable AUBM, *Int J Infrared Millim Waves* **19** (1998), 95–106.

84. A. Lakhtakia and W.S. Weiglhofer, Maxwell Garnett formalism for cubically nonlinear, gyrotropic, composite media, *Int J Electron* **84** (1998), 285–294.
85. W.S. Weiglhofer, Scalar Green functions and superpotentials of a Faraday chiral medium, *Arch Elektron Übertrag* **52** (1998), 109–112.
86. W.S. Weiglhofer and A. Lakhtakia, The correct constitutive relations of chiroplasmas and chiroferrites, *Microw Opt Technol Lett* **17** (1998), 405–408.
87. B. Michel, A. Lakhtakia and W.S. Weiglhofer, Homogenization of linear bianisotropic particulate composite media—Numerical studies, *Int J Appl Electromag Mech* **9** (1998), 167–178; erratum: **10** (1999), 537–538.
88. A. Lakhtakia and W.S. Weiglhofer, Remarks on the nonexistence of linear Tellegen media (in Russian), *Radiotekhnika i Elektronika* **43** (1998), 494–495.
89. A. Lakhtakia and W.S. Weiglhofer, Remarks on ‘The electromagnetic field in a uniformly moving crystal,’ *Pure Appl Opt* **7** (1998), 657.
90. W.S. Weiglhofer and A. Lakhtakia, The Post constraint revisited, *Arch Elektron Übertrag* **52** (1998), 276–279.
91. W.S. Weiglhofer, New expressions for depolarization dyadics in an axially uniaxial bianisotropic medium, *Int J Infrared Millim Waves* **19** (1998), 993–1005.
92. W.S. Weiglhofer, A. Lakhtakia and B. Michel, On the constitutive parameters of a chiroferrite composite medium, *Microw Opt Technol Lett* **18** (1998), 342–345.
93. W.S. Weiglhofer, Comments on ‘Green’s dyadics for bi-anisotropic media with similar medium dyadics,’ by F. Olyslager and I.V. Lindell, *Microw Opt Technol Lett* **18** (1998), 439–440.
94. W.S. Weiglhofer, Electromagnetic depolarization dyadics and elliptic integrals, *J Phys A: Math Gen* **31** (1998), 7191–7196.
95. A. Lakhtakia and W.S. Weiglhofer, Are field derivatives needed in linear constitutive relations? *Int J Infrared Millim Waves* **19** (1998), 1073–1082.
96. A. Lakhtakia and W.S. Weiglhofer, Rayleigh scattering by an infinitely long tube with helical permittivity dyadic, *J Phys D: Appl Phys* **31** (1998), 2499–2501.
97. A. Lakhtakia and W.S. Weiglhofer, Significance of cross-sectional morphology for Motohiro–Taga interfaces, *Optik* **110** (1999), 33–36.
98. W.S. Weiglhofer and S.O. Hansen, Faraday chiral media revisited—I: Fields and sources, *IEEE Trans Antennas Propagat* **47** (1999), 807–814.
99. W.S. Weiglhofer and A. Lakhtakia, On electromagnetic waves in biaxial bianisotropic media, *Electromagnetics* **19** (1999), 351–362.
100. W.S. Weiglhofer, On anomalous propagation in axially uniaxial bianisotropic mediums, *Int J Infrared Millim Waves* **20** (1999), 1277–1286.
101. W.S. Weiglhofer and A. Lakhtakia, Comment on ‘Reflection and transmission by a uniaxial bi-anisotropic slab under normal incidence of plane waves,’ *J Phys D: Appl Phys* **32** (1999), 2703–2704.

102. W.S. Weiglhofer, Electromagnetic field in the source region: A review, *Electromagnetics* **19** (1999), 563–578.
103. A. Lakhtakia and W.S. Weiglhofer, A comparative study of planewave propagation in helicoidal bianisotropic mediums and isotropic chiral mediums, *J Opt A: Pure Appl Opt* **2** (2000), 107–111.
104. A. Lakhtakia and W.S. Weiglhofer, Evanescent plane waves and the far field: Resolution of a controversy, *J Mod Opt* **47** (2000), 759–763.
105. A. Lakhtakia and W.S. Weiglhofer, Comments on ‘Comparison of long-wavelength T-matrix multiple-scattering theory and size-dependent Maxwell–Garnett formula,’ *Microw Opt Technol Lett* **25** (2000), 228–229.
106. W.S. Weiglhofer, On anomalous propagation in transversely uniaxial bianisotropic mediums, *Int J Infrared Millim Waves* **21** (2000), 895–904.
107. W.S. Weiglhofer, Comments on ‘On the constitutive relations of chiral media and Green’s dyadics for an unbounded chiral medium,’ *Microw Opt Technol Lett* **26** (2000), 277.
108. T.G. Mackay and W.S. Weiglhofer, Homogenization of biaxial composite materials: Dissipative anisotropic properties, *J Opt A: Pure Appl Opt* **2** (2000), 426–432.
109. W.S. Weiglhofer, The connection between factorization properties and closed-form solutions of certain linear dyadic differential operators, *J Phys A: Math Gen* **33** (2000), 6253–6261.
110. W.S. Weiglhofer and T.G. Mackay, Numerical studies on the constitutive parameters of a chiroplasma composite medium, *Arch Elektron Übertrag* **54** (2000), 259–265.
111. A. Lakhtakia and W.S. Weiglhofer, Maxwell Garnett formalism for weakly nonlinear, bianisotropic, dilute, particulate composite media, *Int J Electron* **87** (2000), 1401–1408.
112. T.G. Mackay, A. Lakhtakia and W.S. Weiglhofer, Strong-property-fluctuation theory for homogenization of bianisotropic composites: Formulation, *Phys Rev E* **62** (2000), 6052–6064; erratum: **63** (2001), 049901.
113. A. Lakhtakia and W.S. Weiglhofer, Electromagnetic waves in a material with simultaneous mirror-conjugated and racemic chirality characteristics, *Electromagnetics* **20** (2000), 481–488.
114. W.S. Weiglhofer, Scalar Hertz potentials for nonhomogeneous uniaxial dielectric-magnetic mediums, *Int J Appl Electromag Mech* **11** (2000), 131–140.
115. T.G. Mackay and W.S. Weiglhofer, Homogenization of biaxial composite materials: Bianisotropic properties, *J Opt A: Pure Appl Opt* **3** (2001), 45–52.
116. B. Michel, A. Lakhtakia, W.S. Weiglhofer and T.G. Mackay, Incremental and differential Maxwell Garnett formalisms for bi-anisotropic composites, *Compos Sci Technol* **61** (2001), 13–18.

117. T.G. Mackay and W.S. Weiglhofer, Homogenization of biaxial composite materials: Nondissipative dielectric properties, *Electromagnetics* **21** (2001), 15–26.
118. W.S. Weiglhofer, On the inverse homogenization problem of linear composite materials, *Microw Opt Technol Lett* **28** (2001), 421–423.
119. A. Lakhtakia and W.S. Weiglhofer, Comments on ‘Returning to the Post constraints,’ *Microw Opt Technol Lett* **29** (2001), 363.
120. T.G. Mackay, A. Lakhtakia and W.S. Weiglhofer, Ellipsoidal topology, orientation diversity and correlation length in bianisotropic composite mediums, *Arch Elektron Übertrag* **55** (2001), 243–251.
121. W.S. Weiglhofer and A. Lakhtakia, Comment on ‘Rigorous solution for transient propagation of electromagnetic waves through a medium: Causality plus diffraction in time,’ *Opt Lett* **26** (2001), 1218–1219.
122. T.G. Mackay, A. Lakhtakia and W.S. Weiglhofer, Homogenisation of similarly oriented, metallic, ellipsoidal inclusions using the bilocally-approximated strong-property-fluctuation theory, *Opt Commun* **197** (2001), 89–95.
123. A. Lakhtakia and W.S. Weiglhofer, Further comments on ‘Returning to the Post constraints,’ *Microw Opt Technol Lett* **31** (2001), 402.
124. T.G. Mackay, A. Lakhtakia and W.S. Weiglhofer, Third-order implementation and convergence of the strong-property-fluctuation theory in electromagnetic homogenisation, *Phys Rev E* **64** (2001), 066616.
125. W.S. Weiglhofer, Comment on ‘Post’s constraint for electromagnetic constitutive relations,’ *J Opt A: Pure Appl Opt* **4** (2002), 213–214.
126. W.S. Weiglhofer and T.G. Mackay, Needles and pillboxes in anisotropic mediums, *IEEE Trans Antennas Propagat* **50** (2002), 85–86.
127. T.G. Mackay, A. Lakhtakia and W.S. Weiglhofer, Homogenisation of isotropic, cubically nonlinear, composite mediums by the strong-permittivity-fluctuation theory: Third-order considerations, *Opt Commun* **204** (2002), 219–228.
128. M.W. McCall, A. Lakhtakia and W.S. Weiglhofer, The negative index of refraction demystified, *Eur J Phys* **23** (2002), 353–359.
129. W.S. Weiglhofer and A. Lakhtakia, Remarks on ‘Analytical investigation of electromagnetic waves in bianisotropic media,’ *IEE Proc Microw Antennas Propagat* **149** (2002), 138–139.
130. A. Lakhtakia, M.W. McCall and W.S. Weiglhofer, Brief overview of recent developments on negative phase-velocity materials (alias left-handed materials), *Arch Elektron Übertrag* **56** (2002), 407–410.
131. N. Georgieva and W.S. Weiglhofer, Electromagnetic vector potentials and the scalarization of sources in a nonhomogeneous medium, *Phys Rev E* **66** (2002), 046614.
132. W.S. Weiglhofer, A. Lakhtakia and M.W. McCall, Optical switching in nonlinear chiral distributed Bragg reflectors with defect layers: Comment, *J Opt Soc Am B* **19** (2002), 3042–3043.

133. W.S. Weiglhofer and N.K. Georgieva, Vector potentials and scalarization for nonhomogeneous isotropic mediums, *Electromagnetics* **23** (2003), 387–398.
134. N. Georgieva and W.S. Weiglhofer, Electromagnetic vector potentials in isotropic nonhomogeneous materials: Mode equivalence and scalarization, *IEE Proc Microw Antennas Propagat* **150** (2003), 164–170.
135. T.G. Mackay, A. Lakhtakia and W.S. Weiglhofer, The strong-property-fluctuation theory for cubically nonlinear, isotropic chiral composite mediums, *Electromagnetics* **23** (2003), 455–479.

5 Replies to comments

1. W.S. Weiglhofer and A. Lakhtakia, Response to ‘Too-tight mathematical constraints lead to loss of physics,’ by A. Sihvola, *IEEE Antennas Propagat Mag* **37** (1995), 96–99.
2. W.S. Weiglhofer and A. Lakhtakia, Response to ‘Is reciprocal discussion of non-reciprocal materials allowed by Maxwell’s equations?’ by A. Sihvola, *IEEE Antennas Propagat Mag* **37** (1995), 113.
3. W.S. Weiglhofer and A. Lakhtakia, Response to ‘Anything wrong with the naturally non-reciprocal media?’ by S. Tretyakov, *IEEE Antennas Propagat Mag* **38** (1996), 85–86.
4. W.S. Weiglhofer and A. Lakhtakia, Response to ‘Remark on the constitutive relations of linear biisotropic media’ by P. Hillion, *IEEE Antennas Propagat Mag* **38** (1996), 94–95.
5. A. Lakhtakia and W.S. Weiglhofer, Response to ‘Comment on: Electromagnetic waves in a material with simultaneous mirror-conjugated and racemic chirality characteristics’ by A.A. Sochava, C.R. Simovski and S.A. Tretyakov, *Electromagnetics* **21** (2001), 509–511.

6 Letters to the Editor

1. W.S. Weiglhofer and A. Lakhtakia, Thirty-five is enough: A new constraint for linear, bianisotropic media, *Radio Sci Bull* **273** (1995), 4.
2. W.S. Weiglhofer, What book review? *Radio Sci Bull* **280** (1997), 3.

7 Book reviews

1. J.A. Kong, *Electromagnetic wave theory*, Wiley, New York, 1986; *Arch Elektron Übertrag* **41** (1987), 127.
2. R. Forward and J. Davis, Mirror matter, *New Sci* **1656** (1989), 65.
3. A. Lakhtakia, V.K. Varadan and V.V. Varadan, *Time-harmonic electromagnetic fields in chiral media*, Springer-Verlag, Berlin, Germany, 1989; *URSI News*, December (1989), 4–5.

4. A. Lakhtakia, V.K. Varadan and V.V. Varadan, *Time-harmonic electromagnetic fields in chiral media*, Springer-Verlag, Berlin, Germany, 1989; *Arch Elektron Übertrag* **44** (1990), 420.
5. I.V. Lindell, *Methods of electromagnetic field analysis*, Clarendon Press, Oxford, UK, 1992; *Arch Elektron Übertrag* **47** (1993), 155.
6. A. Lakhtakia, *Beltrami fields in chiral media*, World Scientific, Singapore, 1994; *Arch Elektron Übertrag* **50** (1996), 53.
7. I.V. Lindell, A.H. Sihvola, S.A. Tretyakov and A.J. Viitanen, *Electromagnetic waves in chiral and bi-isotropic media*, Artech House, Boston, MA, USA, 1994; *Arch Elektron Übertrag* **50** (1996), 58, 69.
8. L.B. Felsen and N. Marcuvitz, *Radiation and scattering of waves*, IEEE Press, Piscataway, NJ, USA, 1994; *Radio Sci Bull* **276** (1996), 46.
9. A. Priou, A. Sihvola, S. Tretyakov and A. Vinogradov (eds), *Advances in complex electromagnetic materials*, Kluwer, Dordrecht, The Netherlands, 1997; *Radio Sci Bull* **285** (1998), 41–42.
10. E.J. Post, *Formal structure of electromagnetics*, Dover Press, New York, NY, USA, 1997; *IEEE Antennas Propagat Mag* **40** (1998), 92–93.
11. A. Sihvola, *Electromagnetic mixing formulas and applications*, IEE Press, London, UK, 1999; *Arch Elektron Übertrag* **54** (2000), 266.

8 Conference reports

1. W.S. Weiglhofer, Bianisotropics'97, *IEEE Antennas Propagat Mag* **39** (1997), 74–75.
2. W.S. Weiglhofer, Bianisotropics'97, *Radio Sci Bull* **282** (1997), 20–22.
3. W.S. Weiglhofer, Bianisotropics'97, *Newsletter*, The University of Glasgow, Great Britain **193** (1997), 16.
4. W.S. Weiglhofer, Bianisotropics'97, *Arch Elektron Übertrag* **51** (1997), 322–323.

9 Conference proceedings: full papers

1. W. Weiglhofer and W. Papousek, Zur Darstellung elektromagnetischer Felder in einem anisotropen Plasma durch skalare Potentiale, *Kleinheubacher Berichte* **29** (1986), 131–140.
2. W. Weiglhofer and W. Papousek, Scalar Hertz potentials and dyadic Green's functions for gyrotropic media, *Proceedings of the URSI international symposium on electromagnetic theory*, Budapest, Hungary, Akademiai Kiado (1986), 615–617.
3. W. Weiglhofer, Scalarization of Maxwell's equations in anisotropic media, *Kleinheubacher Berichte* **30** (1987), 465–474.
4. W. Weiglhofer, Scalar Hertz potentials—a new technique for the solution of microwave problems in inhomogeneous gyrotropic media, *SBMO International microwave symposium proceedings*, Rio de Janeiro, Brazil (1987), 691–696.

5. W. Weiglhofer, Field representation in warm magnetoplasmas by scalar potentials, *International symposium on radio propagation*, Beijing, China, International Information Exchange Service Company, Kowloon, Hong Kong (1988), 417–420.
6. W. Weiglhofer, Delta-function identities in electromagnetic radiation problems, *Kleinheubacher Berichte* **31** (1988), 347–353.
7. W.S. Weiglhofer, Inversion of dyadic differential operators with simple matrix methods, *Kleinheubacher Berichte* **32** (1989), 239–246.
8. W.S. Weiglhofer, A general inversion technique for a class of dyadic differential operators of electromagnetic theory, *Proceedings of the URSI international symposium on electromagnetic theory*, Royal Institute of Technology, Stockholm, Sweden (1989), 10–12.
9. W.S. Weiglhofer, Chiral media: New developments in an old field, *Proceedings of the URSI international symposium on electromagnetic theory*, Royal Institute of Technology, Stockholm, Sweden (1989), 271–273.
10. W.S. Weiglhofer, Field representation techniques for cold and warm magnetized plasmas, *Kleinheubacher Berichte* **33** (1990), 75–83.
11. W.S. Weiglhofer, Theory and applications of chiral media, *Kleinheubacher Berichte* **33** (1990), 287–296.
12. A. Lakhtakia and W.S. Weiglhofer, Bianisotropic sphere in a uniaxial medium with application to complex composites, *Proceedings of the URSI international symposium on electromagnetic theory*, Sydney, Australia (1992), 149–151.
13. W.S. Weiglhofer, Analytic techniques for Green's functions, *Proceedings of the URSI international symposium on electromagnetic theory*, Sydney, Australia (1992), 237–239.
14. W.S. Weiglhofer, MHD instabilities in rapidly rotating systems, *Current topics in astrophysical and fusion plasma research, Proceedings of the international workshop on plasma physics* (M.F. Heyn, W. Kernbichler and H.K. Biernat, eds), Pichl/Schladming, Austria (1993), 5–10.
15. I.V. Lindell and W.S. Weiglhofer, Analytic solutions for fields in certain anisotropic chiral media, *Proceedings of Chiral'94, 3rd international workshop on chiral, bi-isotropic and bi-anisotropic media* (F. Mariotte and J.-P. Parneix, eds), Perigueux, France (1994), 131–135.
16. W.S. Weiglhofer and A. Lakhtakia, Exact solutions of Maxwell's equations for propagation in helicoidal bianisotropic media, *Proceedings of Chiral'94, 3rd international workshop on chiral, bi-isotropic and bi-anisotropic media* (F. Mariotte and J.-P. Parneix, eds), Perigueux, France (1994), 155–159.
17. W.S. Weiglhofer and A. Lakhtakia, Dyadic Green functions, radiation and scalar potentials in the time-dependent Beltrami formulation for free space, *Proceedings of Chiral'94, 3rd international workshop on chiral, bi-isotropic and bi-anisotropic media* (F. Mariotte and J.-P. Parneix, eds), Perigueux, France (1994), 335–340.

18. A. Lakhtakia and W.S. Weiglhofer, On the causal constitutive relations of magnetoelectric media, *Proceedings of the 1995 IEEE international symposium on electromagnetic compatiblity*, Atlanta, Georgia (1995), 611–613.
19. A. Lakhtakia and W.S. Weiglhofer, Source-region electromagnetic field in a uniaxial dielectric-magnetic medium, *Asia-Pacific microwave conference*, New Delhi, India (1996), 217–220.
20. W.S. Weiglhofer, Bianisotropy and bianisotropics'97—a perspective, *Proceedings of Bianisotropics'97, international conference and workshop on electromagnetics of complex media* (W.S. Weiglhofer, ed), Glasgow, UK (1997), 23–26.
21. W.S. Weiglhofer and A. Lakhtakia, Homogenization of cubically nonlinear, gyrotropic, composite media, *Proceedings of Bianisotropics'97, International conference and workshop on electromagnetics of complex media* (W.S. Weiglhofer, ed), Glasgow, UK (1997), 111–114.
22. W.S. Weiglhofer, Electromagnetic fields in source regions of bianisotropic media, *Proceedings of Bianisotropics'97, International conference and workshop on electromagnetics of complex media* (W.S. Weiglhofer, ed), Glasgow, UK (1997), 169–172.
23. W.S. Weiglhofer, Dyadic Green functions in complex media, singularities and applications, *Proceedings of the URSI international symposium on electromagnetic theory*, Thessaloniki, Greece (1998), 211–213.
24. B. Michel, W.S. Weiglhofer and A. Lakhtakia, Homogenization of bianisotropic particulate composites, *Proceedings of the URSI international symposium on electromagnetic theory*, Thessaloniki, Greece (1998), 220–222.
25. W.S. Weiglhofer, A. Lakhtakia and B. Michel, On chiroplasmas and chiroferrites, *Proceedings of Bianisotropics '98, 7th international conference on complex media* (A.F. Jacob and J. Reinert, eds), Braunschweig, Germany (1998), 9–12.
26. W.S. Weiglhofer, Canonical problems for radiation and propagation in complex media, *Proceedings of Bianisotropics '98, 7th international conference on complex media* (A.F. Jacob and J. Reinert, eds), Braunschweig, Germany (1998), 215–218.
27. W.S. Weiglhofer and A. Lakhtakia, The Post constraint revisited, *Proceedings of Bianisotropics '98, 7th international conference on complex media* (A.F. Jacob and J. Reinert, eds), Braunschweig, Germany (1998), 241–244.
28. W.S. Weiglhofer and B. Michel, Explicit expressions for depolarization dyadics of biaxial anisotropic media, *Proceedings of Bianisotropics '98, 7th international conference on complex media* (A.F. Jacob and J. Reinert, eds), Braunschweig, Germany (1998), 245–248.
29. W.S. Weiglhofer, Electromagnetic theory of complex materials, *Engineered Nanostructural Films and Materials, Proceedings of SPIE international symposium on optical science, engineering, and instrumentation*

- (A. Lakhtakia and R.F. Messier, eds), Denver, Colorado, USA, SPIE Volume 3790 (1999), 66–76.
30. W.S. Weiglhofer, Homogenization of particulate materials, *Complex Mediums, Proceedings of SPIE international symposium on optical science, engineering, and instrumentation* (A. Lakhtakia, W.S. Weiglhofer and R.F. Messier, eds), San Diego, California, USA, SPIE Volume 4097 (2000), 146–154.
 31. W.S. Weiglhofer and A. Lakhtakia, Waves and fields: From uniaxial to biaxial mediums, in between and beyond, *Proceedings of Bianisotropics 2000, 8th international conference on electromagnetics of complex media* (A.M. Barbosa and A.L. Topa, eds), Lisbon, Portugal (2000), 3–6.
 32. A. Lakhtakia, M.N. Lakhtakia and W.S. Weiglhofer, Homogenization formalisms for nonlinear particulate composite mediums, *Proceedings of Bianisotropics 2000, 8th international conference on electromagnetics of complex media* (A.M. Barbosa and A.L. Topa, eds), Lisbon, Portugal (2000), 23–26.
 33. B. Michel, A. Lakhtakia, W.S. Weiglhofer and T.G. Mackay, The incremental and differential Maxwell Garnett formalisms for bianisotropic composites, *Proceedings of Bianisotropics 2000, 8th international conference on electromagnetics of complex media* (A.M. Barbosa and A.L. Topa, eds), Lisbon, Portugal (2000), 27–30.
 34. T.G. Mackay, A. Lakhtakia and W.S. Weiglhofer, Strong-property-fluctuation theory for homogenization of bianisotropic composites, *Proceedings of Bianisotropics 2000, 8th international conference on electromagnetics of complex media* (A.M. Barbosa and A.L. Topa, eds), Lisbon, Portugal (2000), 31–34.
 35. W.S. Weiglhofer, Hertz potentials in complex medium electromagnetics, *Proceedings of Bianisotropics 2000, 8th international conference on electromagnetics of complex media* (A.M. Barbosa and A.L. Topa, eds), Lisbon, Portugal (2000), 107–110.
 36. T.G. Mackay and W.S. Weiglhofer, Numerical homogenization studies of biaxial bianisotropic composite materials, *Proceedings of Bianisotropics 2000, 8th international conference on electromagnetics of complex media* (A.M. Barbosa and A.L. Topa, eds), Lisbon, Portugal (2000), 237–240.
 37. W.S. Weiglhofer, Depolarization dyadics, *Proceedings of Bianisotropics 2000, 8th international conference on electromagnetics of complex media* (A.M. Barbosa and A.L. Topa, eds), Lisbon, Portugal (2000), 415–420.
 38. W.S. Weiglhofer, A. Lakhtakia and T.G. Mackay, Propagation and radiation in a material combining mirror conjugation and racemic chirality, *Proceedings of the URSI international symposium on electromagnetic theory*, Victoria, Canada (2001), 193–195.
 39. T.G. Mackay, A. Lakhtakia and W.S. Weiglhofer, Homogenization of bianisotropic composites via strong-property-fluctuation theory, *Proceedings of the URSI international symposium on electromagnetic theory*, Victoria, Canada (2001), 377–379.

40. T.G. Mackay, A. Lakhtakia and W.S. Weiglhofer, The strong-property-fluctuation theory applied to the homogenisation of linear bianisotropic composites, *Complex mediums II: Beyond linear isotropic dielectrics* (A. Lakhtakia, W.S. Weiglhofer and I.J. Hodgkinson, eds), Proc SPIE **4467** (2001), 243–255.
41. W.S. Weiglhofer, Constitutive relations, *Complex mediums III: Beyond linear isotropic dielectrics* (A. Lakhtakia, G. Dewar and M.W. McCall, eds), Proc SPIE **4806** (2002), 67–80.
42. T.G. Mackay, A. Lakhtakia and W.S. Weiglhofer, Trilocally approximated strong-permittivity-fluctuation theory for nonlinear composite mediums, *Complex mediums III: Beyond linear isotropic dielectrics* (A. Lakhtakia, G. Dewar and M.W. McCall, eds), Proc SPIE **4806** (2002), 100–111.
43. A. Lakhtakia, M.W. McCall, W.S. Weiglhofer, J. Gerardin and J. Wang, On mediums with negative phase velocity: A brief overview, *Complex mediums III: Beyond linear isotropic dielectrics* (A. Lakhtakia, G. Dewar and M.W. McCall, eds), Proc SPIE **4806** (2002), 176–180.

10 Conference proceedings: abstracts and extended abstracts

1. W.S. Weiglhofer, Analytic field representations in inhomogeneous isotropic chiral media, *XXIII General assembly of URSI*, Prague, Czechoslovakia **2** (1990), 409.
2. W.S. Weiglhofer, Symbolic methods versus scalarization techniques, *Kleinheubacher Berichte* **34** (1991), 291–292.
3. W.S. Weiglhofer, Representation of fields in Faraday chiral media, *Progress in electromagnetics research symposium 1993*, Pasadena, CA, USA (1993), 2P5/1.
4. I.V. Lindell and W.S. Weiglhofer, Construction of the Green dyadic for axially chiral uniaxial media, *Proceedings of Chiral'94, 3rd international workshop on chiral, bi-isotropic and bi-anisotropic media* (F. Mariotte and J.-P. Parneix, eds), Perigueux, France (1994), 303.
5. W.S. Weiglhofer and A. Lakhtakia, Electromagnetic fields in source regions of anisotropic media, *1994 URSI radio science meeting abstracts*, Seattle, WA, USA (1994), 79.
6. W.S. Weiglhofer and A. Lakhtakia, Supercholesteric materials, *1994 URSI radio science meeting abstracts*, Seattle, WA, USA (1994), 356.
7. I.V. Lindell and W.S. Weiglhofer, Green dyadic of a uniaxial chiral medium, *1994 URSI radio science meeting abstracts*, Seattle, WA, USA (1994), 424.
8. W.S. Weiglhofer and A. Lakhtakia, Optical propagation in helicoidal bianisotropic media, *1994 URSI radio science meeting abstracts*, Seattle, WA, USA (1994), 426.

9. W.S. Weiglhofer and I.V. Lindell, Scalar potential formalism for uniaxial bianisotropic media, *1994 IEEE antennas and propagation society international symposium digest*, Seattle, WA, USA (1994), 1586–1589.
10. W.S. Weiglhofer, On factorization of differential operators and analytic solutions in electromagnetic theory, *1994 IEEE antennas and propagation society international symposium digest*, Seattle, WA, USA (1994), 1598–1601.
11. W.S. Weiglhofer, Closed-form expressions for one-dimensional dyadic Green functions of general homogeneous and some inhomogeneous linear bianisotropic media, *Progress in electromagnetics research symposium*, Seattle, WA, USA (1995), 164.
12. W.S. Weiglhofer, S.F. Nagle and A. Lakhtakia, Axial propagation in helicoidal bianisotropic media: dependence of mode handedness on pitch, *Progress in electromagnetics research symposium*, Seattle, WA, USA (1995), 288.
13. W.S. Weiglhofer and A. Lakhtakia, Linear constitutive relations and Maxwell's equations: the case for mathematical consistency, *Progress in electromagnetics research symposium*, Seattle, WA, USA (1995), 1103.
14. A. Lakhtakia and W.S. Weiglhofer, Constraint on the constitutive relations of a general linear, spatiotemporally nonlocal medium, *1996 URSI radio science meeting abstracts*, Baltimore, MD, USA (1996), 205.
15. A. Lakhtakia and W.S. Weiglhofer, Exact analytic solution for oblique propagation in helicoidal bianisotropic mediums, *1996 URSI radio science meeting abstracts*, Baltimore, MD, USA (1996), 206.
16. W.S. Weiglhofer and A. Lakhtakia, Dilute nonlinear particulate composites, *XXV General assembly of URSI*, Lille, France (1996), 36.
17. W.S. Weiglhofer and A. Lakhtakia, On constitutive relations for linear materials, *XXV General assembly of URSI*, Lille, France (1996), 40.
18. W.S. Weiglhofer and A. Lakhtakia, Wave propagation in helicoidal bianisotropic media, *XXV General assembly of URSI*, Lille, France (1996), 82.
19. A. Lakhtakia and W.S. Weiglhofer, The Fikioris approach for source-region fields in an uniaxial dielectric-magnetic medium, *1997 URSI radio science meeting abstracts*, Montreal, Canada (1997), 92.
20. A. Lakhtakia, B. Michel and W.S. Weiglhofer, Bruggeman formalism for composite media with aligned short fibers, *1997 URSI radio science meeting abstracts*, Montreal, Canada (1997), 142.
21. W.S. Weiglhofer, Constitutive relations, *NATO Advanced Research Workshop Bianisotropics 2002, 9th international conference on electromagnetics of complex media* (S. Zouhdi and M. Arsalane, eds), Marrakech, Morocco (2002), 26.
22. T.G. Mackay, A. Lakhtakia and W.S. Weiglhofer, Homogenisation via the strong-property-fluctuation theory: Third-order implementation and convergence, *NATO advanced research workshop Bianisotropics 2002, 9th in-*

- ternational conference on electromagnetics of complex media* (S. Zouhdi and M. Arsalane, eds), Marrakesh, Morocco (2002), 28.
23. A. Lakhtakia and W.S. Weiglhofer, Promiscuity-promoting perambulations in the polarizabilities of poly-isotropic spheres, *NATO advanced research workshop Bianisotropics 2002, 9th international conference on electromagnetics of complex media* (S. Zouhdi and M. Arsalane, eds), Marrakesh, Morocco (2002), 63.

Index

A

A waves, 253–254
 aberration, 357
 ablation, 560, 562
 absorbers, saturable, 107
 absorption, 126, 356, 677, 680
 – negative, 552
 add-drop filter, 387
 address pulse, 582
 Aharonov–Bohm integral, 13
 Aizu species, 179
 aluminum, 438, 441
 analysis of variance (ANOVA), 594
 anapole (moment), 187
 angle of rise, 483
 angle selectivity, 427, 442
 angular frequency, 403
 anholonomic references, 12
 anisotropic medium, xxvi, 29, 37–38, 51, 246, 255, 425, 612
 anisotropy, 229–230, 236–239, 550, 677–678
 anisotropy energy density, 228–230, 236, 238
 antimony tin oxide (ATO), 420, 435
 aperture in a screen, 401
 armchair carbon nanotube, 510, 512
 asymmetric magnetoimpedance (AMI), 275
 atomic force microscope (AFM), 560, 562
 attenuation, 353
 autocorrelation, 134, 136
 axial conductivity of a carbon nanotube, 515
 axial excitation, 488
 axial propagation, 479–480, 486, 491–492
 axial propagation in CSTFs, 480
 axial tensor, 650
 axial vector, 8

B

B waves, 253–254
 backscattering, enhanced, 551
 backward medium, 357
 ballistic aggregation, 451–452
 BaMnF₄, 660
 band diagram, 370
 band structure, 111
 band-to-band transitions in spontaneously ordered semiconductors, 677, 692
 bandgap, 106

bandwidth, 484, 490
 barium titanate, 122
 Bateman, H., 19
 beam combiner, 501
 beam fanning, 123
 beam router/combiner, 500
 beam, optical, 551
 beamsplitter, 464
 beta distribution, 604
 bianisotropic medium, xxvi, xxvii, 29, 37, 43, 51, 349, 458, 467, 481, 679
 bianisotropy, 319
 biaxial medium, 45, 50–51, 454
 bifurcation, 574
 bifurcation diagram, 370
 biisotropic medium, 51
 bilocal approximation, 326
 binary diagram, 370
 biochip, 466
 bioluminescence, 466
 biophysics, xxvii
 biopolymer, 661
 birefringence, 109, 129, 454, 677, 679
 – linear, 110
 BisA, 125
 bisector angle, 132
 bismuth silicon oxide (BSO), 123
 bismuth tin oxide, 122
 bispherical analysis, 438
 bistability, 389, 415, 574
 Bloch waves, 379
 Boltzmann constant, 128
 Born, M., 8
 boundary conditions, 404
 bounded modes, 374
 Boys–Post representation, 42, 51, 54
 Bragg filter, 358, 370, 479–481 483–485, 492, 497, 552
 Bragg phenomenon, 461, 480, 496
 – circular, 461, 464. *See also* circular Bragg phenomenon.
 Bragg regime, 462–463, 467, 480, 490, 492, 494, 497, 501
 Bragg wavelength, 492–494, 496, 501–502
 Brillouin zone, 112, 379, 511
 Bruggeman, 322, 432, 436
 Bruggeman formalism, 53–54, 318, 451
 Busse balloon, 579

C

C_{60} , 125
 C_{153} , 125
 Cartan, E., 10, 12, 22
 causality, xxvii, 36, 350
 cavity, 385
 cavity solitons, 581
 center of symmetry, 8
 characteristic functions, 320
 characteristic waves, 256
 charge density, electric, 31, 348
 charge density, electron, 37
 charge density, magnetic, 30
 charge generation (CG), 124
 charge-transport (CT), 122
 chiral liquid crystals, 697
 chiral medium, Faraday, 53
 chiral medium, isotropic, 45, 51, 53, 358
 chiral morphology, 480
 chiral sculptured thin film (CSTF), 479–481, 483–504
 chirality, xxvi, 550
 – index, 667
 – parameter, 45
 chiroferrite, 54
 chiroplasma, 54
 chirping, 463
 cholesteric liquid crystal, 480
 Christiansen filter, 563
 circular birefringence, 68, 645–646
 circular Bragg phenomenon (CBP), 480, 489–490, 493, 502. *See also* Bragg phenomenon, circular
 circular dichroism, 69, 358, 645–646
 cluster, 425, 440–441, 452
 coherence, 550, 554
 coherence length, 557
 coherent light propagation, 684
 cold plasma, 248
 collimation, 563
 collisional cooling, 561
 columnar morphology, 448–449
 columnar nanostructure, 422
 columnar thin film, 449, 457
 comparison medium, 325
 compensational charge, 128
 complex field structure, 634
 composite material, xxv, 295–298, 300–301, 308–309, 311–312, 350, 398
 – ring-wire, 354
 – unusual, 358
 Condon, E.U., 21
 conduction band, 370
 conduction current, 128
 conductivity, 398
 conformal group, 19
 constitutive relations, 4–6, 8, 29, 32, 41, 246, 248, 250–251, 255, 319, 349, 458, 481–482, 731
 – Boys–Post, 42–43, 51
 – constraints on, 34–35, 50–52
 – Tellegen, 42, 44, 51, 458
 constitutive tensor, 19

continuity condition, 32
 continuum, xxvi, 448, 458, 461
 – electromagnetic theory, 64
 – nonhomogeneous, 449
 convolution, 35–36, 39, 43
 correlation, 133
 – edge-enhanced, 133
 – all-optical, 134
 – edge-enhanced, 134
 – joint transform, 134
 correlation length, 326
 coupled-wave theory, 460, 479–480, 484, 491–492, 501, 504
 coupler, 3dB, 501
 coupling, 600
 covariance function, two-point, 327
 covariance function, three-point, 327
 cross-anisotropy films, 268
 cross-correlation, 134, 136
 cross-polarized fields, 493
 cross-polarized reflectivities, 492, 495
 crosstalk voltage, 597
 crystallinity, 448
 crystallographic, 110
 Cunningham, E., 19
 current density, electric, 31, 348
 current density, electron, 37
 current density, magnetic, 30
 curvilinear coordinates, 12
 cutoff frequency, 249, 255

D

damping, 39
 DANS/MMA, 127
 dark current, 130
 de Rham, G., 16
 de Rham cohomology, 18
 defect, 372
 defect modes, 385
 defocusing, 132
 degenerate four-wave mixing (DFWM), 133
 degradation, 467
 DEH, 125, 127
 ΔE effect, 239–240
 demagnetization energy, 228, 232, 236, 242
 dendrimer, 149, 155
 density anisotropy, 453
 depolarization, 563
 depolarization dyadic, 321, 731
 depolarizing optical systems, 682
 deposition, pulsed laser, 559–560
 Dexter exchange mechanism, 148
 dielectric displacement, 31, 458
 dielectric medium, xxvi, 35, 39
 dielectric thin film, 480
 dielectric-magnetic medium, 35–36, 45, 50, 350
 Difféo hierarchy, 23
 Difféo(3), 12
 Difféo(4), 10–11, 23
 diffraction, 571
 diffuser, 549, 563, 565
 – designer, 563

diffusion, 112, 122, 128
 diffusion equation, 552
 dipole, 556
 dipole-dipole interaction, 145–147
 directional coupler, 499–501
 dispersion, xxvii, 29, 249, 404
 – spatial, 38
 – temporal, 38, 461
 dispersion law for graphene, 511
 dispersion law for nanotubes, 512
 dispersion relation, 257, 374
 displays, 461, 465
 dissipation, xxvii, 50, 350, 355
 dissipative system, 577
 domain average engineering, 94
 Doppler effect, 246
 double negative medium, 357
 drift, 122
 Drude model, 701
 Drude–Born–Fedorov relations, 72
 duality, 30
 dyads, 30
 dyadics, 30, 41, 459
 – rotation, 459
 dyadic Green function, 321, 326, 533
 dynamic measurements, 236, 238–239, 241
 Dyson equation, 326

E

ECZ, 124
 edge effects in nanotubes, 520
 edge enhancement, 131, 134
 – real-time, 131
 edge scattering pattern, 523
 effective boundary conditions for carbon nanotubes, 517
 effective medium, 301, 423–424
 eigenpolarization, 109, 612, 615
 – p and s polarized, 685
 Einstein, A., 13
 Einstein–de Hass effect, 242
 elastic constants, 227–228
 elastic energy density, 225, 227, 235, 238
 elastic waves in solids, 612
 elastodynamics, 466, 468
 electric field, 31, 458
 electro-optic effect, 122, 129
 electroluminescence, 467
 electromagnetic compatibility (EMC), 592
 electromagnetic composite, 295, 306–307, 310
 electromagnetic field, xxvi
 electromagnetic field quantization, 532
 electromagnetic metamaterials, 306, 311
 electron, 106
 electron confinement, 466
 electron localization, 398
 electron recombination rate, 128
 electron velocity, 37
 electron-beam evaporation, 562
 electronic chip, 465
 electrostrictive medium, 330
 electrotoroidic effect, 187

ellipsoidal particle, 320
 ellipsometry data, 66, 689, 694
 ellipticity, 66
 emission, amplified spontaneous (ASE), 553, 555, 558
 emission, pulse, 553
 emission, stimulated, 552
 emittance, 434
 enantiomorphism, 21, 23
 energy conservation, 262, 353
 energy exchange, 130
 energy harvesting, 141
 energy pooling, 142, 145, 151
 energy-efficient paint, 426
 environmental stability, 467
 epitaxial growth, 112
 equilibrium Fermi distribution function, 515
 equilibrium, lack of, 550
 Euler equations, 37
 evanescent, 370
 Ewald sphere, 129
 exact forms, 18
 exchange energy, 228
 exciton, 116, 148, 158
 experiment design, 593
 exterior differentiation, 18
 extinction position, 655
 extraordinary light transmittance, 397, 405, 409, 413

F

Fabry–Pérot resonator, 494, 496, 552
 fanning, 127, 130
 – scattering coefficient, 130
 far zone, 356
 Faraday, M., 12, 22, 197, 203
 Faraday chiral medium, 53–54
 Faraday rotation, 117
 fast Fourier transform (FFT), 132
 FDEA-MNST, 125
 finite-difference time domain (FDTD), 246, 263
 feedback, nonresonant, 552
 feedback, positive, 408
 feedback, resonant, 552
 ferrite, 53
 ferroelasticity, 178–181
 ferroelectricity, 178–181
 ferroics, 181–182
 ferromagnetism, 178–181
 ferromagnetic resonance (FMR), 237–238
 ferromagnetoelectricity, 169, 180
 ferrotoroidics, 180
 fiber Bragg gratings, 482, 484
 – filters, 498
 fiber, optical, 551
 field enhancement, 397
 figure of merit (FOM), 123, 298–300, 303–304, 306, 309, 311
 finite-sized device, 380
 fixed point, 573
 flash ionization, 249
 Floquet–Bloch theorem, 368

Föppl, A., 11
 Förster radius, 147
 Fourier series, 129, 405
 Fourier transform, 108
 – combined, 41
 – spatial, 40
 – temporal, 31, 35, 43, 349
 free space, 28, 348, 350, 454
 frequency shift, 246, 262
 frequency transformer, 246, 254
 frequency, emission, 552
 Fresnel reflections, 481, 494, 563
 Fresnel zone, 356
 Fresnel–Fizeau, 8
 FRET, 145
 fullerene, 510

G

Galilei, G., 20
 gallium arsenide, 106, 123
 Gamow vector, 384–385
 gas sensor, 495
 Gaussian units, 400
 generalized ellipsometry, 677
 Generalized Ohm's Law (*see also* Ohm's Law), 430, 441
 Gibbs, J.W., 11
 Ginzburg–Landau equation, 573
 Giorgi rationalization, 9–10
 glass, 550
 – ground, 563
 – opal, 563
 gold, 434, 559–560
 Gothic symbols, 7
 Graf's formula, 381–382
 granular metals, 440–441
 graphene, 511
 gratings, 126, 380
 – absorption, 126
 – higher-order, 127, 130
 – primary, 127
 Green function, 381
 – dyadic, 49
 group velocity, 356
 gyration tensor, 648
 gyrofrequency, 38, 255–256, 261
 gyrotropic medium, 39, 45, 50, 53

H

Hamilton–Jacobi, 12
 handed-reflector, 490
 handedness, 66, 357, 462, 466
 handedness-inverter, 463
 harmonic generation, 113
 Heaviside function, 327
 Heisenberg, W.K., 12
 Heisenberg uncertainty principle, 19
 helical morphology, 487
 helicoidal bianisotropic medium, 47, 489
 – thin-film, 457, 459
 helicon mode, 258
 helix, 664

Helmholtz equation, 129, 381, 479, 481–482
 Hermitian properties, 21–22
 hexagon, 576
 high-accuracy universal polarimeter (HAUP), 645, 648
 high-order harmonics, 524
 higher-order generation, 130
 histogram, 593
 hole, 110
 homogeneous medium, 482, 487, 489, 495
 homogenization, xxvii, 53, 347, 355, 461, 731
 – local, 461
 hopping, 122

I

impair (forms), 18
 impedance, intrinsic, free-space, 350
 incoherent light propagation, 686
 incommensurate crystal, 648
 index matching, 495
 indium tin oxide (ITO), 127, 435
 induced chiral inhomogeneities, 634
 induced electrostatic field, 128
 induced refractive index change, 129
 induction fields, 31, 42, 348
 information processing, 571
 initial value problem, 259, 261
 instability (in roll solutions), 579
 insulator, 448
 integrated optics and telecommunications, 495
 intensity, 128
 – dark, 128
 interaction, resonant, 113
 interbeam angle, 132
 interference, 557
 interferometry, 551
 interlayer dielectric, 460, 466
 intersecting spheres, 438
 intramolecular relaxation, 145
 inversion symmetry, 21
 ion beam, 560
 ion bombardment, 450
 ion thruster, 468
 ionization front, 249
 isolator, 197
 isotropic chiral medium, 64, 328
 isotropic constitutive equations, 5
 isotropic elastic medium, 615, 621
 isotropic medium, 45, 350, 483, 498
 isotropic plasma, 249, 251, 254–258

J

Jacobi technique, 323
 Jones matrix, 656, 678
 Joule effect, 224

K

Kennard, E.H., 20
 Kerr effect, 116, 389, 403–404, 408, 415, 578
 Kottler, F., 10, 12, 22
 Kramers–Kronig relations, 36, 70
 kriging, 593, 599–600

L

LaB₆, 435
 Lambertian scatter, 563
 Laplace operator, 40
 laser, 106, 408, 551–552
 – random, 553–554, 557
 – polymer, 555
 – powder, 549, 553
 – ring, 555
 – self-organized, 552
 laser ablation, 560
 laser mirror, 463
 Laurent series, 384
 law of mixtures, 298–301, 308, 310
 layered systems, 267
 left-handed material (misnomer), 349, 357
 Levi-Civita unit tensors, 15, 17
 light-induced modulation, 404
 linear birefringence, 68, 110, 645–646
 linear dichroism, 69, 646
 liquid crystal, 449, 465, 482
 lithium niobate, 122
 lithography, 560
 localization, 556
 – electron, 398
 – plasmon, 398
 localized fields, 372, 609, 624, 632
 localized structures, 571
 long-wavelength approximation, 326
 longitudinal propagation, 255–257, 578
 Lorentz, H.A., 19
 Lorentz covariance, xxvi
 Lorentz force, 37
 Lorentz model, 352, 356
 Lorentz plasma, 248–249
 Lorentz reciprocity, 34, 50
 Lorentz–Heaviside electromagnetics, xxvi
 losslessness, 50
 low dimensionality, 295, 296, 299, 301, 303–305
 luminescence, 557
 lysozyme, 663

M

macroscopic field, 348
 macroscopic material, 297
 magnetic anisotropy, 267
 magnetic energy density, 225, 228
 magnetic field, 31, 458
 – quasi-static, 37
 magnetic induction, 31, 458
 magnetic medium, 35
 magnetic point groups, 80
 magnetic sensor, 267
 magnetization, 47, 348
 magneto-optic effect, linear, 182, 187
 magneto-optics, 38, 49, 197
 magnetoelastic energy density, 224–225, 231, 233–235, 238
 magnetoelastic parameter, 231, 233–234
 magnetoelasticity, 223–231, 233–242
 magnetoelectric coefficients
 – tensor form, bilinear effects, 173

– tensor form, linear effect, 170–171
 magnetoelectric dyadics, 43
 magnetoelectric effects, 48, 167, 458, 467
 – applications of, 181–188
 – bilinear, 172, 176
 – induced, 172, 176
 – linear, 172
 – ME_H and ME_E effects, 177
 – measurement, 177
 – measuring units, 173
 – spontaneous, 178
 – theory, 188
 – toroidal contributions, 185
 magnetoimpedance, 267
 magnetoionic theory, 256, 259
 magnetoplasma medium, 246, 255
 magnetostriction, 223–231, 233–242
 magnetostriction parameter, 231, 234, 237
 magnetostrictive strain, 231, 233
 magnetotoroidic effect, 187
 manufacturing, 468
 mass operator, 326
 MathCad, 132
 matrix differential equation, 460
 matrix polynomial series, 460
 Matteucci effect, 240
 Maxwell, J.C., 5, 28
 Maxwell equations, 30, 32, 49, 55, 348, 400, 449, 459, 481, 486, 556
 Maxwell Garnett formalism, 53, 318, 322, 409, 432, 436, 438, 550
 – differential, 323–324
 – incremental, 323
 mechanical loading, 467
 medium, 28
 mesomaterial, 296
 metal, 403, 448, 559
 metal shells, 441
 metal vapor, 559
 metal-dielectric film, 398
 metallic film, 397, 415
 metamaterial, 106, 295–301, 303–304, 306, 310, 312–313
 metaparticle, 301, 306–307, 311
 micelle, 559
 microcavity, 578
 microscopic field, 348
 Mie, G., 425, 430, 441
 Minkowski, H., 7, 10
 mirrored pairs, 21
 mobility, 128
 mode, coherent, 556, 558
 modelocking, 107
 modulational instabilities, 580
 molecular dynamics, 450
 momentum, 112
 – conservation of, 404
 monoclinic crystal, 648
 Monte Carlo sampling, 551, 593
 Moore's law, 551
 morphology, xxv, 447–448, 452, 458, 468
 – helicoidal, 456–457, 466

- matchstick, 452
- nematic, 455
- c surface, 559
- moth-eye coating, 495
- Mueller matrix, 678
- multipass filtering, 479
- multichromophore array, 155–156. *See also* dendrimers
- multilayer filter, 499
- multiply transmissive filter, 497
- multipole, 147

N

- NAN, 125
- nanocrystal, 559–560
- nanoengineering, xxv, xxvii, 47, 448
- nanohole, 413, 415, 422, 430
 - light-circuiting in, 413
- nanohole array, 404, 409
- nanoholes in metal, 440
- nanoparticle, 422, 429, 434
- nanoscale imaging, 427
- nanostar, 149
- nanowaveguide, 520
- narrow tapered grooves, 430
- narrow-band filtering, 479, 496–497
- NAT, 125
- natural optical activity, 21–22
- near field, 556
- near-field scanning optical microscope, 414
- negative differential conductivity, 528
- negative group-velocity medium, 357
- negative phase-velocity medium, 347, 349, 357–358
- negative-index medium, 357
- nematic morphology, 455
- Neumann, K.G., 9
- Neumann principle, 21, 23
- NNDN, 125
- noise current, 533
- noise, immunity to, 549–550
- Noll, W., 13
- non-Bragg order, 131
- nondepolarizing optical systems, 682
- nonhomogeneity, xxvi, xxvii, 29, 43, 458, 461, 480
- nonlinear chromophore, 129
- nonlinear dynamics, 550
- nonlinear medium, 47
- nonlinear optics, 106, 403, 551, 571
- nonlinearity, xxvi, xxvii
- nonlinearity enhancement, 335
- nonlocality, 40, 43
- nonreciprocal medium, 50–51
- normal modes, 256
- nucleation, 350

O

- O wave, 256
- observable, 593
- Ohm's Law, generalized, 397, 399, 414
- one-dimensional photonic crystals, 367
- opal glass, 563

- optical activity, 45, 68, 456, 645, 647
- optical bandwidth, 495
- optical bistability, 415
- optical cavity, 578
- optical constants, 677
- optical filter, 398, 461, 464, 467
- optical homogeneity, 64
- optical indicatrix, 648, 651
- optical interconnect, 465
- optical properties, uniformization of, 550, 563
- optical rotation, 68, 75, 457, 651
- optical switching, 398, 415
- optical system, 677, 682
- optoelectronic network, 495
- orientability, 23
- orientational enhancement, 129
- orthonormal beams, 609, 611, 617, 621, 623, 628–629
- orthorhombic crystal, 453
- oscillation, relaxation, 553
- Oseen transformation, 486

P

- p-dci, 125
- pair, 18
- pattern formation, 571
- PBPES, 125
- percolation, 398, 409
- percolation threshold, 398, 415, 560
- perfect lens, 347, 357
- permeability, 34–35, 398
 - relative, 35, 355
 - – negative real, 349, 356
- permittivity, 34–35, 398
 - relative, 35, 354
 - – negative real, 349, 356
- permittivity dyadic, 43
- permittivity modulation, 405
- permittivity tensor, 678–680
- phase discontinuity, 497
- phase grating, 483
- phase retarder, 455
- phase-velocity, 351, 404
 - negative, 351, 353
- phase-conjugation, mutual, 551
- phasor, 32, 41, 350, 459
- phonon mode, 677, 692
- photoactive material, 141
- photoconductor, 122
- photoelasticity, 620
- photon, 550, 558
- photon statistics, 552
- photonic bandgap material, chiral, 464
- photonic bandgap, 370
- photonic bandgap material, 461
- photonic crystals, 366
 - nonreciprocal magnetic, 182
- photorefractive polymer, 122–123, 551
 - fully functionalized, 126
- photosensitive center, 128
- photosynthesis, 142–143
- photovoltaic effect, 122, 142

- physical vapor deposition, 447–448, 455
 - piezoelectricity, 461
 - piezomagnetolectric effect, 178
 - piezotoroidic effect, 187
 - pitchfork, 574
 - pixel, 584
 - plane wave, 109, 350, 353, 399–400, 610, 616, 618, 621, 627
 - plaser (powder laser), 550
 - plasma, 37, 40, 246, 253–254
 - plasma current, 250, 252, 262
 - plasma frequency, 38, 249, 252–253, 255, 433
 - plasma, magnetized, 28
 - cold, 38, 54
 - warm, 41
 - plasmon localization, 398
 - PMMA, 125
 - PNA, 125
 - PNP, 124
 - point group, 80
 - symbols, 96
 - Poisson's ratio, 234
 - polar tensor, 649
 - polar vector, 8
 - polarimeter, 645, 647
 - polarizability dyadic, 321–322
 - polarization, 108, 563
 - circular, 462
 - degree of, 563
 - electric, 36, 47, 348
 - magnetic, 36. *See also* magnetization
 - polarization crosstalk, 490
 - polarization filters, 455, 463, 467, 499
 - polarization management, 495
 - polarization multiplexing, 496
 - polarization routing, 479, 499
 - polarization sensitivity, 495
 - polarization-maintaining optical fiber, 495
 - polarization-sensitive filter, 492
 - polarization-specific filter, 498
 - polarizer, 455
 - polarizing beamsplitter, 499
 - pole, 384–385
 - poly-L-lactic acid, 661
 - polymer, 124, 440, 555, 557
 - population, statistical, 592
 - porosity, 449, 466
 - positive phase-velocity medium, 358
 - Post constraint, 51–52, 54
 - potential, scalar, 402
 - powder laser, 549, 553
 - power intensification, 245, 261–263
 - Poynting vector, 383
 - time-averaged, 42, 347, 350, 354
 - precipitation, arrested, 559
 - preprocessing, 134
 - pressure, electron, 37
 - pressure, radiation, 354
 - primitive fields, 31, 42, 349
 - principal value, 36
 - probability density function, 593
 - propagation, 109, 370
 - guided-wave, 110, 117
 - – paraxial, 117
 - property tensor, 82
 - protein, 647
 - pseudo-isotropy, 466
 - pseudoscalar, 20–21
 - pseudoscalar-valued integral, 10
 - pulse detection, 106
 - pulse generation, 106
 - pulse modulation, 106
 - pulse propagation, 106
 - pulse bleeding, 465
 - pulse shaper, 465
 - pulse shaping, 107
 - Purcell effect, 537
 - PVK, 124
- ## Q
- quantum dot, 107, 550, 562
 - quantum electrodynamics (QED), 145, 148, 152, 157
 - quantum mechanics, 28, 49
 - quantum well, 105
 - semiconductor, 105
 - multiple, 112
 - quantum-well intermixing, 105
 - quasi-normal modes, 384
 - quasi-patterns, 577
 - quasi-phase matching, 105
 - quasi-phase matching grating, 105
- ## R
- R* wave, 256, 257
 - Raman spectroscopy, 415, 549, 559, 562
 - surface-enhanced, 559
 - random cell, 430, 436
 - random laser, 553–554, 557
 - liquid, 554
 - random material, 398
 - random preorganization, 550
 - randomness, 549–550, 559, 564
 - rare-earth material, 151
 - Rayleigh scattering, 403
 - Rayleigh–Wood anomaly, 464
 - reciprocity, 34, 50, 52
 - reflection, 8, 353, 375, 376, 402, 461, 681
 - reflector, 455, 464
 - refraction, 353
 - refractive index, 122, 258–259, 261, 349, 401
 - complex, 350–351
 - negative, 349, 355
 - – condition for, 351
 - regression, 594
 - relative permittivity tensor, 483
 - residue operator, 384
 - resonance, 111, 353, 385, 397, 404, 408, 411
 - resonance energy transfer, 145
 - resonance linewidth, 352
 - resonance wavelength, 352
 - resonant frequency, 255
 - resonant interaction, 117
 - resonant mode, 385

resonator, stochastic, 558
 Rham cohomology, 10
 Ricci-Curbastro, G. (Ricci), 12
 Riemann–Christoffel tensor, 10, 15, 20
 Riemannian world, 13
 ringlaser, 22
 rugate filter, 464

S

Sagnac effect, 8, 20
 sample, statistical, 592
 saturable absorber, 578
 scalar Bragg grating, 497
 scanning electron microscope (SEM), 564
 scatter, Lambertian, 563
 scattering, 356, 367, 398, 400
 – multiple, 555
 – random, 551
 scattering losses, 326
 scattering matrix, 380, 382, 384
 Schouten, J.A., 8, 18
 Schrödinger, E., 12
 Schultz, A.K., 22
 sculptured nematic thin film, 457, 459, 464
 sculptured solid-state dielectric films, 699
 sculptured thin film, 47, 448, 456, 480, 482, 502.
 See also chiral sculptured thin film
 – chiral, 456, 459, 461
 – slanted, 464
 second harmonic, 105, 467
 self-assembly, 559
 self-conjugation, 551, 558
 self-focusing, 107
 self-organization, 549–551, 556, 564, 585
 self-shadowing, 450, 452, 454
 semiconductor, 49, 106, 448, 578
 – compound, 113
 sensor, fluid, 449, 460, 464
 sensor, optical, 461
 separatrix, 582
 shape dyadic, 320
 sieve, 467
 silicon, 450
 silver, 408, 429, 559
 silver thiogallate, 669
 simple medium, 34
 simulation, 580
 single feedback mirror, 577
 skin effect, 404
 slow-wave coefficient, 519
 slowly varying amplitude approximation, 110
 Snoek's relation, 305
 solar cell, 449
 solar energy, 142, 425
 solar selective absorbers, 427, 433
 soliton, spatial, 107, 197, 571
 Šolc filter, 464, 495
 Sommerfeld, A., 12
 sound waves in liquids, 615
 SP resonance, 436
 space-charge field, 122, 127
 space-guide, 465
 spacer layers, 498
 spatial discontinuity, 246–247
 spatial soliton, 107, 197, 571
 spatial symmetry, 571
 speckle, 555, 563
 spectral hole filter, 464
 spectroscopic gradient, 144–145, 150
 spectroscopy, surface-enhanced Raman, 398, 415
 spinor, 23
 splashing effect, 561
 symmetry-breaking, 576
 spontaneous radiation, 533, 539
 sputtering, 457
 SR(3), 8–9, 23
 stability, 573
 static measurements, 236–237, 239, 241
 statistics, 591
 stimulated emission, 553
 stochastic resonance, 550
 strain, 225–226, 233
 Stranski-Krastanow growth, 108
 stress, 228, 232, 235
 stripes, 576
 strong-property-fluctuation theory (SPFT), 53, 325–336
 strontium barium niobate (SBN), 123
 Structure Zone Model, 453
 subcritical bifurcation, 574
 substrate, 559
 – fabrication of, 560
 subwavelength holes, 397, 409
 sudden collapse, 261–262
 sudden creation, 249, 251, 258
 super-exchange interaction, 148
 supercriticaln bifurcation, 574
 superfluorescence, 554, 558
 superlattice, 105–106
 – semiconductor, 105
 surface current density, 514
 surface impedance tensor, 271
 surface plasmon, 425, 428
 surface plasmon polariton (SPP), 397, 403–404, 412, 415, 425, 430
 surface waves, 518
 susceptibility, 646
 – electric, 35
 – linear, 110
 – magnetic, 35
 – nonlinear, 48, 111, 408
 SVEA, 129
 switching, 106
 symmetry, crystallographic, 113
 symmetry, cubic, 113
 symmetry breaking, 112, 576
 synchronization, 558
 systematic errors, 653

T

tantalum oxide, 455, 462
 Tellegen representation, 42, 51, 54. *See also* constitutive relations, Tellegen
 temporal discontinuity, 246–248, 252–253, 258

tensor, 30
 tensor distinction, domain pair, 86
 tensor distinction, domains, 83
 tensor distinction, global, 85
 tensor invariants, 92
 Terfenol-D, 224, 237, 240
 thermal barrier, 467
 thermoelectric material, 304
 thin film, 267
 third harmonic, 116, 526
 three-layer film, 269
 tight-binding approximation, 511, 513
 time-varying medium, 245–246
 titanium oxide, 455, 462
 TNF, 124, 125
 topological concepts, 10
 toroidal moments, spontaneous, 180, 185–187
 toroidic effects, 187
 touching particles, 438, 440
 transfer matrix, 686
 transition energy, interband, 116
 transition frequency, 112
 – interband, 112
 transmission, 355, 461
 – light-induced, 408
 – resonant, 404, 415
 transmission matrix, 368, 373
 transmittance, 375, 402
 – extraordinary light. *See* extraordinary light transmittance
 – nonresonant, 406
 transverse propagation, 256
 trilocal approximation, 327
 Truesdell, C.A., 13
 tunability, 461
 tunneling, 106, 122, 253
 twinning group, 88
 – completely transposable, 90
 two-beam coupling, 123, 127, 130
 two-dimensional photonic crystals, 378
 two-wave approximation, 400

U

ultrasonic applications, 466
 ultrasonic diffraction gratings, 636
 uniaxial anisotropy, 45, 50, 228–229, 231, 324, 454, 467

uniformization, 550, 563
 upconversion, 152, 155

V

vacuum, 28, 34, 454
 van Dantzig, D., 10, 12, 22
 vapor flux, 450
 vector analysis, 8
 vibration ellipse, 65
 Villari effect, 239–240
 virtual photon, 146, 157
 virus entrapment, 467
 void, 398, 452, 456
 Voigt, W., 16, 197, 199
 Voigt notation, 226
 volume fraction, 321
 volume magnetostriction, 230, 234

W

wave mixing, 127
 wave propagation, 108, 459, 731
 waveguide, 107
 – planar, 116
 – rib-loaded, 110
 wavelength division multiplex, 106
 wavelength, free-space, 352
 wavenumber, 403
 – free-space, 350
 wavepacket, 376
 waveplate, 456
 wavevector, 110
 whistler mode, 258, 261
 Wiedemann effect, 240–242
 wiggler magnetic field, 245, 252, 260, 262

X

X wave, 256

Y

Young's modulus, 234, 239–240

Z

Zeeman energy, 228, 236
 zigzag carbon nanotube, 510, 512
 zinc oxide, 555
 zirconium oxide, 455