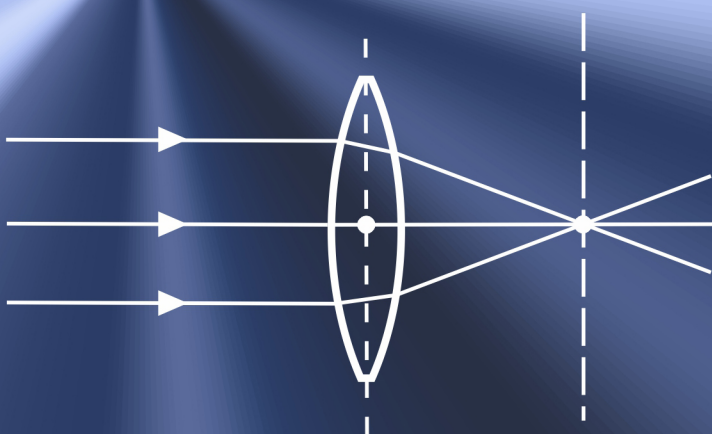


Sergey Y. Yurish
Editor

Advances in OPTICS

Reviews



1

Advances in Optics: Reviews

Book Series, Volume 1

Sergey Y. Yurish
Editor

Advances in Optics: Reviews

Book Series, Volume 1



International Frequency Sensor Association Publishing

Sergey Y. Yurish
Editor

Advances in Optics: Reviews
Book Series, Vol. 1

Published by International Frequency Sensor Association (IFSA) Publishing, S. L., 2018
E-mail (for print book orders and customer service enquires): ifsa.books@sensorsportal.com

Visit our Home Page on <http://www.sensorsportal.com>

Advances in Optics: Reviews, Vol. 1 is an open access book which means that all content is freely available without charge to the user or his/her institution. Users are allowed to read, download, copy, distribute, print, search, or link to the full texts of the articles, or use them for any other lawful purpose, without asking prior permission from the publisher or the authors. This is in accordance with the BOAI definition of open access.

Neither the authors nor International Frequency Sensor Association Publishing accept any responsibility or liability for loss or damage occasioned to any person or property through using the material, instructions, methods or ideas contained herein, or acting or refraining from acting as a result of such use.

ISBN: 978-84-697-9435-7,
e-ISBN: 978-84-697-9436-4
BN-20180415-XX
BIC: TTB

Acknowledgments

As Editor I would like to express my undying gratitude to all authors, editorial staff, reviewers and others who actively participated in this book. We want also to express our gratitude to all their families, friends and colleagues for their help and understanding.

Contents

Contents.....	7
Contributors.....	15
Preface	19
 1. Multilayer Mirrors for Attosecond Pulses	 21
1.1. Introduction	21
1.1.1. Attosecond Pulses (Isolated and Pulse Trains)	22
1.1.2. HHG as Major/Only Attosecond Pulse Source	22
1.1.3. What Is It Good for? The Applications	23
1.2. Reflective Optics in the EUV/Soft X-Ray Range	24
1.2.1. Interaction of Radiation with Matter	24
1.2.2. Single Interface Reflectance (and Transmission)	26
1.2.3. Multilayer Mirrors	27
1.2.3.1. Principal of Multilayer Mirrors	28
1.2.3.2. Periodic and Aperiodic Multilayer Systems	29
1.2.3.3. Imperfect Interfaces	32
1.3. Simulations of Multilayer Systems - Their Theoretical Optical Performance	34
1.4. Fabrication of Multilayer Coatings	36
1.5. Achieved Results and Examples for Attosecond Multilayer Mirrors	37
1.5.1. A Normal Incidence Broadband 30-60 eV Mirror	38
1.5.2. Short Pulses in the 60-100 eV Range	39
1.5.3. Above the Silicon L-Edge in the 100 eV – 150 eV Photon Range	40
1.5.3.1. Attosecond Dispersion Control by Multilayer Mirrors above 100 eV	40
1.5.3.2. Chromium/Scandium Multilayer Mirror for Attosecond Pulses at 145 eV	43
1.5.4. Multilayer Mirrors for Attosecond Pulses in the Water Window Soft X-Ray Range	47
1.5.4.1. Aperiodic Multilayer Mirrors	47
1.6. Metrology and Application	50
1.6.1. Streaking Measurements	50
1.6.2. Transient Absorption Measurements	51
1.6.3. The RABBITT Technique	51
1.7. Outlook	52
1.7.1. The Way Toward Ever Shorter Pulses - Approaching the Atomic Unit of Time	52
1.7.2. Multilayer Mirrors for HHG Photon Energies up to the keV Spectral Range	53
1.8. Conclusions	54
Acknowledgements	54
References	54
 2. Moiré Effect in 3D Structures	 61
2.1. Introduction	61
2.2. Moiré Effect around Us	64
2.3. Plain Coplanar Case	69
2.3.1. Projected Mesh	69
2.3.2. Moiré Wavevector	72
2.3.3. Spectral Trajectories	76

2.4. Moiré in Regular 3D Bodies	78
2.4.1. Parallelepipied (Parallel Planes).....	78
2.4.2. Prism (Wedge)	80
2.4.3. Cylinder	82
2.5. Interesting Moiré Issues.....	85
2.5.1. Moiré Statistics	85
2.5.2. Moiré Lens	87
2.5.3. Square and Octagon.....	88
2.6. Conclusion	90
References	90
3. Numerical Method for Diffraction by Multilayered Dielectric Gratings Using Scattering Factors.....	95
3.1. Introduction.....	95
3.2. Matrix Eigenvalues Method.....	97
3.3. Application of Shadow Theory to All Regions.....	100
3.3.1. In the Case of Uniform Regions	100
3.3.2. In the Case of Periodic Regions.....	104
3.4. An Excitation Source and Boundary Conditions	106
3.5. Diffraction Efficiencies and Optical Theorem	108
3.6. Analysis of Low Grazing Scattering by Dielectric Gratings.....	112
3.7. Analysis of Eigenvalues Degeneracy in the Middle Region	114
3.8. Conclusions.....	117
References	118
4. An Alternative Model of the Spatial Light Intensity Distribution	121
4.1. Introduction.....	121
4.2. Motivation and Related Previous Work.....	122
4.3. Mathematical Model	124
4.4. Experiment – a Comparison of Two Error Measures	127
4.4.1. The Algorithms Used in Previous Studies	128
4.4.2. The Algorithms IF-RMS and IF-MAX.....	130
4.4.3. Setup	131
4.4.4. Results.....	131
4.5. Conclusion	134
References	136
5. Optical Beams in Linear and Nonlinear Media	137
5.1. Introduction.....	137
5.2. Nonparaxial Beams in Linear Media	138
5.2.1. Nonparaxial Radially Polarized Beams	138
5.2.2. Nonparaxial Elegant Hermite-Laguerre Gaussian Beams.....	142
5.2.3. Nonparaxial Pearcey Gaussian Beams	146
5.2.4. Nonparaxial Parabolic Rotational Coordinate Beams	147
5.3. Paraxial Beams in Nonlocal Nonlinear Media.....	148
5.3.1. Ince-Gaussian Breathers and Solitons.....	150
5.3.2. Complex Variable Function Breathers and Solitons.....	152

5.4. Conclusions	154
Acknowledgements	155
References	155
6. Incoherently Coupled Soliton Families in Photorefractive Media.....	161
6.1. Introduction	161
6.2. Theoretical Model	163
6.2.1. <i>Dynamical Evolution Equation for Soliton families in Centrosymmetric Photorefractive Material</i>	163
6.2.2. <i>Modulation Instability</i>	167
6.2.3. <i>Dynamical Evolution Equation for Soliton Pairs in aPyroelectric Photorefractive Material</i>	169
6.3. Discussion	171
6.3.1. <i>Coupled Solitons in Centrosymmetric Photorefractive Material</i>	171
6.3.1.1. Dark-Dark Soliton Pair	171
6.3.1.2. Bright-Bright Soliton Pair	175
6.3.1.3. Grey-Grey Soliton Pair	178
6.3.1.4.2N-Component Dark Solitons	182
6.3.1.5.2N-Component Bright Solitons	183
6.3.1.6. N Component Grey Solitons.....	184
6.3.1.7. Modulation Instability	186
6.3.2. <i>Coupled Solitons in Pyroelectric Photorefractive Materials</i>	186
6.3.2.1. Bright Soliton Pair	186
6.3.2.2. Dark Soliton Pair	188
6.3.2.3. Grey Soliton Pair	191
6.3.3. <i>Bright-Dark Soliton Pair</i>	193
6.3.4. <i>Observation of Separate Components of Incoherently Coupled Solitons</i>	196
6.4. Conclusions	197
Acknowledgement.....	198
References	199
7. Vectorial Complex Ray Model for Light Scattering of Nonspherical Particles	203
7.1. Introduction	203
7.2. Fundamentals of Geometrical Optics	204
7.2.1. <i>Snell Laws and Fresnel Formulas</i>	205
7.2.2. <i>Light Scattering by a Sphere and a Circular Cylinder</i>	206
7.2.2.1. Deviation of Rays on Particle Surface	207
7.2.2.2. Amplitudes of Reflected and Refracted Rays	208
7.2.2.3. Phases of Rays.....	208
7.2.2.4. Scattering of an Infinite Circular Cylinder	210
7.2.2.5. Scattering of a Sphere.....	211
7.2.3. <i>Comparison of Scattering Diagrams with Lorenz - Mie Theory</i>	212
7.3. Vectorial Complex Ray Model.....	213
7.3.1. <i>Snell Law and Fresnel Formulas in Vector Form</i>	214
7.3.2. <i>Wave Front Equation</i>	215
7.3.3. <i>Amplitude and Phase of a Ray</i>	216
7.3.3.1. Amplitude.....	217
7.3.3.2. Phase	219
7.3.4. <i>Simple Applications of the Wave Front Equation</i>	220
7.3.4.1. Image Formation by a Plane Diopter	221

7.3.4.2. Image Formation by a Spherical Diopter	222
7.3.4.3. Divergence Factor of a Circular Cylinder	223
7.3.4.4. Divergence Factor of a Sphere	223
7.4. Applications of VCRM in Light Scattering	224
7.4.1. Revisit of Airy Theory in Term of VCRM	224
7.4.2. Scattering by an Elliptical Cylinder	225
7.4.3. Scattering of the Plane Wave by an Ellipsoidal Particle	227
7.4.4. Software VCRMell2D	227
7.4.5. Hyperbolic Umbilic Foci of an Oblate Particle and Experimental Validation	227
7.4.6. Dependence of Two Rainbow Intensity Ratio on the Aspect Ratio of a Prolate Particle	229
7.5. Conclusions	230
Acknowledgements	231
References	231
8. Magneto-Optical Effects Arising from Coupling of Magnetic and Dielectric Properties for Colloidal Particle System	233
8.1. Introduction	233
8.2. Theoretical Framework	234
8.2.1. Origin of Magneto-Optical Effects for Colloidal Particle System	234
8.2.2. Magneto-Dielectric Properties of Crystals and Colloids	235
8.2.3. Magneto-Optical Birefringence and Dichroism in Colloids	239
8.3. Experiment Description and Results	241
8.3.1. Sample Description	241
8.3.2. The Optical Experiments and Results	242
8.3.2.1. The Optical Effect of Light Beam Perpendicular to Magnetic Field	242
8.3.2.2. The Optical Effect of Light Beam Parallel to Magnetic Field	244
8.4. Conclusions	247
Acknowledgement	247
References	248
9. General Overview of Coherent X-Ray Diffraction Imaging and Ptychography and Their Developments and Applications	251
9.1. Coherent X-Ray Diffraction Imaging	251
9.1.1. Overview and Introduction	251
9.1.2. Bragg CXDI	253
9.2. Ptychography	257
9.2.1. Overview and Introduction	257
9.2.2. Forward Ptychography	258
9.2.3. Bragg Ptychography	259
9.2.3.1. Bragg Ptychography on Single Crystals	259
9.2.3.2. Bragg Ptychography on Thin Film Phase Domains	259
9.2.4. Partial Coherence and Multimodes in CXDI Diffraction Imaging	260
9.2.4.1. Transverse Partial Coherence	260
9.2.4.2. Longitudinal Partial Coherence	262
9.2.5. Resolution Limit and Inverse 4 th Power Law (Flux and Resolution)	262
9.3. Future Developments Aspects of Single-Shot Coherent Diffraction Imaging	264
9.3.1. Diffract-and-Destroy Serial Femtosecond Nanocrystallography in 4 th Generation Free Electron Laser Facilities	264
9.3.2. Single-Shot Coherent Modulation Imaging (CMI) for Materials Sciences Application	265

9.4. Conclusion and Future Outlooks.....	265
Acknowledgements	265
References	265
10. The Optical Anderson Localization in Three-Dimensional Percolation System	269
10.1. Introduction	269
10.2. Basic Equations	271
10.3. Lasing in Percolating System with Incorporated Emitters	273
10.3.1. <i>The Case of Small Time</i>	278
10.3.2. <i>Numerical 3D Simulations</i>	278
10.4. Condition of Optical Localization in Percolating System	282
10.5. Inverse Participation Ratio	284
10.6. Localization for Large Time	288
10.7. Discussion	298
10.8. Conclusions	298
References	298
11. A Highly Directional Supercontinuum in the Visible upon Filamentation in Air.....	301
11.1. Introduction	301
11.2. Experimental Setup and Research Methods	303
11.3. Experimental Results and Discussion	303
11.4. Conclusion	313
References	314
12. Study of Third-Harmonic Generation at Interfaces Taking into Account the Contribution of Self-Focusing Effect.....	317
12.1. Introduction	317
12.2. Basic Theory of Nonlinear Polarization and Nonlinear Refractive Index	319
12.3. THG under Influence of SF Effect	321
12.4. THG at Optical Glasses Interfaces	324
12.4.1. <i>Experiments</i>	325
12.4.2. <i>Results and Discussion</i>	326
12.5. THG at the Interfaces of a Cuvette Filled with Organic Solvents	330
12.5.1. <i>Experiments</i>	331
12.5.2. <i>Results and Discussion</i>	332
12.6. THG at Organic Solvents Interfaces as a Function of Pulse Duration	338
12.6.1. <i>Result and Discussion</i>	339
12.7. Conclusions	339
Acknowledgements	341
References	341
13. Transformations and Evolution of Phase Singularities in Diffracted Optical Vortices.....	345
13.1. Introduction	345

13.2. Experimental Setup.....	347
13.3. Description of the Diffraction Model.....	350
13.3.1. General Principles of the Singular Skeleton Analysis	350
13.3.2. Description of the Incident OV Beams	351
13.3.3. Migration of Singularities in the Diffracted OV Beams: Experimental Data Compared with Theory	353
13.4. Mathematical Model of the Singular Skeleton Evolution in Diffracted OV Beams	357
13.4.1. Asymptotic Analytical Model	357
13.4.2. Refined Analytical Model.....	358
13.5. Theoretical Study of the OV Migration: LG Beams.....	361
13.5.1. OV Displacements: Incident Beam with Plane Wavefront	362
13.5.2. OV Displacements: Incident Beam with Spherical Wavefront	364
13.5.3. Incident LG Beam with the Second-Order OV.....	368
13.6. Discontinuities of the OV Trajectories and Topological Reactions in the Diffracted OV Beams	370
13.6.1. The 'Jump' Description: Kummer Beams	370
13.6.2. Discontinuities in the Laguerre-Gaussian Beams' Diffraction	374
13.7. OV Jumps in the z-Dependent Singular Skeleton Evolution	375
13.7.1. Kummer Beams	376
13.7.2. Laguerre-Gaussian Beams.....	376
13.7.3. 3D Trajectories and the Nature of Discontinuities.....	378
13.8. Conclusion	380
Acknowledgements.....	382
References	382
Appendix	385
14. Wavefront Reconstruction with Rotational Fields	391
14.1. Introduction.....	391
14.2. Theoretical Framework.....	393
14.2.1. Two-Dimensional Signal Processing	393
14.2.2. Tree Structure and Two-Dimensional Quadrature Mirror Filters	395
14.2.3. Wavelets in Tree Structure and Factoring Wavelets	396
14.2.4. Relationship of Wavelets to Shack-Hartmann Measurements.....	397
14.3. Phase Reconstruction Algorithm	398
14.3.1. Iteration for Level 1	398
14.3.2. Iteration for Level 2	399
14.3.3. Further Iterations.....	401
14.3.4. Setting the Mean and Waffle Values.....	401
14.3.5. Synthesis Section.....	402
14.4. Rotational Phase Fields.....	403
14.4.1. Continuous Vector Fields.....	403
14.4.2. Discrete Vector Fields.....	403
14.5. Conclusion	407
References	407
Appendix. Further Proof Details.....	409
A. High-Order Wavelet Simplification Proof.....	409
B. Iteration for Level k.....	410
C. Branch Point Boundary Condition Proof.....	412

15. Fresnel Nearfield Space-Grating Optics in the Human Retina Explains Human Color and Dimlight Vision	415
15.1. The Gap in the Physics of Colors	415
15.2. Human Color and Dimlight Vision	416
15.3. What a Space Grating Optical Explanation of Vision Ought To Be Able To Achieve	417
15.4. The Development of the Cortical Retina to Become a Three-Layered Cell Body Grating	418
15.5. The Layered Processing of Information in the Cortical Visual Centers	420
15.6. The Calculation of the Achievements of the ONL Space Grating of the Retina	421
15.7. The Space Grating Optical Explanation of the Purkinje-Shift: RGB(B) in Dimlight Vision	430
15.8. The Space Grating Chromatic Adaptivity in Daylight Vision	431
15.9. The Aperture Angle of the Light Cones of the Diffracted Light	431
15.10. The Retinal Space Grating Becomes a ‘Living Crystal’	432
15.11. The Individual and the Whole: The Relativization of Color of the Local on to the Color of the Global	433
15.12. Summary and Outlook	433
References	435
 16. Characteristics of Unbalanced Mach-Zehnder Interferometers in Metal/Insulator/Metal Plasmonic Waveguides	437
16.1. Introduction	437
16.2. Design and Analysis of the Unbalanced MZI	438
16.3. Fabrication and Experimental Results in Straight MIM PWGs	440
16.4. Fabrication and Experimental Results in MZI Based on MIM PWGs	443
16.5. Conclusion	447
References	447
 17. Stable Solitons of Higher Order Cubic Quintic Nonlinear Schrödinger Equation with a \mathcal{PT}-Symmetric Potential	451
17.1. Introduction	451
17.2. \mathcal{PT} -Symmetric Optical Lattices	453
17.3. Numerical Solution of Higher Order (1+1)D CQNLS Equation	455
17.4. Linear Stability Analysis via Linear Spectrum	457
17.5. Nonlinear Stability Analysis	462
17.6. Cubic Nonlinear Schrödinger Equation with Fourth Order Dispersion	463
17.7. Cubic Quintic Nonlinear Schrödinger Equation with Fourth Order Dispersion	464
17.8. Conclusion	470
References	471
 Index	473

Contributors

O. Angelsky

Chernivtsi National University, Kotsyubinsky Str. 2, Chernivtsi 58012, Ukraine

Travis W. Axtell

Department of the Navy, USA

İlkay Bakırtaş

Department of Mathematics, Istanbul Technical University, Maslak 34469, Istanbul, Turkey, ilkayb@itu.edu.tr

Emerson Cristiano Barbano

São Carlos Institute of Physics, PO Box 369, 13560-970, São Carlos, SP, Brazil

A. Bekshaev

Odessa I. I. Mechnikov National University, Dvorianska 2, Odessa 65082, Ukraine

Gennadiy Burlak

CIICAp, Universidad Autónoma del Estado de Morelos, Av. Universidad 1001, Cuernavaca, Morelos, 62210, México

Roberto Cristi

Naval Postgraduate School, Monterey, California, USA

Dongmei Deng

Guangdong Provincial Key Laboratory of Nanophotonic Functional Materials and Devices, South China Normal University, Guangzhou 510631, China

Alfredo Díaz-de-Anda

Instituto de Física, Benemérita Universidad Autónoma de Puebla, P.O. Box J-48, 72570 Puebla, Puebla, México

İzzet Göksel

Department of Mathematics, Istanbul Technical University, Maslak 34469, Istanbul, Turkey

Alexander Guggenmos

Faculty of physics, Ludwig-Maximilians-University Munich, Am Coulombwall 1, D-85748 Garching, Germany

Max Planck Institute of Quantum Optics, Hans-Kopfermann-Str. 1, D-85748 Garching, Germany

University of California Berkeley, Department of Chemistry, Berkeley, CA-94720, USA

Qi Guo

Guangdong Provincial Key Laboratory of Nanophotonic Functional Materials and Devices, South China Normal University, Guangzhou 510631, China, guoq@scnu.edu.cn

S. G. Hanson

DTU Fotonik, Department of Photonics Engineering, DK-4000 Roskilde, Denmark

Masanobu Haraguchi

Haraguchi Laboratory, Tokushima University, Japan

Ivanov N.G.

Institute of High Current Electronics SB RAS, Akademichesky ave. 2/3, 634055, Tomsk, Russia

David Kaljun

FME, University of Ljubljana, Aškerčeva 6, 1000 Ljubljana, Slovenia

Shun Kamada

Haraguchi Laboratory, Tokushima University, Japan

Aavishkar Katti

Department of Physics, Banasthali University, Rajasthan, India

Department of Physics, Institute of Science, Banaras Hindu University, U.P, India

Ulf Kleineberg

Faculty of physics, Ludwig-Maximilians-University Munich, Am Coulombwall 1, D-85748 Garching, Germany

Max Planck Institute of Quantum Optics, Hans-Kopfermann-Str. 1, D-85748 Garching, Germany

Norbert Lauinger

Institut für Optosensorik, Wetzlar (Germany)

Jian Li

School of Physics Science and Technology, Southwest University, PR China

Losev V. F.

Institute of High Current Electronics SB RAS, Akademichesky ave. 2/3, 634055, Tomsk, Russia

National Research Tomsk Polytechnic University, Lenin Avenue 30, 634050, Tomsk, Russia

Erika Martínez-Sanchez

Facultad de Ingeniería, Universidad Autónoma de Coahuila, Blvd. Fundadores Km. 13 Ciudad Universitaria, Arteaga, Coahuila 25354, México.

Xiangshen Meng

School of Physics Science and Technology, Southwest University, PR China

Lino Misoguti

São Carlos Institute of Physics, PO Box 369, 13560-970, São Carlos, SP, Brazil

Toshihiro Okamoto

Haraguchi Laboratory, Tokushima University, Japan

Prokop'ev V. E.

Institute of High Current Electronics SB RAS, Akademichesky ave. 2/3, 634055,
Tomsk, Russia

Kuan Fang Ren

CORIA / UMR 6614, CNRS – INSA & Université de Rouen Normandie, 76801
Saint Etienne du Rouvray, France

Claude Rozé

CORIA / UMR 6614, CNRS – INSA & Université de Rouen Normandie, 76801
Saint Etienne du Rouvray, France

Vladimir Saveljev

Department of Physics, Myongji University, 116 Myongji-ro, Cheoin-gu, Yongin,
Gyeonggi-do, 17058, Korea

Xiaowen Shi

Diamond Light Source, Oxfordshire, United Kingdom, xiaowen.shi@cantab.net

Sitnik K. A.

Institute of High Current Electronics SB RAS, Akademichesky ave. 2/3, 634055,
Tomsk, Russia

Hideaki Wakabayashi

Faculty of Computer Science and Systems Engineering, Okayama Prefectural
University, Soja, Okayama, 719–1197 Japan

R. A. Yadav

Department of Physics, Institute of Science, Banaras Hindu University, U.P, India

Álvaro Zamudio-Lara

CIICAp, Universidad Autónoma del Estado de Morelos, Av. Universidad 1001,
Cuernavaca, Morelos, 62210, México

Janez Žerovnik

FME, University of Ljubljana, Aškerčeva 6, 1000 Ljubljana, Slovenia

Institute of Mathematics, Physics and Mechanics, Jadranska 19, Ljubljana, Slovenia

Sergio Carlos Zilio

São Carlos Institute of Physics, PO Box 369, 13560-970, São Carlos, SP, Brazil

Preface

It is my great pleasure to introduce the first volume of new Book Series '*Advances in Optics: Reviews*' started by the IFSA Publishing in 2018. Three volumes were published in this year.

The '*Advances in Optics: Reviews*' Book Series is published as an Open Access Books in order to significantly increase the reach and impact of these volumes, which also published in two formats: electronic (pdf) with full-color illustrations and print (paperback).

The first three volumes of this Book Series has organized by topics of high interest. In order to offer a fast and easy reading of each topic, every chapter in this book is independent and self-contained. All chapters have the same structure: first an introduction to specific topic under study; second particular field description including sensing or/and measuring applications. Each of chapter is ending by complete list of carefully selected references with books, journals, conference proceedings and web sites.

The Vol.1 is devoted to various topics of optics and optic instrumentation, and contains 17 chapters written by 36 experts in the field from 15 countries: Brazil, China, Denmark, France, Germany, India, Japan, Mexico, Russia, Turkey, Slovenia, South Korea, UK, Ukraine and USA.

'*Advances in Optics: Reviews*' Book Series is a comprehensive study of the field of optics, which provides readers with the most up-to-date coverage of optics, photonics and lasers with a good balance of practical and theoretical aspects. Directed towards both physicists and engineers this Book Series is also suitable for audiences focusing on applications of optics. A clear comprehensive presentation makes these books work well as both a teaching resources and a reference books. The book is intended for researchers and scientists in physics and optics, in academia and industry, as well as postgraduate students.

I shall gratefully receive any advices, comments, suggestions and notes from readers to make the next volumes of '*Advances in Optics: Reviews*' Book Series very interesting and useful.

Dr. Sergey Y. Yurish

*Editor
IFSA Publishing*

Barcelona, Spain

Chapter 1

Multilayer Mirrors for Attosecond Pulses

Alexander Guggenmos and Ulf Kleineberg

1.1. Introduction

Multilayer-based optics have been developed into well-established optical components for steering [1], focusing [2], polarizing [3] and spectral filtering [4] of soft x-ray radiation since its first description by Spiller in 1976 [5].

Their significant advantage compared to single-surface grazing incidence optics working in external total reflection lays in an increased glancing angle, even up to near-normal incidence, larger collection angles, better imaging performance due to decreased astigmatism, and tailor-made spectral filtering. However, in many cases multilayer reflectors do not reach the high reflectance values of grazing incidence mirrors and are limited to peak reflectances slightly above 70 % at best, mainly due to absorption and scattering losses [6].

Due to remarkable improvements in various deposition technique like electron beam evaporation [7, 8], magnetron [9-11] and ion beam sputtering [12-14] or pulsed laser deposition [15-17], multilayers of various metal or semiconductor materials with layer thicknesses ranging down to less than 1 nm, layer numbers of many hundreds to thousands, and layer precisions in terms of thickness and interface accuracies on an atomic length scale (0.1 nm) have been realized.

While most multilayer optics development in the past was directed toward improved reflectivities for various multilayer material combinations and various photon energies from extended ultraviolet to the soft x-ray regime, and such multilayer optics have been employed as imaging devices in solar astronomy [18], extreme ultraviolet lithography [19] and microscopy [20], plasma diagnostics [21], Synchrotron Radiation optics [22] and many more, their application to the formation and handling of ultrashort soft x-ray radiation bursts with a few femtosecond (10^{-15} s) to attosecond (10^{-18} s) pulse durations, as they have become accessible since a few years by advanced laser technology in High

Alexander Guggenmos
Faculty of physics, Ludwig-Maximilians-University Munich, Am Coulombwall 1, D-85748 Garching, Germany

Harmonic Generation (HHG) [23] or advanced electron accelerator technology in Free Electron Lasers [24], has only recently been discussed in literature, however gaining increasing attention from the scientific community.

This article is a review about the status of (our) multilayer attosecond optics development in these fields.

1.1.1. Attosecond Pulses (Isolated and Pulse Trains)

Attosecond pulses of soft x-ray radiation bursts, either as isolated single pulses (where one attosecond pulse is generated by one laser pulse) or as pulse trains (where multiple about equal spaced attosecond pulses are generated by one laser pulse) have become possible by the invention of femtosecond laser amplifiers delivering few-cycle laser pulses with controlled electric field under the pulse envelope (Carrier Envelope Offset phase stabilization) and their frequency conversion in High Harmonic Generation (HHG). They constitute as of today one of the shortest electromagnetic pulses in time and provide the basis for time-resolved experiments characterizing the temporal dynamics of electron wave packets in atoms, molecules or even condensed matter systems.

While HHG is delivering extremely short pulses in time, their very small conversion efficiency limits their applicability currently to the extreme ultraviolet (lower soft x-ray) range and to very low pulse energies.

Pulses of higher photon energy in the soft- to hard x-ray range and much larger pulse energies can be delivered by Free Electron X-ray Laser sources, however their pulse duration is currently limited to some tens of femtoseconds. Very recent experiments performed at LCLS however might have an experimental way to further compress such pulses into the few femtosecond, or even sub-femtosecond, regime [25], which might then outperform current HHG sources in the foreseen future.

However, both sources in common is the generation of broad bandwidth radiation at short spectral wavelength, which requires proper handling of the spectral amplitude as well as of the spectral phase by subsequent optics components. Multilayer optics are proven to play a dominant role in controlling the amplitude and phase of the reflected radiation, thus providing a powerful toolbox for temporal pulse shaping. These properties are described in the following article.

1.1.2. HHG as Major/Only Attosecond Pulse Source

There are many publications on the generation of attosecond pulses by High Harmonic Generation, delivering a very intuitive theoretical description [26] of this process as well as describing the required experimental setups [27] and generated radiation spectra and pulses [28].

The process of HHG in atoms and molecules is theoretically best described by the three-step model of Paul Corkum [29], separating the sub-steps into laser-field induced tunnel

ionization of electrons during a fraction of laser period, followed by the acceleration of the free charges in the ponderomotive laser field, and finally the re-scattering of the returning electrons with their mother cores.

It is important to note, that this process delivers a coherent (spatially and temporally) radiation spectrum with multiple odd overtones (harmonics) of the driving laser frequency starting from the lasers fundamental wavelength (with a Ti: sapphire laser usually in the near-infrared at 780 nm) ranging over the strongly modulated and plateau area toward the weakly modulated cut-off regime, constituting the shortest HHG wavelengths.

The spectral position of the HHG cut-off is determined by the lasers pulse intensity and frequency as well as the ionization potential of the target atom and is given by the following formula:

$$E_{cut-off} = I_p + 3.17 \times U_p, \quad (1.1)$$

with the target atom ionization potential I_p and the ponderomotive potential of the laser:

$$U_p = \frac{Ie^2}{2m_0\epsilon_0 c\omega_L^2}. \quad (1.2)$$

The forming of single isolated attosecond pulses from such broad band HHG radiation spectra is performed by filtering out the shortest-wavelength part of the HHG spectrum in the cut-off regime, which is typically achieved by transparent filter foils, grazing incidence or multilayer reflectors, or a combination of those elements.

While thin-film filters and grazing incidence optics are not very versatile for tailor-made spectral filtering and spectral phase manipulation, multilayer optics as such described herein allow for a wide range of tailor-made spectra and phases, thus providing best flexibility for temporal pulse shaping.

1.1.3. What Is It Good for? The Applications

Coherent extreme ultraviolet pulses from High Harmonic Generation and their optics components find a wide range of applications in experiments, where some make use of the almost perfect coherence of the generated radiation, such as in experiments on extreme-ultraviolet interferometry [30], HHG microscopy such as Coherent Diffractive Imaging [4], or the generation of EUV frequency combs for advanced metrology [31].

However, the majority of applications of HHG radiation utilizes the attosecond temporal structure, e.g. in pump-probe experiments on electron streaking providing insights in the electron wave packet dynamics in atoms [32] and molecules [33] as well as giving new insights in the photoemission process from solid surfaces [34] or nanostructures [35].

While most of the experiments utilize single isolated attosecond pulses, as they are required for clean electron streaking experiments, recently also the use of attosecond pulse

trains has been established in RABBITT measurements and might provide a path to attosecond angle-resolved photoemission spectroscopy (ARPES) characterization of solid surfaces in the near future [36].

It is important to note, that all those different experiments impose differing requirements on the inevitable spectral filtering of the HHG spectrum by the extreme ultraviolet optical components, which can in many cases be best achieved by proper multilayer reflectors.

1.2. Reflective Optics in the EUV/Soft X-Ray Range

1.2.1. Interaction of Radiation with Matter

The interaction of electromagnetic waves with matter can be described by the index of refraction $n = c/v_p$, where c is the speed of light and v_p is the phase velocity in the medium. The phase velocity is usually decelerated in a medium, since the electric field disturbs the charges of each atom proportional to the electric susceptibility of the medium. This disturbance decreases with increasing frequency. Fig. 1.1 illustrates the progression of the index of refraction dependent on the electromagnetic wave's frequency ω .

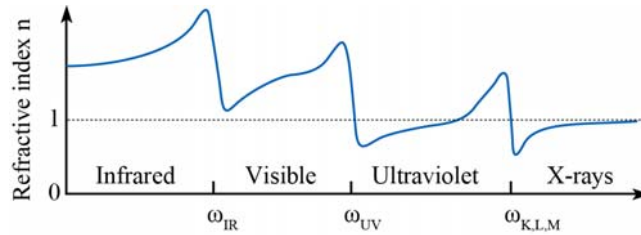


Fig. 1.1. Real part of the index of refraction as a function of the frequency with strong variations near the resonance frequencies [37].

It shows strong variations near IR, UV and x-ray resonances (ω_r) and the tendency toward unity for very short wavelengths. Only the real part of the index of refraction is shown in Fig. 1.1. Its adaption to a complex refractive index is necessary in case of attenuation in the medium. When electromagnetic waves get attenuated in a medium one considers this absorptive effect and defines the complex index of refraction as:

$$\tilde{n} = n + i\kappa, \quad (1.3)$$

where κ is the extinction coefficient. The real part of Eq. (1.3) approaches unity for electromagnetic waves with wavelengths in the EUV/soft x-ray range, since the dispersion of the medium strongly decreases. One considers the low dispersion background in that energy range by rewriting Eq. (1.3) to the EUV/soft x-ray form:

$$\tilde{n}(\omega) = 1 - \delta + i\beta, \quad (1.4)$$

with the dispersion δ and the extinction coefficient β . The influence of δ and β on the light-matter interaction becomes clear, when one considers the propagation of a plane wave in z -direction through a material with a refractive index n . The electric field can be written as:

$$E(z, t) = E_0 e^{-i(\omega t - z/r)}. \quad (1.5)$$

The complex index of refraction is considered by:

$$\frac{\omega}{k} = \frac{c}{n} = \frac{c}{1 - \delta + i\beta} \Leftrightarrow k = \frac{\omega}{c} (1 - \delta + i\beta). \quad (1.6)$$

Using Eq. (1.6) in Eq. (1.5) leads to:

$$E(z, t) = \underbrace{E_0 e^{-i\omega\left(t - \frac{z}{c}\right)}}_{\text{vacuum propagation}} \underbrace{e^{-i\left(\frac{2\pi\delta}{\lambda}\right)z}}_{\text{phase shift}} \underbrace{e^{-\left(\frac{2\pi\beta}{\lambda}\right)z}}_{\text{decay}}. \quad (1.7)$$

The first term in Eq. (1.7) describes the propagation of a plane wave through vacuum, the second term contributes a phase shift dependent on the dispersion δ and the last term is a damping term, accounting for absorption within the medium. The extinction coefficient β is related to the absorption coefficient α by squaring Eq. (1.7):

$$I = I_0 e^{-\alpha z} \Leftrightarrow \alpha = \frac{4\pi\beta}{\lambda}, \quad (1.8)$$

where I represents the Intensity and α is the inverse of the absorption length $l_{abs} = \alpha^{-1}$.

Eq. (1.4) can also be written as a function of the real part (f_1) and imaginary part (f_2) of the atomic scattering factor:

$$n(\omega) = 1 - \frac{r_e n_a \lambda^2}{2\pi} [f_1(\omega) - if_2(\omega)], \quad (1.9)$$

with the classical electron radius $r_e = e^2/(4\pi\epsilon_0 mc^2)$ and the electron density per unit volume n_a . Comparing Eq. (1.9) and Eq. (1.4) yields the relation between the atomic scattering factor and the dispersion and extinction, respectively:

$$\delta = \frac{n_a r_e \lambda^2}{2\pi} f_1(\omega), \quad (1.10)$$

and

$$\beta = \frac{n_a r_e \lambda^2}{2\pi} f_2(\omega). \quad (1.11)$$

One can thus calculate the dispersion and the extinction by making use of the measured tabulated values of the atomic scattering factors by Henke et al. [38]. The index of refraction approaches unity due to these low values (10^{-2} - 10^{-7}) where the values of δ and β are typically on the same order. This approach and the high absorption for EUV/soft x-ray wavelengths (Eq. (1.8)) results in low refraction at an interface between two materials (due to the similarity of the optical constants) and a limited penetration depth (on the order of a few hundred nm). This limits the degree of freedom of EUV/soft x-ray optics, since they must be either very thin, be used under small grazing angles in reflection geometries or use multiple reflections from different interfaces as utilized in multilayer mirrors.

1.2.2. Single Interface Reflectance (and Transmission)

One can use e.g. metal bulk mirrors for a broadband and high reflective guidance of visible light under nearly arbitrary angles of incidence. Unlike the visible range, no comparable optics can be used in the EUV/soft x-ray range. If an electromagnetic wave incidents under an angle α_i on an interface which is formed by two layers with the refractive indexes n_1 and n_2 , it is usually reflected and transmitted (Fig. 1.2).

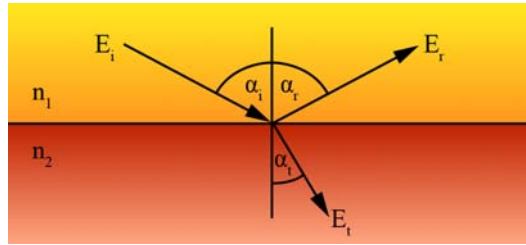


Fig. 1.2. Reflected and transmitted electric field at an interface, formed by two materials with the index of refraction n_1 and n_2 .

The degree of reflection and transmission depends on the incident angle α , on the index of refraction of both media and on the polarization state s or p. The characteristics are described by the Fresnel equations [39]:

$$r_s = \frac{E_{rs}}{E_{is}} = \frac{\tilde{n}_1 \cos \alpha_i - \tilde{n}_2 \cos \alpha_t}{\tilde{n}_1 \cos \alpha_i + \tilde{n}_2 \cos \alpha_t}, \quad (1.12)$$

$$t_s = \frac{E_{ts}}{E_{is}} = \frac{2\tilde{n}_1 \cos \alpha_i}{\tilde{n}_1 \cos \alpha_i + \tilde{n}_2 \cos \alpha_t}, \quad (1.13)$$

$$r_p = \frac{E_{rp}}{E_{ip}} = \frac{\tilde{n}_2 \cos \alpha_i - \tilde{n}_1 \cos \alpha_t}{\tilde{n}_1 \cos \alpha_i + \tilde{n}_2 \cos \alpha_t}, \quad (1.14)$$

$$t_p = \frac{E_{ip}}{E_{ip}} = \frac{2\tilde{n}_1 \cos \alpha_i}{\tilde{n}_1 \cos \alpha_i + \tilde{n}_2 \cos \alpha_i}. \quad (1.15)$$

Here, r and t are the reflection and transmission coefficients, respectively and E represents the complex electric field amplitude, while the indices determine the polarization (s or p) and whether the incident (i), the reflected (r) or the transmitted (t) fraction is considered. The different index of refraction is considered by n_1 and n_2 . Throughout this article, normal incidence angles are referred to as α and grazing angles as θ . The single interface reflectance for angles much larger than the critical angle θ_c ($\theta = 90^\circ - \alpha$) can be approximated to:

$$R_{i,2} = \frac{\Delta\delta^2 + \Delta\beta^2}{4\cos^4(\alpha_i)}, \text{ for } \beta \ll \delta \ll 1, \theta \gg \theta_c \left(\sin \theta_c \approx \sqrt{2\delta} \right). \quad (1.16)$$

Summarized, the single interface reflectivity is given by the difference of the optical constants of both media and the normal incidence angle α_i and is a rather low value in the EUV/soft x-ray wavelengths range.

1.2.3. Multilayer Mirrors

Multilayer mirrors address the issue of low single interface reflectivity as was described in the previous section. Creating a stack of alternating layers of varied materials makes use of constructive interference of partially reflected electromagnetic waves at each interface. Dependent on the working energy range and bandwidth 10-1000 interfaces are utilized. Each single interface reflectivity is still small but adding up each interface reflectivity can result in a normal incidence EUV reflectivity above 70 % [6] near the Si L_3 -edge. Constructive interference of the waves reflected at each interface can be achieved by a proper material choice and thickness design yielding a tailored reflectivity characteristic. This allows for a very precise spectrally filtering and shaping of attosecond pulses. Multilayer can be coated on arbitrary surfaces which allows for various shapes like parabolas, ellipsoids or the most commonly used spheres. Summarized, multilayer mirrors can tailor attosecond pulses and can be used for beam guidance, imaging or focusing and stand out as key components in attosecond science. The development of multilayer mirrors for the EUV started in the 1970s [5], it also pushed the development in the VIS/IR as the EUV range is more challenging and achieved results in that energy range can be easily adapted to the VIS/IR range. Nowadays it's state-of-the-art in the VIS/IR range to calculate [40] or even realize [41] very complex designs by the usage of various techniques [42, 43] and fulfill different requirements such as enough photon throughput or dispersion control at the same time [44]. This development and its results were extended to the EUV/soft x-ray range in the last years for various applications such as soft x-ray microscopy [20], EUV lithography [19], astronomy [18] or time resolved attosecond spectroscopy [45]. The last-mentioned attosecond science will be extended from the EUV range to the soft x-ray water window range in the near future.

1.2.3.1. Principal of Multilayer Mirrors

In Section 1.2.2 the reflectance and transmission of a single interface was described. This is now extended to a multilayer system, where multiple interface reflections and transmissions occur. One starts with the j -th layer within a stack consisting of one interface at the top and one at the bottom. The corresponding reflection coefficient reads as:

$$r_j = r_{(j-1)j} + t_{(j-1)j} t_{j(j-1)} r_{j(j+1)} e^{-2i\varphi_j} + t_{(j-1)j} t_{j(j-1)} r_{j(j-1)} r_{j(j+1)}^2 e^{-4i\varphi_j} + \dots, \quad (1.17)$$

where r and t are the reflection and transmission coefficients, respectively. Their indices determine at which interface the reflection and transmission occurs. The phase difference between the waves reflected at the bottom and the top interface is referred to as φ_j and can be written as:

$$\varphi_j = \frac{2\pi d_j}{\lambda} n_j \sin \theta_j, \quad (1.18)$$

where n_j is the index of refraction and θ_j the angle of incidence of the j -th layer. Fig. 1.3 illustrates all these parameters.

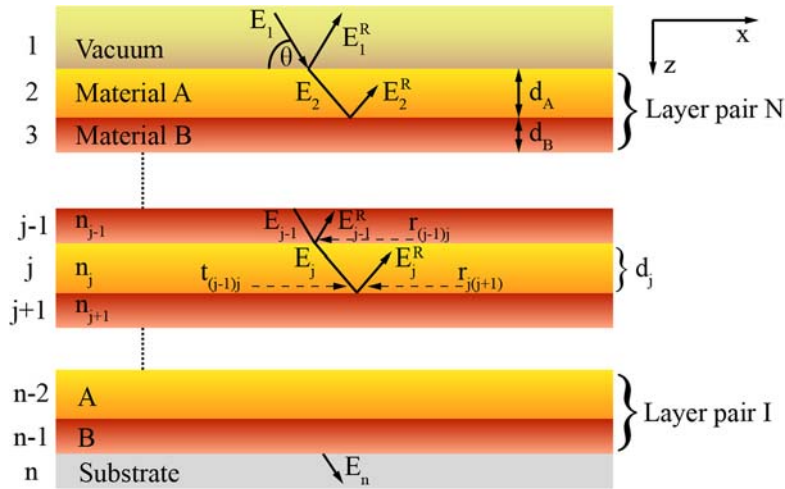


Fig. 1.3. Principle of multiple reflections and transmissions within a multilayer stack. The incident electromagnetic wave is partly reflected and transmitted at each interface (figure is related to [46]).

The sum in Eq. (1.17) converges to [47]:

$$r_j = \frac{r_{(j-1)j} + r_{j(j+1)} e^{-2i\varphi_j}}{1 + r_{(j-1)j} r_{j(j+1)} e^{-2i\varphi_j}}. \quad (1.19)$$

With this recursive formula, one can calculate the reflectivity of the overall multilayer stack. One starts from the lowermost layer of the stack, where the reflectivity is given by the single interface reflectivity of the substrate, and the first layer, and ‘walks’ through the stack. The reflectivity is calculated layer by layer and one ends up with the overall reflectivity $r_{ML} = r_I$. Its intensity is given by $R = |r_I|^2$.

1.2.3.2. Periodic and Aperiodic Multilayer Systems

The main classification for multilayer mirrors is to distinguish between periodic and aperiodic systems. Periodic mirrors usually consist of two (three) varied materials in alternating order where one usually refers to the low absorbing material as *spacer* whereas the high absorbing material as *absorber*. Aperiodic multilayer systems have individual layer thicknesses without any periodicity. Periodic stacks are characterized by the following parameters:

- The period thickness d which is constant throughout the stack.
- The period number N is the number of bilayers (trilayers) in a two (three) material stack. In a more general way: How often is the period used within a stack ?
- The fraction of the different layer thicknesses within one period, denoted as γ ratio. In a two-material stack: $\gamma = d_{bottom}/d \leftrightarrow d_{top} = d(1-\gamma)$. In the case of more than two materials per period one defines one γ_A per material A.

Fig. 1.4 illustrates a typical Gaussian like reflectivity and flat phase (within the bandwidth) shape of a periodic multilayer mirror.

The characteristic values such as maximum reflectivity, center energy or bandwidth are influenced by the listed multilayer parameters.

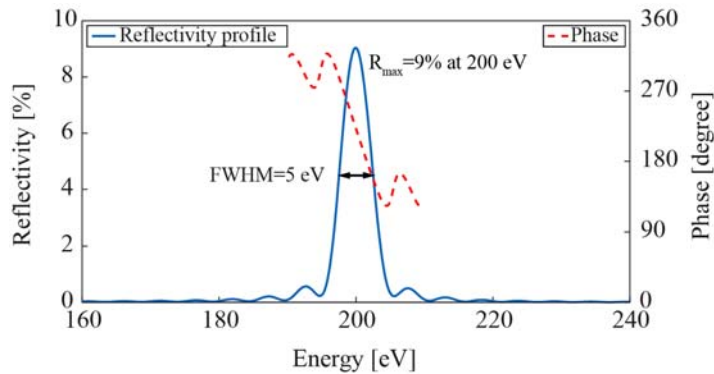


Fig. 1.4. Reflectivity and phase shape of a periodic multilayer mirror. Simulated reflectivity (solid blue) and phase (dashed red) of a periodic Cr/Sc multilayer mirror. Following parameters have been used for the simulation: spacer Sc, absorber Cr, $d = 3.144$ nm, $\gamma = 0.6$, $d_{Sc} = 1.886$ nm, $d_{Cr} = 1.258$ nm, $N = 37$, $\alpha = 5^\circ$.

The period thickness determines the center wavelength of the mirror according to the modified Bragg equation accounting for refraction:

$$m\lambda = 2d \cdot \sin \theta \sqrt{1 + \frac{\bar{n}^2 - 1}{\sin^2 \theta}}, \quad (1.20)$$

where m is the diffraction order, λ is the wavelength of the incident light, θ is the grazing angle and n the mean real part of the complex refractive index including the averaged dispersion within one period d consisting of k materials:

$$\bar{\delta} = \sum_k \frac{d_k \delta_k}{d}, \quad (1.21)$$

where d_k is the individual layer thickness and δ_k the corresponding dispersion within the period thickness d . The square root in Eq. (1.20) takes refraction effects into account and is sometimes neglected at normal incidence as the denominator approaches unity and $n \approx 1$. Neglecting the square root in Eq. (1.20) yields the condition for maximum normal incidence reflectivity: the optical period thickness is equal to $\lambda/2$, i.e. $d \approx 6$ nm at 100 eV.

The period number N influences the overall reflectivity, since more contributing layers increase this value, but its continuous increment is limited by the absorption of the materials. At the point of saturation more periods will not contribute to the overall reflectivity. The number of contributing periods can be increased by adapting the gamma ratio for reducing the amount of absorbing material. But this procedure is limited in case of mirrors for attosecond pulses. For the support of attosecond pulses the reflectivity is a very important factor but comes always along with the pulse duration support. This point is linked by the bandwidth support and can be approximated by the multilayer parameters by:

$$\frac{\Delta\lambda}{\lambda} = \frac{\Delta E}{E} \approx \frac{1}{mN}, \quad (1.22)$$

where m is the diffraction order. The bandwidth is linked to the support of the temporal pulse duration by

$$\Delta E[eV] \cdot \tau_0[as] \approx 1824[eV \cdot as], \quad (1.23)$$

which is known as the time-bandwidth product and is a result of the Fast Fourier Transform (FFT) of a Gaussian pulse and its full width at half maximum (FWHM) consideration.

Periodic multilayer mirrors have an almost linear phase within the bandwidth, since the strong periodicity ensures that each period contributes equally throughout the stack. Consequently, periodic multilayer mirrors do not temporal broaden an incoming pulse, but are therefore also not able to compress an incoming pulse. The influence on pulse shaping is limited by periodic systems.

A higher degree of freedom for attosecond pulse shaping is achieved by utilizing aperiodic multilayer mirrors in contrast to periodic stacks. While each period is equal in a periodic stack, i.e. the thickness for a certain material is constant, the thicknesses are arbitrary [46] in an aperiodic stack. This facilitates addressing different frequency components by individual layer thicknesses. This allows for spectral and dispersion shaping as different frequency components are reflected at different penetration depths within the stack. Fig. 1.5 illustrates the principle of aperiodic multilayer stacks.

In Fig. 1.5 (a) longer wavelengths are reflected on the top whereas short wavelengths are reflected at the bottom. Therefore, the short wavelengths are temporally delayed in respect to the longer wavelengths due to the longer penetration distance. This is defined as positive chirp; the mirror introduces a positive group delay dispersion (GDD). Vice versa in case of a negative chirp (Fig. 1.5 (b)). This facilitates shaping the dispersion of an incoming attosecond pulse and tailoring the reflected pulse. Aperiodic multilayer systems in the visible/infrared spectral range (e.g. (dielectric) multilayer mirrors in ‘chirped pulse amplification’ (CPA) systems) show an almost linear decreasing or increasing (dependent on the introduced GDD) of the layer thicknesses [41], due to missing absorption in dielectric materials in this spectral range, whereas aperiodic EUV/soft x-ray multilayer mirrors exhibit usually a random stack arrangement. Therefore Fig. 1.5 reflects only a useful presentation of delaying different wavelengths at different penetration depths, a real multilayer stack design for the EUV differs from this presentation, since the influence of strong absorptions leads to a rather random like stack design, which can be seen in the next figure. Fig. 1.6(a) depicts the design of a complex aperiodic multilayer stack consisting of 48 different layers and five varied materials (MoSi_2 is formed after deposition [48]).

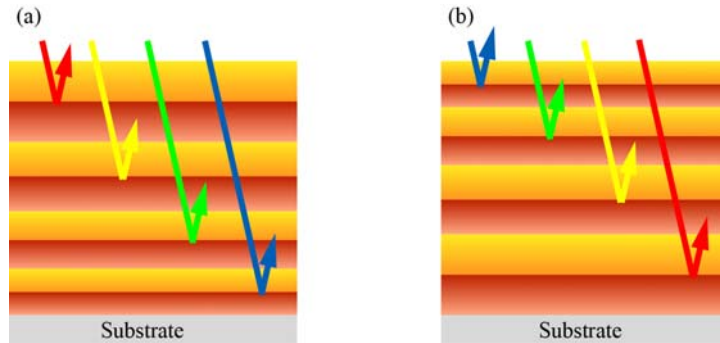


Fig. 1.5. Aperiodic multilayer stack. Sketch of aperiodic multilayer stacks, where different wavelengths are reflected at different penetration depths. The almost linear decreasing/increasing of the layer thickness is used for the VIS/IR range whereas it is rather a random stack arrangement in the EUV/soft x-ray range. (a) A positively chirped multilayer stack where short wavelengths are temporally delayed in respect to the longer wavelengths due to the longer penetration distance; (b) Vice versa in case of a negatively chirped mirror.

This complex design has been developed for the reflection of a FWHM > 40 eV and a close to Gaussian spectral profile, when it is convoluted with the transmission of a 150 nm thick

Pd filter (Fig. 1.6 (b)). A periodic stack supporting a bandwidth of 40 eV at 145 eV would, according to Eq. (1.22), have only four periods. That are eight layers in contrast to the 48 layers of the aperiodic system. This demonstrates why aperiodic multilayer mirrors exhibit a much higher degree of freedom in contrast to periodic stacks. Besides, one would expect that eight layers of the periodic case have a lower overall reflectivity as only a sixth of the aperiodic interface number is used. This is another advantage of aperiodic multilayer mirrors in addition to dispersion control. One can e.g. increase the reflectivity of a periodic system by switching to an aperiodic design while keeping the bandwidth constant.

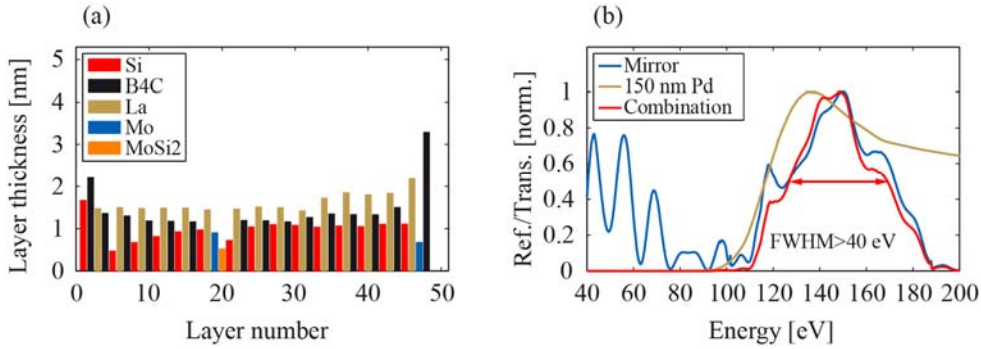


Fig. 1.6. A complex aperiodic La/Mo/Si/B4C mirror design: (a) Design of a complex aperiodic multilayer consisting of 48 different layers and five varied materials; (b) Reflectivity (blue), transmission of a 150 nm thick Pd filter (brown) and their convolution (red) resulting in a bandwidth of more than 40 eV. A normal incidence angle of 5 degree and a constant interface roughness of $\sigma = 0.5$ nm was assumed for the simulation.

1.2.3.3. Imperfect Interfaces

It is impossible to realize multilayer mirrors with perfectly abrupt and flat interfaces as finite boundaries occur during the fabrication. Interface imperfections are mainly caused by interface roughness [39] from fringed interfaces (Fig. 1.7 (a)) or atomic displacement as intermixing or interdiffusion (Fig. 1.7 (b)).

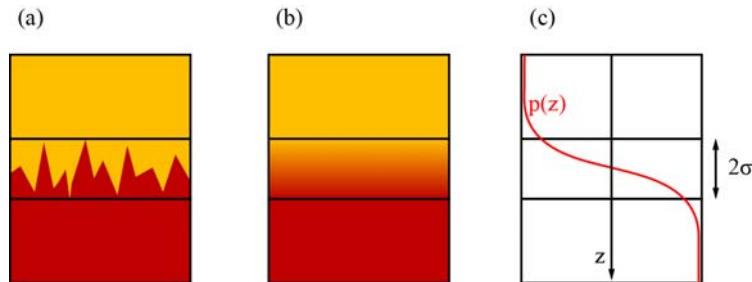


Fig. 1.7. The interface roughness σ : (a) Corrugated interface, (b) Intermixing and interdiffusion of both materials, (c) Interface profile function $p(z)$ for modeling both boundary imperfections.

Also, chemical reactions are possible forming an interlayer, as can be observed in Mo/Si multilayer where thin MoSi₂ layers are formed [48, 49]. This interlayer can be considered by including an additional layer with the corresponding refractive index of the compound. One should find a way to model these described imperfections and include it in simulations of multilayers, since it will be shown that they have a crucial effect on the reflectivity performance. This is mathematically done by the interface roughness value σ . One could now include the transition zone in simulations by representing the interface by a stack of very thin layers of continuously varying refractive index [39], but this would be a very time-consuming calculation especially when more than two materials are used or aperiodic stacks with a large layer number are simulated. Debye [50] and Waller [51] developed a rather elegant model to implement the roughness via Fourier transformation. An interface profile function $p(z)$, like the one-dimensional error function, is used to characterize the interface width for both roughness and diffusion (Fig. 1.7(c)):

$$p(z) = \text{erf}\left(\frac{z}{\sqrt{2}\sigma}\right), \quad (1.24)$$

with the one-dimensional Gauss error function $\text{erf}(x)$ and its FWHM σ . Using this profile function the reflection coefficient of the interface can be written as a Gaussian distribution:

$$r(z) = \frac{r_0}{\sigma\sqrt{2\pi}} e^{-z^2/(2\sigma^2)}. \quad (1.25)$$

A Fourier transformation of Eq. (1.25) yields the reflection coefficient in the momentum space:

$$r(q) = r_0 e^{-\frac{q^2\sigma^2}{2}} = r_0(\lambda) \exp\left(-\frac{8\pi^2 n^2 \sin^2(\theta)\sigma^2}{\lambda^2}\right), \quad (1.26)$$

with the wavelength dependent Fresnel coefficient $r_0(\lambda)$, the amplitude reflectivity of a perfectly abrupt boundary, and the momentum transfer $q = 4\pi n \sin(\theta)/\lambda$ between the boundary and the photon. Névot and Croce extended this method of Debye and Waller to an even more realistic description [52], considering the different refractive indices n and the resulting different propagation grazing angles θ of material 1 and material 2:

$$r(\lambda) = r_0(\lambda) \exp\left(-\frac{8\pi^2 n_1 n_2 \sin \theta_1 \sin \theta_2 \sigma^2}{\lambda^2}\right). \quad (1.27)$$

The Névot-Croce model is used for all simulations within this article to account for interface imperfections. Squaring Eq. (1.27) yields the intensity dependence:

$$R = R_0 \exp\left[-n_1 n_2 \sin \theta_1 \sin \theta_2 \left(\frac{4\pi\sigma}{\lambda}\right)^2\right]. \quad (1.28)$$

The influence of roughness on the reflectivity performance of a multilayer mirror for attosecond pulses is shown in Fig. 1.8.

The two plots of Fig. 1.8 are simulations of a molybdenum/silicon multilayer mirror ($N = 26$, $d = 6.7$ nm, $\gamma = 0.4$, $\alpha = 5^\circ$), including no molybdenum silicide layers, with a bandwidth of 5 eV in case of no interface roughness. According to Eq. (1.23) this would support ~ 400 as pulses. Eq. (1.28) has shown that an increased roughness value reduces the overall throughput, which is simulated in Fig. 1.8(a). Due to an increased diffuse scattering, this is accompanied by a decrease of the supported bandwidth. This is directly linked to the supported pulse duration, which is increased ($\sigma = 0$ nm supports a FWHM of 5 eV and thus ~ 400 as pulses whereas in case of $\sigma = 1.5$ nm only a bandwidth of 3.2 eV is possible resulting in a pulse duration support of ~ 600 as). Summarized a bad interface roughness reduces the overall photon throughput and increases the supported pulse duration. In terms of attosecond pulses both criteria are important and consequently the interface roughness should be minimized.

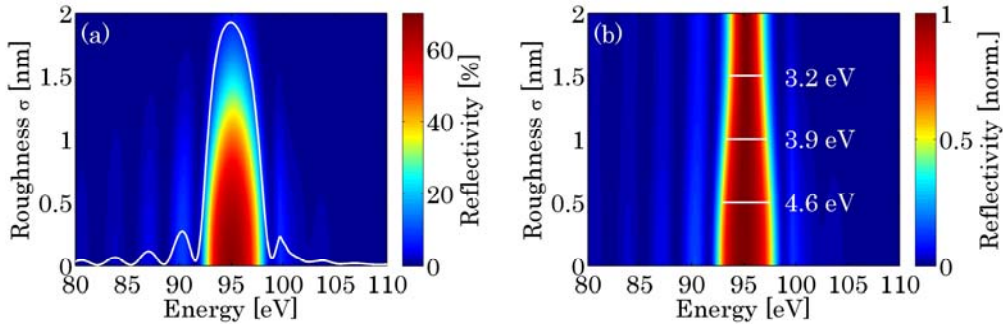


Fig. 1.8. (a) The influence of interface roughness σ on the reflectivity for a typical MoSi multilayer system for different σ values. The lineout indicates the reflectivity profile for $\sigma = 0$ nm (the reflectivity colormap is used as y-axis for the lineout); (b) Normalized reflectivity for each roughness value to demonstrate the decrease of bandwidth with increasing roughness.

1.3. Simulations of Multilayer Systems - Their Theoretical Optical Performance

For reflectivity and transmission calculations we use a self-written multilayer Fresnel code algorithm written with the software MATLAB. The overall reflectivity (mirror) or transmission (polarizer, filter) is calculated together with the relevant phase based on the multiple interface reflections and transmissions described in Section 1.2.3.1. The code uses the tabulated values of the atomic scattering factors from Henke and Gullikson [38] above 30 eV to calculate the dispersion (Eq. (1.10)) and the extinction (Eq. (1.11)) necessary for the Fresnel equations of Section 1.2.2. Realistic imperfect interfaces (Section 1.2.3.3), top layer oxidation or interlayer formation like molybdenum silicide [53] are considered as the stack design demands. We can thus identify various parameters which are used as input for the simulations. These parameters are either tabulated from

former characterization measurements or theoretical studies or should be extracted from own research studies and different characterization or analysis methods:

- **The optical constants.** Tabulated like Henke and Gullikson [38], updated self-determined tabulated values at certain absorption edges [54] or directly measured by spectral ellipsometry [55] for performance approximations in other spectral ranges.
- **Realistic interface roughness values.** Extracted from soft [56] or hard x-ray reflectometry [57] measurements as well as cross section transmission electron microscopy [58].
- **Top layer oxidation.** Analysis methods like for the roughness and additionally spectral ellipsometry. The thickness loss of the top material as well as the thickness of the grown oxide layer should be determined, since those two layers are responsible for a small aperiodicity within the stack as well as those are the first layers the radiation penetrates through, a very important fact in case of EUV radiation.
- **Interlayer formation.** Analysis methods like for the roughness.

The self-written code is used for simulating the optical performance of periodic multilayer systems or of already calculated aperiodic designs. The calculations of aperiodic designs are previously realized with the software *Optilayer*, a Fresnel code coupled to a 'needle optimization' algorithm [42, 43]. This software is used for the optimization (mainly chirped mirrors) of multilayer systems in combination with the self-written Fresnel code. The most important prerequisite for a reasonable optimization is the properly chosen start design. One usually starts with a periodic multilayer system which consists of the material combination best suitable for the aimed spectral range and adjusts the layer thicknesses (optimized γ -ratio) to the central energy E of interest. The period number N of the start design is chosen according to the required spectral bandwidth ΔE and is estimated by:

$$N = \frac{E}{\Delta E}. \quad (1.29)$$

Finally, one defines one or more target functions for the algorithm for which the start design should be optimized. The optimization can, for example, include a higher peak reflectivity, a spectral broader mirror reflectivity or in case of chirped mirrors an optimized spectral phase being realized by aperiodic layers, thus introducing a group delay dispersion in contrast to a flat phase in case of a periodic system. The needle optimization algorithm [42] starts with the periodic design and inserts small needlelike additional layers at certain positions in the stack, defined by the best benefit for the merit function. The thicknesses of these layers are consecutively increased or decreased until their best thickness value is achieved. Each optimization procedure consists of a certain number of iterations in which the current design is changed by random walk within user defined limits. The quality of a design $D_i(\omega)$ is quantized by the merit function MF including weights w_i (which must be predefined) that define the importance of the individual target function $T_i(\omega)$.

$$MF = \sum_i w_i [D_i(\omega) - T_i(\omega)]^2. \quad (1.30)$$

The figure-of-merit function further rules whether the modern design is taken as start design for the next iteration step or not. Thus, every calculated design is a trade-off between different target functions and is a consequence of the start design. Finally, the implementability of a final design should be checked and converted into deposition times using a self-written Fresnel-code algorithm. Interlayer formation, structural changes yielding thickness losses and a realistic Névot-Croce roughness is included in the code and both the spectral and the temporal structure of the pulse is analyzed including both the source and the filter characteristics to be able to compare the most realistic pulse simulation with the experimental constraints. Ultrathin layers because of the needle optimization procedure are either eliminated or broadened to a realizable value of at least 1 nm. If the final design does not fulfill the earlier formulated requirements an innovative design optimization is necessary either with a better start design or fresh design target points and weights. This procedure is used for aperiodic multilayer mirrors for shaping attosecond pulses.

1.4. Fabrication of Multilayer Coatings

Multilayer coatings are realized by various deposition techniques. Every technique has advantages and disadvantages and one should evaluate the coating priorities according to numerous criteria such as overall cost, deposition speed, layer density (which is directly connected to damage thresholds), thickness accuracy, number of target materials, availability of target materials, vacuum requirements, in-situ characterization techniques, coating homogeneity, to name only some of them. Nowadays following techniques are used for the realization of multilayer coatings:

- **Electron beam physical vapor deposition (EBPVD)** [7, 8], is a form of physical vapor deposition in which a target anode is bombarded with an electron beam given off by a charged tungsten filament under high vacuum. The electron beam causes atoms from the target to transform into the gaseous phase. These atoms then precipitate into solid form, coating everything in the vacuum chamber (within line of sight) with a thin layer of the anode material.
- **Magnetron sputtering** [9-11] is a Plasma Vapor Deposition (PVD) process in which a plasma is created and positively charged ions from the plasma are accelerated by an electrical field superimposed on the negatively charged electrode or ‘target’. The positive ions are accelerated by potentials ranging from a few hundred to a few thousand electron volts and strike the negative electrode with sufficient force to dislodge and eject atoms from the target. These atoms will be ejected in a typical line-of-sight cosine distribution from the face of the target and will condense on surfaces that are placed in proximity to the magnetron sputtering cathode.

- **(Dual) ion beam deposition (D-IBD)** [12-14] apparatus typically consists of an ion source, ion optics and the deposition target. Optionally a mass analyzer can be incorporated. In the ion source, materials in the form of a gas, an evaporated solid, or a solution (liquid) are ionized. The ions are then accelerated, focused or deflected using high voltages or magnetic fields. Optional deceleration at the substrate can be employed to define the deposition energy. This energy usually ranges from a few eV up to a few keV. At low energy, molecular ion beams are deposited intact (soft landing), while at a high deposition energy molecular ions fragment and atomic ions penetrate further into the material, a process known as ion implantation. The ion beam current, which is quantitative measure for the deposited amount of material, can be monitored during the deposition process.
- **Pulsed laser deposition (PLD)** [15-17] is a physical vapor deposition (PVD) technique where a high-power pulsed laser beam is focused inside a vacuum chamber to strike a target of the material that is to be deposited. This material is vaporized from the target (in a plasma plume) which deposits it as a thin film on a substrate. This process can occur in ultra-high vacuum or in the presence of a background gas, such as oxygen which is commonly used when depositing oxides to fully oxygenate the deposited films.
- **Metalorganic vapor phase epitaxy (MOVPE)**, also known as organometallic vapor phase epitaxy (OMVPE) or metalorganic chemical vapor deposition (MOCVD) [59, 60], is a chemical vapor deposition method used to produce single or polycrystalline thin films. It is a highly complex process for growing crystalline layers to create complex semiconductor multilayer structures. In contrast to molecular beam epitaxy (MBE) the growth of crystals is by chemical reaction and not physical deposition. This takes place not in a vacuum, but from the gas phase at moderate pressures (10 to 760 Torr). As such, this technique is preferred for the formation of devices incorporating thermodynamically metastable alloys, and it has become a major process in the manufacture of optoelectronics.

1.5. Achieved Results and Examples for Attosecond Multilayer Mirrors

This section will aim on the results which have been achieved so far in different spectral ranges. Ever shorter isolated attosecond pulses enable experimental access of absolute or relative timing of electronic processes with never achieved temporal resolution [26, 61, 62]. The filtering, reflection and shaping of attosecond pulses requires phase-correct broadband optics. Especially in normal incidence setups, the question of the shortest possible attosecond pulses is thus directly connected to the availability of appropriate optics, namely EUV multilayer mirrors and filters. One can estimate the bandwidth ΔE of an N -period multilayer stack with central energy E from Eq. (1.29): $\Delta E/E = 1/N$. An interference coating consists of at least two periods ($N \geq 2$), thus in principle, the maximum bandwidth of a periodic multilayer mirror is about half its central energy. Higher central energies relax the constraints on the bandwidth of optics and thus enable shorter reflected pulses but suffer usually from lower reflectivity.

1.5.1. A Normal Incidence Broadband 30-60 eV Mirror

Many attosecond experiments are performed in the spectral range between 30 eV and 60 eV, which is easy to access with HH sources [63-65]. The shortest attosecond pulses in this spectral range could be generated by *polarization gating*. Sansone et al. realized 130 as long pulses around 40 eV [64]. This and other experiments may benefit from broadband normal-incidence EUV optics suitable for attosecond beam guidance and focusing. An aperiodic broadband mirror reflecting between 35 eV and 50 eV with a reflectivity up to 20 % has been designed and published by Morlens et al. [66]. The here presented periodic Si/Sc mirror, centered at nearly the same central energy of ≈ 40 eV exceeds the bandwidth of Morlens' mirror by far. Its design is chosen such that the high single interface reflectivity of Si and Sc is extended by an appropriate quarter-wave design to a bandwidth of more than 20 eV. Until now, Sc/Si multilayers have mainly been used at energies between 25 eV and 35 eV [67] just below the Sc $M_{2,3}$ -edge. Above this energy, both Sc and Si are highly absorbing, and the number of contributing periods in Si/Sc multilayers is very limited. The normal incidence reflectivity saturates between 35 eV and 45 eV for only ≈ 5 periods. The relatively high reflectivity between 5-7 % is mostly independent from the mirror design due to a high single interface reflectivity of about 2 %. This property has been utilized for the design of the following ultrabroadband mirror. Its design is chosen such that the already high Si, Sc reflectivity is extended on the high energy side for a larger overall bandwidth. A 10 period Sc/Si coating with a period thickness of 14 nm (for a central energy of ≈ 50 eV) has been chosen (Fig. 1.9 (b)).

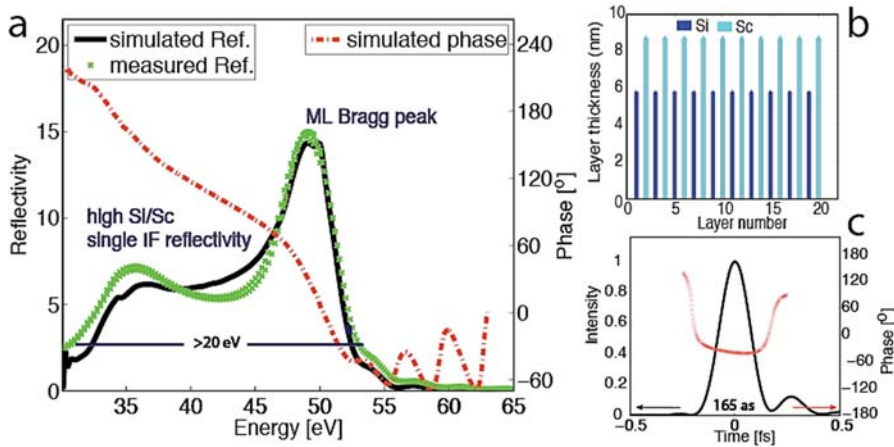


Fig. 1.9. Broadband normal incidence EUV multilayer mirror between 30 eV and 55 eV. Panel a): simulated (black) and measured (green) EUV reflectivity and spectral phase (red dash-dotted line). An $\sigma = 1.3$ nm has been used in the simulation. Panel b) shows the appropriate design of the 10 periods, $d = 14$ nm, $\gamma = 0.4$ Si/Sc stack. Panel c): Temporal structure of the mirror calculated as the Fourier transform of its spectral shape: It supports the reflection of 165 as sub-2 cycle EUV pulses.

Its designed and measured reflectivity is plotted together with the calculated phase in Fig. 1.9 (a). The measured EUV reflectivity confirms a successful implementation, a bandwidth of more than 20 eV and a reflectivity between 5 and 15 %. Simulations show

that this mirror is capable to support the reflection of Fourier limited 165 as long pulses what corresponds to only 1.7 EUV field oscillations at 42 eV central energy (Fig. 1.9 (c)). This mirror thus allows for the phase correct reflection of sub-2 cycle EUV pulses and is therefore at the possible duration limit at this energy.

1.5.2. Short Pulses in the 60-100 eV Range

Shorter pulses require even more bandwidth and thus necessarily higher central energies. The development of ultra-broad HH cutoff spectra with sufficiently high photon flux in the spectral range between 60 eV and 100 eV and the availability of appropriate thin metal filters and broadband high-reflectivity multilayer optics suggest this energy range for the generation of ever shorter ultrashort pulses. The HH generated EUV cut-off spectrum at about 100 eV matches the high reflectivity spectral band of Mo/Si multilayer mirrors; this material combination has thus been chosen. The generation of 80 as long pulses [68] requires at least 25 eV (FWHM) bandwidth (for a Gaussian spectrum) and is given by a multiplication of both the spectrum of the optics and the HH spectrum, from which the single pulse is being filtered. As will be presented in the following paragraph, this requires a HHG spectrum which exceeds the chirp-free cut-off region, ranging to the lower energy plateau region which is known to be positively chirped. This phase behavior (positive chirp at low energies, negligible chirp at high energies) exactly opposes that of a thin Zr filter, used in this experimental set-up anyway to filter the laser light from the EUV propagation path. As the filter indispensably compensates most of the HHG chirp, a chirp-less but broadband attosecond mirror is required. The penetration depth and thus the maximum number of usable bi-layers is limited by the attosecond pulse-length. A four-layer Mo/Si EUV coating centered at 70 eV, with a bandwidth of 32 eV (FWHM) has been designed for the coating of the inner part of the double mirror (Fig. 1.10(b)). Its reflectivity has been measured (green stars in panel a) and is in good agreement with simulations. Fig. 1.10 concludes the spectral (panel a)) and the temporal (panel c)) characteristics of this bandpass.

Simulations reveal that the mirror alone supports the reflection of nearly Fourier-limited pulses of 64 as (Fourier limit of the mirror: 63 as) calculated from the simulated spectrum (black line in Fig. 1.10 (c)) and the phase (red dash-dotted line). The effective bandpass bandwidth is experimentally limited by the transmission of the Zr filter which is required for suppression of the IR laser light and to tune the spectral phase. Different filter thicknesses (150, 300 and 450 nm) allow for some tunability of the GDD for the sake of reducing the effective bandwidth (the peak of the multiplied spectrum, plotted by the blue dashed line is shifted toward higher energies). It turned out that the combination of the mirror with a 300 nm Zr filter (yellow line, Fig. 1.10 (a)) compresses the incoming pulse to almost its Fourier limit. The goodness of the perfect interplay between the spectral phase of the bandpass and that of the HH source becomes more obvious, when one compares the temporal structure of the bandpass (panel c), blue dashed line) with its Fourier limit (light green dash-dotted line). The bandpass has a Fourier limit of 80 as, its temporal structure reveals 90 as including the phase (pulse broadening can be almost completely related to the phase of the filter). The final pulse-length of 80 as thus is at exactly the Fourier-limit of the bandpass. Please note here, that the final pulse is a

convolution of the bandpass and the incoming pulse. The HH intensity decreases at higher energies in the cut-off range and yields a slight down shift and a slight broadening of the bandpass spectrum. This experiment resulted in the generation of light flashes of no more than 80 as and has been published by E. Goulielmakis et al. in Science [68].

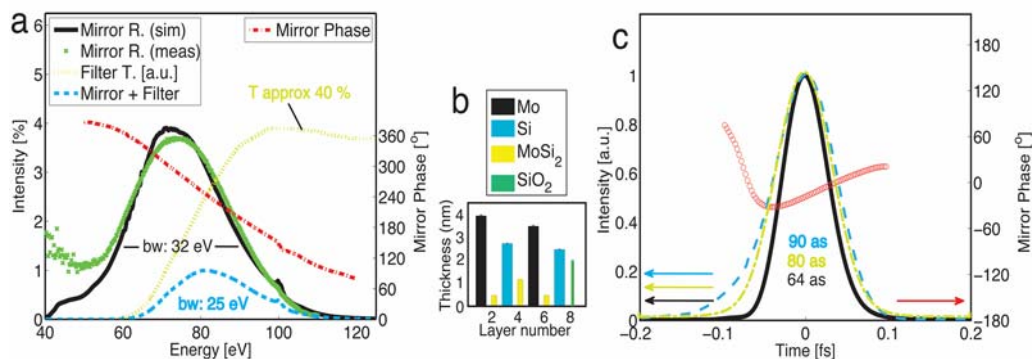


Fig. 1.10. Panel a): Simulated (black line) and measured (green crosses) reflectivity of a 32 eV (FWHM) bandwidth flat phase (red dash dotted line) multilayer mirror (design plotted in panel b), revealing a measured 3.6 % reflectivity centered at 70 eV. Together with a 300 nm thick Zr filter (yellow dotted line) it functions as a 25 eV broad bandpass and compressor (blue dashed line) for the generation of 80 as EUV pulses. Panel c shows the simulated 64 as long pulse (black line) and the temporal phase (red circles) of a pulse supported by the mirror upon reflection as an outcome of Fourier transforming the mirrors spectrum. The blue dashed line shows the temporal structure of the bandpass (mirror + filter), resulting in a 90 as long pulse, compared to its Fourier limit (light green dash dotted line) of 80 as.

1.5.3. Above the Silicon L-Edge in the 100 eV – 150 eV Photon Range

Various multilayer mirrors have been realized for attosecond pulses above 100 eV [69, 70] going from the EUV spectral range to the soft x-ray range for addressing more core states of numerous elements. The next section will describe as examples the attosecond dispersion control by multilayer mirrors, based on Hofstetter et al. [71], as well as a chromium/scandium mirror designed for a central energy around 145 eV and reflecting attosecond pulses on the order of 600 attoseconds. The later experiment was at that time realized at a central energy which was 27 eV higher than the recent energy limit in table-top attosecond pump studies [34].

1.5.3.1. Attosecond Dispersion Control by Multilayer Mirrors above 100 eV

Specially designed aperiodic binary or ternary multilayer-coated EUV mirrors allow for a large degree of freedom in influencing the EUV pulse characteristics by reflection of those mirrors [72]. So far, multilayer mirrors in attosecond physics applications were used as band-pass reflectors preserving the spectral phase of the attosecond pulse that is defined in the generation process and by the dispersion characteristics of all EUV optical elements passed by the pulse. In this work we demonstrate, how specially designed non-periodic

multilayer EUV mirrors furnish attosecond technology with the ability to control the spectral phase in addition to spectral intensity of attosecond pulses, offering a great degree of freedom in influencing the EUV pulse characteristics, such as the pulse shape, duration, frequency sweep and central wavelength.

To prove the validity of the concept for precision chirp control of attosecond EUV pulses, a set of three multilayer mirrors exhibiting distinctly different GDD but similar reflectivity characteristics was developed [71]. Optimized for an incident angle of 45° they exhibit high reflectivity within the range of 100-130 eV with comparable full widths at half maximum energy bandwidth: $\Delta E = 11-13$ eV FWHM and peaking at different central energies within the range of 107-122 eV.

The multilayer designs have been calculated and optimized by a Fresnel equation thin film code coupled to a needle optimization algorithm [42]. The mirrors have been designed such, that the GDD is almost maximum and mostly linear within the final attosecond pulse spectrum. Small shifts of the mirror spectrum due to the thin metal filter or the shape of the high harmonic cut-off spectrum have been estimated and considered in the design. The calculated designs have been optimized for additional suppressed reflectivity contributions in the near vicinity of the main Bragg peak and have been tested for stability against small layer thickness deviations. The final multilayer coating designs are displayed in the lower panels (d-f) of Fig. 1.11.

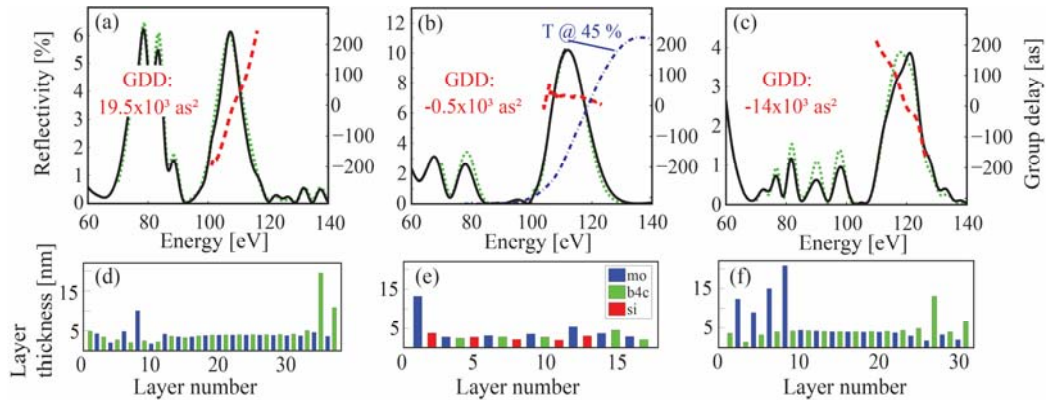


Fig. 1.11. Characteristics of EUV chirped multilayer coatings designed and manufactured for the experiments: Set of EUV multilayers with positive (a), near zero (b) and negative (c) group delay dispersion. The diagrams compare the measured (solid black) with the designed (dotted green) reflectivity curves. The calculated group delays (derivative of the spectral phase with respect to angular frequency) are also displayed (dashed red curve). The transmittivity of a 150 nm palladium foil is additionally qualitatively plotted in panel (b), depicted by the blue dash-dotted line. The lower panels (d), (e) and (f) show the corresponding mirror-designs of the appropriate mirrors above. Each color represents one material as explained by the legend in 1.2 (e).

While the designs of the positively- and the negatively-chirped mirrors contain around 30 layers Mo/B4C the almost unchirped mirror is a 17 layer Mo/Si/B4C stack. The mirrors have been deposited by means of Dual Ion Beam Deposition (DIBD) on flat high-polished

glass substrates. Interface losses and compound formation have been included in the calculation and could be compensated within the coating procedure.

Figs. 1.11 (a)-1.11 (c) shows the calculated reflectivity and group delay of the three multilayer mirrors versus photon energy, compared with the reflectivity measured by x-ray reflectometry using synchrotron radiation [73]. The measured and computed EUV reflectivities are in excellent agreement, revealing maximum peak reflectivities between 4 and 10 percent. Notable deviations in the peak shape appear only in Fig. 1.11 (c) near 120 eV, possibly due to minor uncertainties in the deposition layer thicknesses. In our proof-of-concept experiments, radiation at photon energies below 100 eV is suppressed by a 150 nm thick Palladium (Pd) filter (its transmittivity is shown in Fig. 1.11 (b) by the blue dash-dotted line). The throughput of this high-pass filter combination could be enhanced by fine-tuning the filter thickness and the mirrors' high-reflectivity band. Since isolated attosecond pulses can be extracted by spectral filtering from the cut-off part of the generated HH spectrum [1], the high energy extend of the spectrum must coincide with the high reflectivity range of the mirror. Reflectivity above this energy window thus does not affect the attosecond pulse generation. The three mirrors are designed to introduce substantial positive GDD, negligible GDD, and substantial negative GDD, with computed values of $19.5 \times 10^3 \text{ as}^2$, $-0.5 \times 10^3 \text{ as}^2$, and $14 \times 10^3 \text{ as}^2$, respectively.

We have implemented attosecond streaking by liberating photoelectrons from the 2p sub-shell of an ensemble of neon atoms with sub-300-attosecond EUV pulses filtered by the combination of a Pd foil and one of the band-pass multilayer mirrors described above. The energy distribution of the ejected electrons has been streaked by the controlled linearly polarized electric field of near-single-cycle near-infrared (NIR) laser pulses [68, 74, 75]. The streaked spectra for electrons collected in a narrow cone aligned with the laser polarization were recorded as a function of the delay between the ionizing EUV pulse and the streaking NIR field.

To access the temporal intensity profile of the synthesized attosecond pulses and their frequency sweep we performed a frequency-resolved optical gating (FROG) analysis [76] of the acquired streaking spectrograms. This method gives not only access to a full characterization of the laser vector potential, it allows us as well to characterize the spectral intensity distribution and the group delay variation [77, 78] of the final attosecond EUV pulse reflected from the mirrors after passing the filter, allowing for a direct comparison of its GDD with the design values of the mirrors' GDD. The resultant streaking spectrograms recorded with the three mirrors of Figs. 1.11 (a)-1.11 (c) are displayed in Figs. 1.12 (a)-1.12 (c) along with their corresponding FROG retrievals in Figs. 1.12 (d)-1.12 (f), respectively.

We can now turn our attention to a quantitative evaluation of the chirp carried by the attosecond EUV pulse in the three experiments performed with the three different mirrors. Fig. 1.13 shows the retrieved intensity and GD of the attosecond pulses after reflection off the EUV multilayer and filtering by the transmission filter as evaluated from the photoelectron spectrograms in Fig. 1.12.

The solid black lines in Figs. 1.13 (a) - 1.13 (c) show a direct measurement of the EUV pulse spectrum at the target after passing the filter and the appropriate mirror. Comparison of the retrieved EUV spectra (green dotted lines) with those directly measured photon spectra shows remarkable agreement of the bandwidth, spectral profile and central energy position. This agreement between retrieved and measured spectra highlights the precision with which EUV mirrors are designed and manufactured and demonstrates the power of attosecond streaking measurements coupled with the FROG CRAB retrieval procedure.

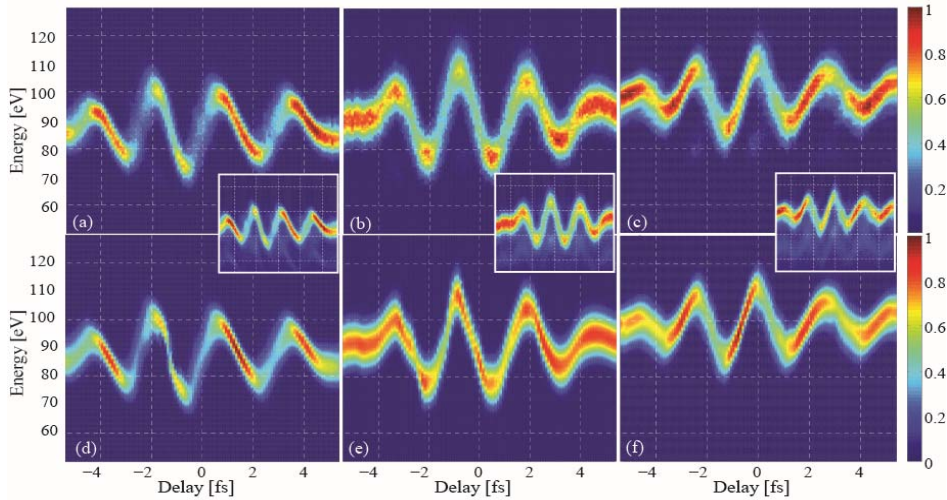


Fig. 1.12. Measured and retrieved attosecond electron streaking spectrograms, upper and lower row, respectively. Panels (a)-(c) and (d)-(f) display spectrograms recorded with attosecond pulses reflected off the mirrors described in Figs. 1.11 (a)-1.11 (c), respectively. The energy shift in the recorded spectra is proportional to the vector potential $A_L(t)$ of the streaking laser field at the instant of photoemission, hence the dependence of this energy shift on the delay between the attosecond EUV pulse and the NIR laser field reflects the temporal evolution of the NIR laser field's vector potential, $A_L(t)$. Narrowing and broadening in the streaked electron spectra at the zero crossings is displayed by enhanced and diminished electron count rates, respectively. The small insets show the energy calibrated raw data before subtraction of the background.

1.5.3.2. Chromium/Scandium Multilayer Mirror for Attosecond Pulses at 145 eV

This section will present the first application of Cr/Sc multilayer mirrors to attosecond science. The material system has been optimized for attosecond pulses and is now applied to these pulses for the first time. Most of the results have been published in [79]. The previous sections have demonstrated that multilayer mirrors provide a unique approach for beam steering, spatial and spectral shaping as well as spectral phase control with reasonably low reflective losses. It will be shown that optimizing the Cr/Sc material system is not only a key to future attosecond experiments in the water window [80], but also facilitates a promising choice for realizing new attosecond experiments at around 130-160 eV, the energy range where attosecond sources with sufficient photon flux are nowadays already available [28], but multilayer optics are very limited. Fig. 1.14 shows a

simulation comparison of certain established multilayer material systems reflecting (attosecond) HHG pulses with a central energy of 145 eV and a FWHM bandwidth of 3 eV at an angle of normal incidence of 5 degree.

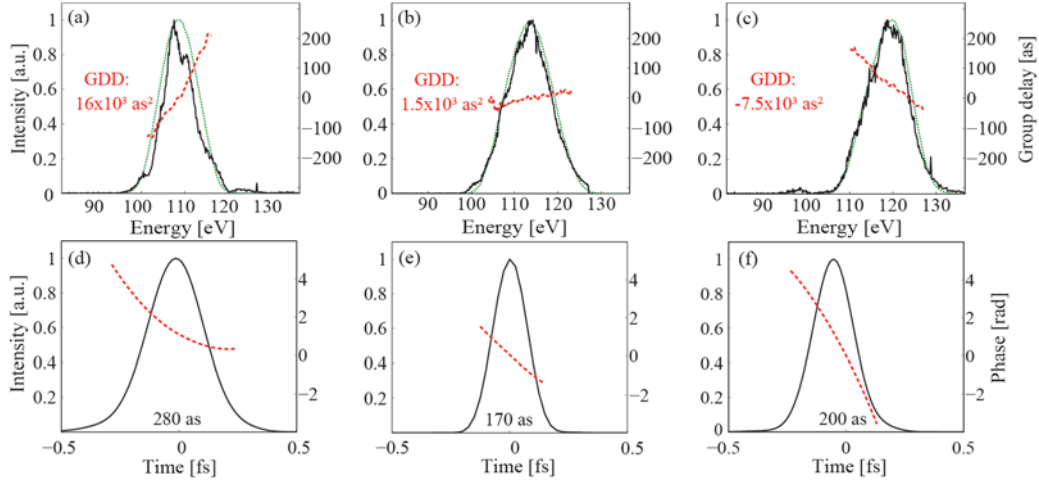


Fig. 1.13. Retrieved spectral and temporal characteristics of the attosecond EUV pulse, upper row and lower row, respectively. Panels (a), (b) and (c) display the intensity spectrum (green dotted line) and group delay (red dashed line) of the attosecond EUV pulse reflected off the mirrors described in Figs. 1.11 (a), 1.11 (b), and 1.11(c), respectively, as retrieved from the measured streaking spectrograms shown in Figs. 1.12 (a), 1.12 (b), and 1.12 (c), respectively. The black full lines depict the corresponding EUV spectra measured directly with an EUV spectrometer. The evaluated effective group-delay dispersions weighted by the final spectral EUV intensity are also shown. Panels (d), (e), (f) show the temporal intensity profile (black line) and temporal phase (red dashed line) retrieved from the streaking spectrograms of Figs. 1.12 (a), 1.12 (b), and 1.12 (c), respectively. The EUV pulse duration (full width at intensity half maximum) has been evaluated and is displayed at the bottom of each panel.

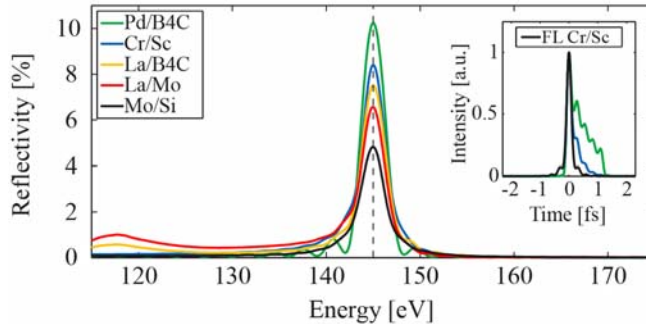


Fig. 1.14. Comparison of certain material systems at 145 eV. Simulation comparison of certain multilayer material systems for the reflection of a FWHM bandwidth of 3 eV centered at 145 eV. The small inset shows a comparison of the Cr/Sc and Pd/B4C system in the time domain including the transmission through a 200 nm thick palladium (Pd) filter. The black line shows the Fourier limit (FL) of the Cr/Sc mirror reflection.

The parameters were chosen as a trade-off between spectral and temporal resolution in high-resolution attosecond experiments. The simulations show only a weak suppression of unwanted low energy out-of band radiation in the range of ≈ 120 eV with lanthanum (La) based multilayer mirrors (La/Mo, La/B4C). The inevitable and commonly used metal filter for blocking the near-infrared (NIR) laser radiation (typically a 200 nm thick palladium (Pd) filter) cannot be used to eliminate the out-of-band radiation due to its transmission properties. As a result, chirped plateau harmonics are not sufficiently suppressed by such multilayer reflectors, which is a prerequisite for filtering isolated single attosecond pulses from the cut-off area of the high harmonic spectrum. This suppression of low energy out-of-band radiation is essential for attosecond spectroscopy experiments for e.g. delay measurements [45] or direct observation of electron propagation [34]. Well established Mo/Si mirrors, which are widely used in attosecond experiments at photon energies below the silicon L_3 -edge at ≈ 100 eV, suffer from very low reflectivity above 100 eV. Other molybdenum based multilayer systems like Mo/B4C, Mo/Y, Mo/Be or Mo/Sr, which on the one hand can provide a higher degree of out-of- band radiation suppression (for Mo/B4C) accompanied with a higher reflectivity (Mo/Y, Mo/Be, Mo/Sr) [81, 82, 83, 84] but on the other hand suffer from strong spectral modulations around the main reflectivity Bragg peak by Kiessig fringes and therefore introduce additional GDD, which broadens the pulse in the time domain. Furthermore, Mo/Sr is not stable and shows long-term degradation, and beryllium is strongly toxic, thus limiting experimental adoption. A reflection comparison in the time domain of multilayer mirrors composed of Cr/Sc and a highly reflective system, here as example Pd/B4C [85], is shown in the small inset of Fig. 1.14, which already takes the transmission and the spectral phase of a 200 nm thick Pd filter into account. Whereas the pulse reflection of the Cr/Sc mirror is close to its Fourier limit and exhibits a Gaussian pulse shape, the Pd/B4C system shows unwanted temporal pulse broadening due to GDD as well as temporal modulations resulting from the multilayer reflectivity fringes. However, the Cr/Sc multilayer mirror system combines all the advantages required for applications with HHG attosecond pulses: Sufficient throughput due to the optimized reflectivity [58], suppression of out-of-band radiation components (in case of the lanthanum based systems a thicker filter can increase the suppression in the 120 eV range but reduces the overall throughput as well) and a nearly (Fourier limited) Gaussian pulse profile, both in the spectral and temporal domain.

The experimental realization of the Cr/Sc attosecond multilayer mirror was performed by dual ion-beam sputtering technique [80] together with a tailored interface polishing process [58] for a higher mirror reflectivity. For a later characterization by attosecond streaking, the mirror was also analyzed by two independent measurement techniques, hard x-ray reflectometry and EUV/soft x-ray reflectometry. Those are not shown here.

To characterize the attosecond pulses upon reflection from the Cr/Sc multilayer mirror, the well-established EUV/soft x-ray pump/NIR probe streaking technique [74] was used. Here, both the attosecond soft x-ray pulse and the NIR laser pulse are focused by a double mirror into neon gas. The soft x-ray pulse photoionizes Ne atoms, which frees photoelectrons from the 2p shell, which are then momentum-streaked by the co-propagating temporally synchronized and phase stabilized NIR laser's electric field. The

inner part of the double mirror can be moved with respect to the outer part, to introduce a temporal delay between the soft x-ray pulse, which is reflected at the mirror core, and the laser pulse, which is reflected at the outer ring. Changing the delay between the laser and the soft x-ray attosecond pulse yields a typical streaking spectrogram (Fig. 1.15 (a)).

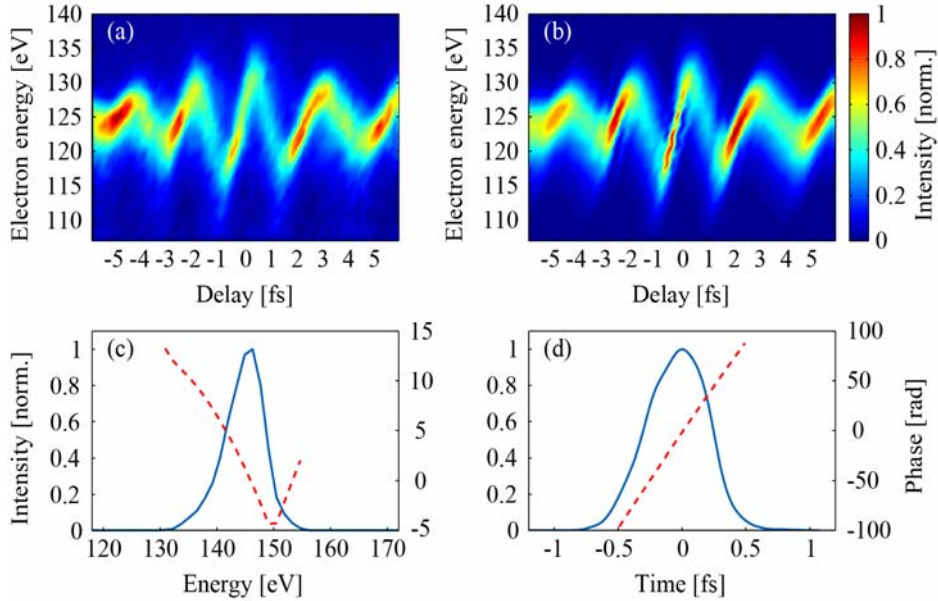


Fig. 1.15. Attosecond electron streaking at a photon energy of 145 eV. Results of an attosecond electron streaking experiment for pulse characterization in neon utilizing the Cr/Sc multilayer mirror. (a) Measured electron streaking trace and (b) the retrieved trace performed by FROG/CRAB analyses. (c) The retrieved soft x-ray pulse (solid blue) and the phase (dotted red) in the spectral domain. (d) Soft x-ray pulse and phase in the temporal domain.

FROG/CRAB [76] analysis allows for a complete reconstruction of both the intensity and the phase of the soft x-ray attosecond pulse, as well as the vector potential of the streaking laser field from a recorded spectrogram (Fig. 1.15(a)). Fig. 1.15 (b) shows the result of the appropriate FROG/CRAB retrieval as described in [78, 86]. The retrieved intensity and phase of the soft x-ray pulse are depicted in Fig. 1.15, once in the spectral (c) and once in the temporal (d) domain. Shifted by the binding energy of the Ne-2p electrons (21.6 eV) the soft x-ray pulse shows a central energy of 145.6 eV which agrees to the target and the previously described results. With the retrieved spectral bandwidth and phase being the most prominent sources of uncertainty, the temporal error can be estimated to be about 20 as. From the retrieved amplitude and phase, an attosecond pulse duration of about 580 as is found in the temporal domain which is in excellent agreement with the previous estimations (based solely on the mirror reflectivity plus the assumption of a flat mirror phase). A perfect Gaussian Fourier limited pulse with 2.9 eV bandwidth has a duration of 629 as. In summary, an optimized Cr/Sc multilayer mirror has been developed and applied for reflecting single isolated attosecond pulses at a photon energy of 145 eV with a pulse duration of 580 as.

1.5.4. Multilayer Mirrors for Attosecond Pulses in the Water Window Soft X-Ray Range

1.5.4.1. Aperiodic Multilayer Mirrors

Time resolved attosecond experiments are nowadays well established in the sub 120 eV EUV photon energy range [45, 64, 87]. With ever improving few cycle laser development toward higher pulse energies, this regime will be extended into the soft x-ray water window spectral range. Using thin filters for attosecond pulse shaping in this energy regime is limited to fixed opening and absorption edges of the used materials and thus restricts the degree of freedom. This leaves multilayer mirrors as the only key components for tailored spectral filtering and shaping of an attosecond water window pulse. For that reason, a negatively chirped aperiodic Cr/Sc multilayer mirror, optimized for reflecting sub-70 as pulses from HHG at a central photon energy of 326.3 eV (Ar L_1 -edge) with a bandwidth of about 30 eV, was developed for future resonant attosecond photo ionization experiments. A positively chirped mirror was realized for comparison reasons [80].

The two different aperiodic Cr/Sc multilayer mirrors have been optimized by the thin film program *Optilayer*, a Fresnel code coupled to a needle optimization algorithm [42, 43] which is described in Section 1.3. While the first multilayer mirror has been designed for the introduction of an averaged negative chirp (the group delay dispersion is the negative second derivative of the spectral phase, $GDD = -d\phi^2/d\omega^2$) of approximately -8000 as^2 to compensate a possible positive chirp of the high harmonic plateau, a second mirror with similar parameters has been designed to introduce an averaged positive chirp of approximately $+8000 \text{ as}^2$. Both mirrors have been designed such that their central energy coincides with the L_1 -edge of Ar (326.3 eV) at an incidence angle of 45 degree. The optimized stack designs are shown in Fig. 1.16, where the negative (a) and the positive (b) design consists of around 95 individual layers.

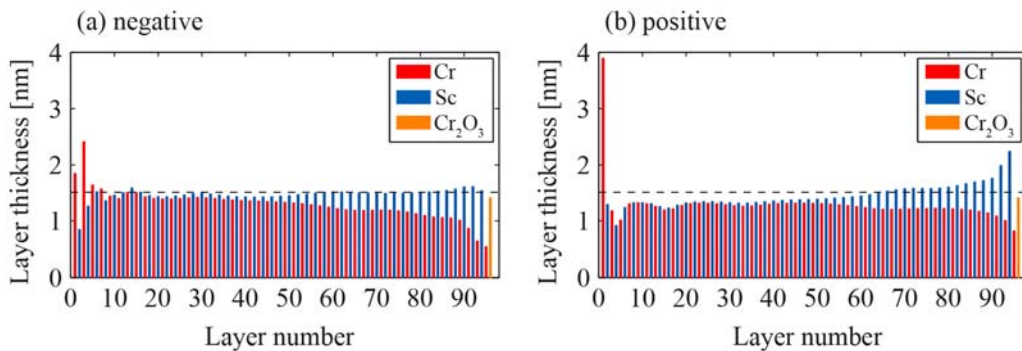


Fig. 1.16. Design of negatively and positively chirped Cr/Sc mirrors. Multilayer stack designs of the aperiodic Cr (red) and Sc (blue) layers for the introduced (a) negative, and (b) positive GDD. An arbitrary line with a layer thickness of $d = 1.5 \text{ nm}$ (dashed black) is depicted for comparison reasons.

The design has been chosen according to their robustness of GDD against small layer thickness errors. Typical layer thicknesses are between 1 nm and 2 nm, with the thinnest layers going down to 0.5 nm. For both designs a top layer of 1.4 nm Cr_2O_3 out of 0.3 nm Cr has been included in the model. These designs principally prove the large degree of freedom in customizing water window attosecond pulses both in space (substrate shape and incidence angle) and time, utilizing aperiodic Cr/Sc multilayer mirrors.

Soft x-ray reflectometry measurements were carried out at the beamline 6.3.2 of the ALS [56] to retrieve the reflectivity profile in the energy range for which the mirror designs aim at an incidence angle of 45 degree. A comparison of the simulated and measured soft x-ray reflectivity for both chirped mirrors is shown in Fig. 1.17.

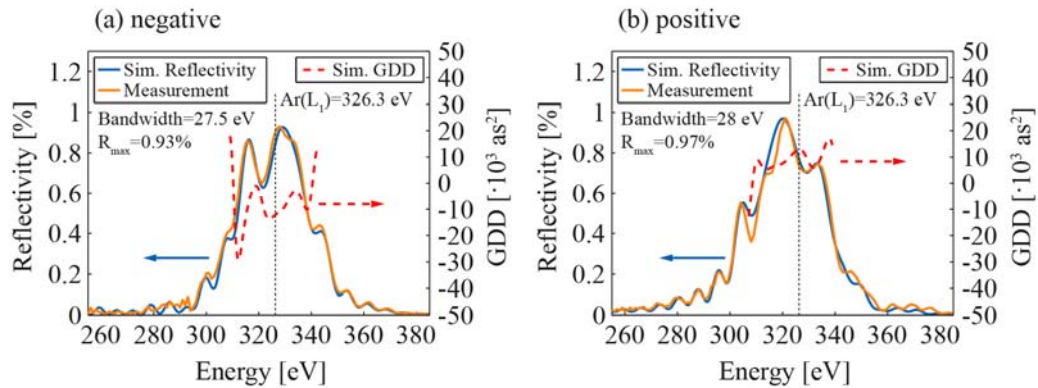


Fig. 1.17. Soft x-ray measurement of aperiodic Cr/Sc mirrors. Soft x-ray reflectivity measurement (brown) and simulation (blue) for the (a) negatively and (b) positively chirped Cr/Sc multilayer together with the corresponding simulated GDD (dashed red).

The reflectivity simulations were performed using the self-written Fresnel code of Section 1.3 and the start designs retrieved from the *Optilayer* optimization procedure [42]. An average Névo-Croce roughness factor of $\sigma = 0.49$ nm has been retrieved from simulations to account for interfacial imperfections. Both the central energy and the side peaks of the measurements and the designs coincide very well and indicate a nearly perfect experimental implementation of the simulated multilayer stack designs of Fig. 1.16. Also shown is the calculated evolution of the GDD within the reflectivity bandwidth of both multilayer systems, indicating an averaged GDD of ≈ -8000 as². Both multilayer systems are very similar in terms of peak energy, spectral bandwidth and peak reflectivity and only differ by the sign of their GDD. Note that the spectral multilayer phase (and thus the GDD) is not accessible by simple reflectivity measurements. Measurements of the spectral phase, by soft x-ray reflectometry, have been reported by detecting the standing-wave assisted total electron yield from the multilayer surface as a function of the photon energy around the Bragg peak [88, 89]. While this method is very suitable for periodic multilayer systems, its accuracy is limited when applied to aperiodic systems with a weak standing wave. On the other hand, one can characterize the spectral phase with two attosecond methods, the RABBITT-technique [63, 90] or by attosecond photoelectron streaking

spectroscopy providing access to a full characterization of the reflected attosecond pulse both in amplitude and phase [68, 72]. However, for the water window the implementation of both attosecond techniques has not been established yet due to a lack of sufficient photon flux from HHG attosecond sources in this spectral range. Here we analyze and estimate the influence of the layer errors on the reflectivity and the GDD via simulations. The rather perfect match of the designed and the measured reflectivity in Fig. 1.17 proofs the correct elimination of systematic deposition errors and leaves only random errors to be analyzed. Both the simulated reflectivity and the GDD of the original design of the negatively chirped multilayer is compared to slightly modified designs. The designs have been chosen randomly by joggling each layer thickness within predefined limits and the merit function (MF) of the reflectivity, as well as the corresponding averaged GDD value, was calculated to analyze the quality of reflectivity and GDD simulations:

$$MF = \sqrt{\frac{1}{N} \sum_{i=1}^N \left(R_{(sim),i} - R_{(softx-ray),i} \right)^2}, \quad (1.31)$$

where N is the total number of wavelength sampling with an integer i representing the position of the sampling equally spaced. 100 designs have been averaged per allowed error and both the reflectivity and the GDD deviations have been calculated. Designs with an overall stack height deviation of more than 0.8 % have been left out due to discrepancy with profilometry investigations. From Fig. 1.18(a) it can be deduced, that the average random thickness error is less than 0.5 %, corresponding to an average layer thickness error of approximately 0.05 angstrom.

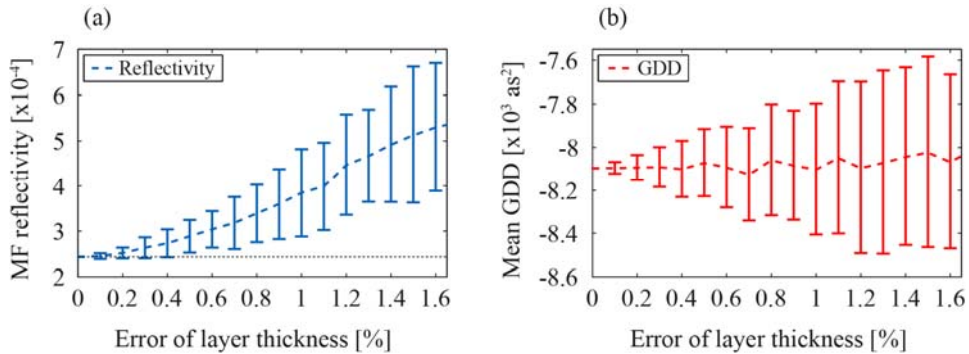


Fig. 1.18. Merit function and GDD dependence on layer thickness errors. (a) Merit function for the simulated reflectivity and the soft x-ray measurement for different layer thickness errors. (b) Mean GDD for different layer errors.

Fig. 1.18(b) shows, that the GDD mean value stays nearly constant independent of the layer errors. The upper limit of the layer errors of 0.5 % corresponds to a maximum mean GDD error of approximately $\pm 150 \text{ as}^2$. To summarize the results, we have a standard deviation of $\approx 0.024 \%$ in the soft x-ray reflectivity and a mean GDD of $-8090 \pm 150 \text{ as}^2$ for the aperiodic negatively chirped Cr/Sc multilayer mirror.

1.6. Metrology and Application

This section will describe three different measurement, characterization and application techniques which are used in attosecond science. The techniques allow for the access to attosecond phenomena in matter and utilize the attosecond pulses being reflected off and shaped by attosecond multilayer mirrors, described in the previous sections. Multilayer mirrors for attosecond pulses are therefore the key components in those experiments gaining deep physical insight into the dynamics of matter. This paves the way toward eventually controlling chemical or biological processes in the future.

1.6.1. Streaking Measurements

The last step after the generation and shaping of attosecond pulses is their measurement and characterization. Indirect measurement techniques must be used as no electronics are fast enough to map their temporal structure. One way for the pulse characterization is to cross-correlate the photoemission of the EUV/soft x-ray pulse (pump) with the electric field of the driving laser (probe). The driving laser pulse is used as a probe, since it is intrinsically synchronized with the EUV pulse by the HHG process. This cross-correlation results in a time delay dependent spectral photoelectron trace which can be used to reconstruct both pulses. This method is called attosecond streak camera or atomic transient recorder [74]. Both the EUV and laser beam are focused into an interaction medium (gas, solid, liquid) where the EUV ionizes the atoms. The ionized electrons are freed and inherit all the characteristics of the attosecond pulse regarding energy, pulse duration and phase nevertheless with a shifted kinetic energy, i.e. reduced by the binding energy W_b of the original electron level (e.g. 21.6 eV for the Ne 2p electrons):

$$W_{kin}(t_0) = \hbar\omega - W_b. \quad (1.32)$$

After the ionization, the electrons are accelerated by the electric field $E(t)$ of the laser resulting in a time dependent momentum:

$$p(t) = p(t_0) - e \int_{t_0}^t E(\tau) d\tau = p(t_0) - eA(t_0) + eA(t), \quad (1.33)$$

where t_0 is the time of ionization, A the vector potential and $p(t_0) = \sqrt{2mW_{kin}(t_0)}$ the instant momentum. After the laser has passed, the electron's final momentum is only dependent on the vector potential at the time of its ionization, since $A(t=\infty) = 0$ when the laser is no longer present at $t = \infty$:

$$p(\infty) = p(t_0) - eA(t_0). \quad (1.34)$$

Measuring the electron spectrum for different time delays between the laser and the attosecond pulse results in a streaking trace like shown in Fig. 1.15. This spectrogram allows for a complete retrieval of both the vector potential of the laser and the characteristic of the attosecond pulse by FROG/CRAB [76, 78, 86, 91]. The analysis

reveals the intensity as well as nonlinear phase terms (e.g. a possible chirp) of the attosecond pulse thus a complete reconstruction of the temporal structure. Summarized, attosecond streaking is sensitive not only to the EUV spectrum, but also to its phase which allows fully characterizing the EUV pulse. As the vector potential of the laser pulse is measured as well, it can also be completely characterized.

1.6.2. Transient Absorption Measurements

Attosecond transient absorption spectroscopy (ATAS) is an extension of absorption spectroscopy. Ultrafast transient absorption spectroscopy, an example of non-linear spectroscopy, measures changes in the absorbance/transmittance in a sample. The sample is excited by a pump pulse (UV–NIR) and the corresponding change in transmittance is probed by an ultrafast attosecond pulse as a function of the temporal delay of both pulses [92]. By means of highly intense few-cycle infrared pulses one modifies for example the electrical conductivity of a dielectric. The strong electric field of the laser changes the population of the conduction band and transfers the dielectric from an insulator to a conductor [93]. The population of the valence and conduction band influences the photoemission of the core states and changes consequently the absorption property of the material in the EUV range at element-specific absorption edges. This measurement technique enabled for the first time the observation of the motion of a superposition of valence electrons in krypton in real time on the attosecond timescale [32]. Applying this pump-probe method e.g. to semiconductors, electron-hole pairs are generated by the laser pulse and the occupancy of connected valence and conduction bands states is probed by the attosecond EUV pulse. The desired electron-hole pair density can be generated by tuning the laser intensity. The underlying dynamics are imprinted in the probing EUV absorption spectrum, since the absorption of the EUV photon is sensitive to the conduction band population via the interaction of the EUV excited electron with electrons injected by the pump pulse. Introducing a controlled time delay between the probe and pump pulse reveals information about the filling/emptying rate of the states by the probe [94]. This ensures a fundamental insight into the predominant carrier dynamics.

1.6.3. The RABBITT Technique

A second technique for the characterization of attosecond pulses is usually referred to as RABBITT (Reconstruction of Attosecond harmonic Beating By Interference of Two-photon Transitions) [63, 90]. It is known that HHG occurs when an intense laser pulse is focused into a gas target. If the focused IR laser pulse consists now of several cycles, in contrast to close to single cycle in Section 1.6.1 (e.g. $\Delta\tau = 30$ fs versus $\Delta\tau = 5$ fs), the emitted harmonics will also be confined in short attosecond pulses, which are continuously generated every half cycle of the IR pulse, but now resulting in an attosecond pulse train (APT) with a strongly modulated EUV spectrum. The number of electric field cycles increases with longer pulses, i.e. more attosecond pulses are generated. The interference of a higher number of pulses is responsible for the strongly modulated spectrum in contrast to a superimposed plateau spectrum for the case of few cycle driver pulses. These APTs can be characterized by RABBITT. Spectrally and temporally filtered

EUV-APTs are focused into a noble gas which gets ionized. The freed electrons exhibit kinetic energies which reflect the corresponding high harmonic photon energies (separated by $2\omega_0$). Sidebands peaks will appear in the electron energy spectrum if a moderately intense IR field is also focused into the noble gas. These sidebands are based on two-photon two-color ionization by one EUV photon and one infrared laser photon. Varying the time delay between the EUV and the IR modulates the sideband signal which originates from the interference between two processes: absorption of an EUV photon and an IR photon ($q\omega_{\text{IR}} + \omega_{\text{IR}}$), and absorption of an EUV photon and emission of an IR photon ($((q+2)\omega_{\text{IR}} - \omega_{\text{IR}})$) [63]. As the width and especially the relative temporal position of a sideband maximum is determined by the phase difference between two neighboring fundamental harmonic peaks, only the oscillation of the high harmonic ionization signal in the delay dependent electron energy spectrum contains the exact spectral phase information of the EUV pulses. This allows a precise characterization of the attosecond pulse trains by RABBITT. However, the RABBITT technique will probably be limited to the lower attosecond energy regime as few cycle laser pulses, exhibiting a high electric field amplitude, are needed to push attosecond science to higher photon energies (e.g. the water window) in the near future.

1.7. Outlook

Big achievements have already been made in terms of the performance of multilayer mirrors for attosecond pulses, but further optimization and application are possible. The realization of the first chirped multilayer mirrors for the visible spectral energy range and their first proof of successful femtosecond pulse compression was the key to implement the idea behind those mirrors in today's high-power laser systems and to make them to working horses in laser technology, since chirped laser amplification would be impossible without such mirrors. The realization of chirped mirrors for attosecond pulses in the EUV and even soft x-ray energy range was possible by fully controlling atomic layer-by-atomic layer deposition and may also open doors to key technologies and new exciting experiments like the mirrors for the visible range already do today.

1.7.1. The Way Toward Ever Shorter Pulses - Approaching the Atomic Unit of Time

The cornerstone of attosecond physics was the key for ever shorter pulses, since the ultrafast femtosecond pulses were superseded after several decades by attosecond pulses. The first experiment for the generation of isolated pulses has proven a temporal duration of 650 attoseconds [1] and lots of subsequent experiments have realized even shorter pulses, resulting in the nowadays shortest flashes of light below 100 attosecond, isolated attosecond pulses with a temporal duration of 67 attoseconds [95] and 80 attoseconds [68].

With the investigation of ever more intense laser systems (HHG cutoff energy: \sim Intensity) and in parallel going toward laser systems with longer wavelengths (HHG cutoff energy: $\sim\lambda^2$) an ever higher cutoff energy, up to the keV, is achieved [96] and as a consequence broader spectra are accessible and therefore shorter pulses can in principle be achieved. A time scale often aimed and given is the time it takes an electron to complete one rad of its

orbit in the ground state of hydrogen [97]: $\tau_{\text{atomic}} = \hbar/2U_{\text{H}} \approx 24$ as (often called the atomic unit of time), where U_{H} is the binding energy of the hydrogen atom (13.6 eV). The key is multicolor high-peak-power excitation in the HHG process [98]. So, the HHG field is aiming toward high-energy few-cycle femtosecond mid-IR lasers as prospective driver sources for strong-field applications, offering a high photon flux in a broad energy window and the opportunity for ever shorter time resolutions, approaching the atomic unit of time. Very recently the first accuracy below the attosecond time scale was achieved. Scientists succeeded in determine the photoemission time in helium with an accuracy of 0.85 attoseconds or 850 zeptoseconds [99].

1.7.2. Multilayer Mirrors for HHG Photon Energies up to the keV Spectral Range

The previous section described the way toward ever higher photon energies in HHG but neglected one important parameter: the high photon number needed for low-cross-section attosecond experiments. Consequently, the sources should offer a high photon number $>10^{10}$ in the soft x-ray range and multilayer mirrors, being used for beam guiding, beam shaping or beam focusing, should ensure a high photon throughput for high attosecond pulse intensities in the target to be investigated. The way toward higher photon energies combined with a high photon flux leaves only one way open for multilayer mirrors: toward grazing incidence optics. Grazing incidence optics are the key for photon energies above the water window (> 543.1 eV= oxygen 1s) up to the keV due to four distinct reasons:

- Eq. (1.16) demonstrates a high single interface reflectivity for grazing incidence.
- The Bragg Eq. (1.20) shows the necessity for grazing optics in case of the keV: A photon energy of 1000 eV corresponds to a wavelength of $\lambda = 1.2398$ nm, which leads in case of normal incidence to a $\lambda/2$ period in a multilayer mirror. Assuming a typical two material systems one ends up in a layer thickness of ~ 0.3 nm, which is close to the limit of technical realization.
- The roughness Eq. (1.28) demonstrates how one can minimize the influence of the roughness on the reflectivity performance: The lower the σ/d -ratio the lower is the reflectivity loss. One thus can minimize the reflectivity loss by reducing the interface roughness σ or increase the period thickness d . Keeping the center energy from a multilayer mirror constant leads in case of a higher period thickness to a lower grazing incidence angle. This overlaps with the previous reason. Typical roughness values are in the range of 0.4-0.6 nm, so comparable thickness values are strongly not recommended.
- The bandwidth Eq. (1.22) shows the advantage for attosecond pulses. In case of grazing incidence only few layers can contribute to the overall reflectivity which results at a certain energy in a higher bandwidth support. According to Eq. (1.23) this offers the opportunity for shorter pulses.

1.8. Conclusions

Multilayer mirrors investigations over the last decades paved their application way from the visible/infrared to the EUV/soft x-ray spectral range in terms of spectral characteristic and from the picosecond/femtosecond down to the attosecond regime in terms of temporal consideration. Those mirrors are nowadays already used by default in attosecond science in the EUV range, thanks to the big improvements being realized by an optimized combination of deposition, simulation and characterization methods. Due to their high degree of freedom multilayer mirrors will remain a keycomponent in attosecond beamlines as pulse shaping element and further investigations will open the way toward attosecond physics at ever higher spectral ranges and with ever shorter pulses.

Acknowledgements

We strongly acknowledge scientific support by Michael Hofstetter (MPQ, LMU) and we thankfully acknowledge scientific support and valuable discussions by Ferenc Krausz (MPQ, LMU). This work was financially supported by the DFG (German Research Foundation) via the Excellence Cluster ‘Munich Centre for Advanced Photonics’ (MAP, EXC158). Alexander Guggenmos acknowledges personal support from the German Research Foundation.

References

- [1]. M. Hentschel, R. Kienberger, Ch. Spielmann, G. A. Reider, N. Milosevic, T. Brabec, P. Corkum, U. Heinzmann, M. Drescher, F. Krausz, Attosecond metrology, *Nature*, Vol. 414, 2001, pp. 509-513.
- [2]. C. Morawe, M. Osterhoff, Hard x-ray focusing with curved reflective multilayers, *X-Ray Optics and Instrumentation*, Vol. 2010, 2010, pp. 1-8.
- [3]. J. Schmidt, A. Guggenmos, M. Hofstetter, S. H. Chew, U. Kleineberg, Generation of circularly polarized high harmonic radiation using a transmission multilayer quarter waveplate, *Optics Express*, Vol. 23, 2015, pp. 33564-33578.
- [4]. M. Zürch, J. Rothhardt, S. Hädrich, S. Demmler, M. Krebs, J. Limpert, A. Tünnermann, A. Guggenmos, U. Kleineberg, C. Spielmann, Real-time and sub-wavelength ultrafast coherent diffraction imaging in the extreme ultraviolet, *Scientific Reports*, Vol. 4, 2014, 7356.
- [5]. E. Spiller, Reflective multilayer coatings for the far UV region, *Applied Optics*, Vol. 15, 1976, pp. 2333-2338.
- [6]. S. Bajt, J. B. Alameda, J. Barbee, W. Troy, W. M. Clift, J. A. Folta, B. Kaufmann, E. A. Spiller, Improved reflectance and stability of Mo-Si multilayers, *Optical Engineering*, Vol. 41, 2002, pp. 1797-1804.
- [7]. I. V. Kozhevnikov, L. L. Balakireva, A. I. Fedorenko, I. A. Kopealets, V. E. Levashov, A. N. Stetsenko, I. I. Struk, A. V. Vinogradov, Synthesis and measurement of Os-Si multilayer mirrors optimized for the wavelength 380 Å, *Optics Communications*, Vol. 125, 1996, pp. 13-17.
- [8]. S. Rajesh, V. Arivazhagan, M. M. Parvathi, Structural, optical and electrical properties of vacuum evaporated PbSe/ZnSe multilayer thin films, *AIP Conference Proceedings*, Vol. 1451, 2012, pp. 197-199.

- [9]. K. M. Skulina, C. S. Alford, R. M. Bionta, D. M. Makowiecki, E. M. Gullikson, R. Soufli, J. B. Kortright, J. H. Underwood, Molybdenum/beryllium multilayer mirrors for normal incidence in the extreme ultraviolet, *Applied Optics*, Vol. 34, 1995, pp. 3727-3730.
- [10]. S. Bajt, D. G. Stearns, P. A. Kearney, Investigation of the amorphous-to-crystalline transition in Mo/Si multilayers, *Journal of Applied Physics*, Vol. 90, 2001, pp. 1017-1025.
- [11]. P. N. Rao, M. Nayak, G. S. Lodha, S. K. Rai, A. K. Srivastava, M. H. Modi, A. Sagdeo, Fabrication and evaluation of large area Mo/Si soft x-ray multilayer mirrors at Indus SR facilities, *Advances in Optical Technologies*, Vol. 2012, 2012, pp. 1-8.
- [12]. D. L. Voronov, P. Gawlitza, R. Cambie, S. Dhuey, E. M. Gullikson, T. Warwick, S. Braun, V. V. Yashchuk, H. A. Padmore, Conformal growth of Mo/Si multilayers on grating substrates using collimated ion beam sputtering, *Journal of Applied Physics*, Vol. 111, 2012, 093521.
- [13]. A. El Hajj, B. Lucas, M. Chakaroun, R. Antony, B. Ratier, M. Aldissi, Optimization of ZnO/Ag/ZnO multilayer electrodes obtained by Ion Beam Sputtering for optoelectronic devices, *Thin Solid Films*, Vol. 520, 2012, pp. 4666-4668.
- [14]. M. Grigonis, É. J. Knystautas, C/Si multilayer mirrors for the 25–30-nm wavelength region, *Applied Optics*, Vol. 36, 1997, pp. 2839-2842.
- [15]. Z. Wang, H. Zhao, Q. Yao, J. Xu, H. Kimura, Structure and magnetism of ZnO/Co multilayers prepared by pulsed laser deposition, *Crystal Research and Technology*, Vol. 47, 2012, pp. 799-803.
- [16]. K. Siraj, M. Khaleeq-ur Rahman, M. S. Rafique, M. Z. Munawar, S. Naseem, S. Riaz, Pulsed laser deposition and characterization of multilayer metalcarbon thin films, *Applied Surface Science*, Vol. 257, 2011, pp. 6445-6450.
- [17]. D. B. Chrisey, G. K. Hubler, Pulsed Laser Deposition of Thin Films, *Wiley*, 1994.
- [18]. D. Martínez-Galarce, R. Soufli, D. L. Windt, M. Bruner, E. Gullikson, S. Khatri, E. Spiller, J. C. Robinson, S. Baker, E. Prast, Multisegmented, multilayer-coated mirrors for the solar ultraviolet imager, *Optical Engineering*, Vol. 52, 2013, 095102.
- [19]. C. Wagner, N. Harned, EUV lithography: Lithography gets extreme, *Nature Photonics*, Vol. 4, 2010, pp. 24-26.
- [20]. W. Chao, J. Kim, S. Rekawa, P. Fischer, E. H. Anderson, Demonstration of 12 nm resolution fresnel zone plate lens based soft x-ray microscopy, *Optics Express*, Vol. 17, 2009, pp. 17669-17677.
- [21]. J.-Ph. Champeaux, Ph. Troussel, B. Villier, V. Vidal, T. Khachroum, B. Vidal, M. Krumrey, Development and realization of non-periodic W/Si multilayer mirrors for 5–14 keV x-ray plasma diagnostic, *Nuclear Instruments and Methods in Physics Research Section A*, Vol. 581, 2007, pp. 687-694.
- [22]. N. V. Andronova, V. G. Kohn, A. I. Chechin, Multilayer mirrors as synchrotron radiation monochromators, *Nuclear Instruments and Methods in Physics Research Section A*, Vol. 359, 1995, pp. 131-134.
- [23]. M. Ferray, A. L'Huillier, X. F. Li, L. A. Lompre, G. Mainfray, C. Manus, Multiple-harmonic conversion of 1064 nm radiation in rare gases, *Journal of Physics B: Atomic, Molecular and Optical Physics*, Vol. 21, 1988, pp. L31-L35.
- [24]. M. Prasciolu, A. F. G. Leontowich, K. R. Beyerlein, S. Bajt, Thermal stability studies of short period Sc/Cr and Sc/B4C/Cr multilayers, *Applied Optics*, Vol. 53, 2014, pp. 2126-2135.
- [25]. W. Helml, A. R. Maier, W. Schweinberger, I. Grguras, P. Radcliff, G. Doumy, C. Roedig, J. Gagnon, M. Messerschmidt, S. Schorb, C. Bostedt, F. Grüner, L. F. DiMauro, D. Cubaynes, J. D. Bozek, T. Tschentscher, J. T. Costello, M. Meyer, R. Coffee, S. Düsterer, A. L. Cavalieri, R. Kienberger, Measuring the temporal structure of few-femtosecond free-electron laser x-ray pulses directly in the time domain, *Nature Photonics*, Vol. 8, 2014, pp. 950-957.
- [26]. M. F. Kling, M. J. J. Vrakking, Attosecond electron dynamics, *Annual Review of Physical Chemistry*, Vol. 59, 2008, pp. 463-492.

- [27]. M. Fieß, M. Schultze, E. Goulielmakis, B. Dennhardt, J. Gagnon, M. Hofstetter, R. Kienberger, F. Krausz, Versatile apparatus for attosecond metrology and spectroscopy, *Review of Scientific Instruments*, Vol. 81, 2010, 093103.
- [28]. W. Schweinberger, A. Sommer, E. Bothschafter, J. Li, F. Krausz, R. Kienberger, M. Schultze, Waveform-controlled near-single-cycle milli-joule laser pulses generate sub-10 nm extreme ultraviolet continua, *Optics Letters*, Vol. 37, 2012, pp. 3573-3575.
- [29]. P. Corkum, Plasma perspective on strong-field multiphoton ionization, *Physical Review Letters*, Vol. 71, 1993, pp. 1994-1997.
- [30]. D. Descamps, C. Lyngå, J. Norin, A. L'Huillier, C.-G. Wahlström, J.-F. Hergott, H. Merdji, P. Salières, M. Bellini, T. W. Hänsch, Extreme ultraviolet interferometry measurements with high-order harmonics, *Optics Letters*, Vol. 25, 2000, pp. 135-137.
- [31]. J. Lee, D. R. Carlson, R. J. Jones, Optimizing intracavity high harmonic generation for XUV fs frequency combs, *Optics Express*, Vol. 19, 2011, pp. 23315-23326.
- [32]. E. Goulielmakis, Z.-H. Loh, A. Wirth, R. Santra, N. Rohringer, V. S. Yakovlev, S. Zherebtsov, T. Pfeifer, A. M. Azzeer, M. F. Kling, S. R. Leone, F. Krausz, Real-time observation of valence electron motion, *Nature*, Vol. 466, 2010, pp. 739-743.
- [33]. M. Nisoli, P. Decleva, F. Calegari, A. Palacios, F. Martín, Attosecond electron dynamics in molecules, *Chemical Review*, 2017, pp. 10760-10825.
- [34]. S. Neppl, R. Ernstorfer, A. L. Cavalieri, C. Lemell, G. Wachter, E. Magerl, E. M. Bothschafter, M. Jobst, M. Hofstetter, U. Kleineberg, J. V. Barth, D. Menzel, J. Burgdörfer, P. Feulner, F. Krausz, R. Kienberger, Direct observation of electron propagation and dielectric screening on the atomic length scale, *Nature*, Vol. 517, 2015, pp. 342-346.
- [35]. S. H. Chew, F. Süßmann, C. Späth, A. Wirth, J. Schmidt, S. Zherebtsov, A. Guggenmos, A. Oelsner, N. Weber, J. Kapaldo, A. Gliserin, M. I. Stockman, M. F. Kling, U. Kleineberg, Time-of-flight-photoelectron emission microscopy on plasmonic structures using attosecond extreme ultraviolet pulses, *Applied Physics Letters*, Vol. 100, 2012, 051904.
- [36]. Z. Tao, C. Chen, T. Szilvasi, M. Keller, M. Mavrikakis, H. Kapteyn, M. Murnane, Direct time-domain observation of attosecond final-state lifetimes in photoemission from solids, *Science*, Vol. 353, 2016, pp. 62-67.
- [37]. D. Attwood, Soft X-Rays and Extreme Ultraviolet Radiation: Principles and Applications, *Cambridge University Press*, 2007.
- [38]. B. L. Henke, E. M. Gullikson, J. C. Davis, X-ray interactions: photoabsorption, scattering, transmission, and reflection at $E=50\text{-}30000$ eV, $Z=1\text{-}92$, *Atomic Data and Nuclear Data Tables*, Vol. 54, 1993, pp. 181-342.
- [39]. E. Spiller, Soft X-Ray Optics, *SPIE Optical Engineering Press*, 1994.
- [40]. M. Trubetskov, A. Tikhonravov, V. Pervak, Time-domain approach for designing dispersive mirrors based on the needle optimization technique. Theory, *Optics Express*, Vol. 16, 2008, pp. 20637-20647.
- [41]. V. Pervak, A. V. Tikhonravov, M. K. Trubetskov, S. Naumov, F. Krausz, A. Apolonski, 1.5-octave chirped mirror for pulse compression down to sub-3 fs, *Applied Physics B*, Vol. 87, 2006, pp. 5-12.
- [42]. A. V. Tikhonravov, M. K. Trubetskov, G. W. DeBell, Optical coating design approaches based on the needle optimization technique, *Applied Optics*, Vol. 46, 2007, pp. 704-710.
- [43]. V. Yakovlev, G. Tempea, Optimization of chirped mirrors, *Applied Optics*, Vol. 41, 2002, pp. 6514-6520.
- [44]. V. Pervak, C. Teisset, A. Sugita, S. Naumov, F. Krausz, A. Apolonski, High-dispersive mirrors for femtosecond lasers, *Optics Express*, Vol. 16, 2008, pp. 10220-10233.
- [45]. M. Schultze, M. Fieß, N. Karpowicz, J. Gagnon, M. Korbman, M. Hofstetter, S. Neppl, A. L. Cavalieri, Y. Komninos, T. Mercouris, C. A. Nicolaides, R. Pazourek, S. Nagele, J. Feist, J. Burgdörfer, A. M. Azzeer, R. Ernstorfer, R. Kienberger, U. Kleineberg,

- E. Goulielmakis, F. Krausz, V. S. Yakovlev, Delay in photoemission, *Science*, Vol. 328, 2010, pp. 1658-1662.
- [46]. J. H. Underwood, T. W. Barbee, Layered synthetic microstructures as Bragg diffractors for x-rays and extreme ultraviolet: theory and predicted performance, *Applied Optics*, Vol. 20, 1981, pp. 3027-3034.
- [47]. L. G. Parratt, Surface studies of solids by total reflection of x-rays, *Physical Review*, Vol. 95, 1954, pp. 359-369.
- [48]. A. L. Aquila, F. Salmassi, F. Dollar, Y. Liu, E. Gullikson, Developments in realistic design for aperiodic Mo/Si multilayer mirrors, *Optics Express*, Vol. 14, 2006, pp. 10073-10078.
- [49]. U. Kleineberg, H.-J. Stock, A. Klödt, B. Schmiedeskamp, U. Heinzmann, S. Hopfe, R. Scholz, Interface stability and silicide formation in high temperature stable MoSi_{1-x}/Si multilayer soft x-ray mirrors studied by means of x-ray diffraction and HRTEM, *Physica Status Solidi (A)*, Vol. 145, 1994, pp. 539-550.
- [50]. P. Debye, Interferenz von röntgenstrahlen und wärmebewegung, *Annalen der Physik*, Vol. 348, 1913, pp. 49-92.
- [51]. I. Waller, Zur frage der einwirkung der wärmebewegung auf die interferenz von röntgenstrahlen, *Zeitschrift für Physik*, Vol. 17, 1923, pp. 398-408.
- [52]. L. Nénot, P. Croce, Caractérisation des surfaces par réflexion rasante de rayons x. Application à l'étude du polissage de quelques verres silicates, *Revue de Physique Appliquée*, Vol. 15, 1980, pp. 761-779.
- [53]. E. Chi, J. Shim, J. Kwak, H. Baik, Silicide formation by solid-state diffusion in MO/Si multilayer thin films, *Journal of Materials Science*, Vol. 31, 1996, pp. 3567-3572.
- [54]. M. Hofstetter, A. Aquila, M. Schultze, A. Guggenmos, S. Yang, E. Gullikson, M. Huth, B. Nickel, J. Gagnon, V. S. Yakovlev, E. Goulielmakis, F. Krausz, U. Kleineberg, Lanthanum-molybdenum multilayer mirrors for attosecond pulses between 80 and 130 eV, *New Journal of Physics*, Vol. 13, 2011, 063038.
- [55]. E. Gerald, J. Jellison, Data Analysis for Spectroscopic Ellipsometry, in Handbook of Ellipsometry, (F. A. Jenkins, H. E. White, Ed.), *William Andrew Publishing*, 2005, pp. 237-296.
- [56]. E. M. Gullikson, S. Mrowka, B. B. Kaufmann, Recent developments in EUV reflectometry at the advanced light source, *Emerging Lithographic Technologies V, Proceedings of SPIE*, Vol. 4343, 2001, pp. 363-373.
- [57]. S. Singh, S. Basu, P. Bhatt, A. K. Poswal, Kinetics of alloy formation at the interfaces in a Ni-Ti multilayer: x-ray and neutron reflectometry study, *Physical Review B*, Vol. 79, 2009, 195435.
- [58]. A. Guggenmos, S. Radünz, R. Rauhut, M. Hofstetter, S. Venkatesan, A. Wochnik, E. M. Gullikson, S. Fischer, B. Nickel, C. Scheu, U. Kleineberg, Ion polished Cr/Sc attosecond multilayer mirrors for high water window reflectivity, *Optics Express*, Vol. 22, 2014, pp. 26526-26536.
- [59]. F. Hamelmann, G. Haindl, J. Schmalhorst, A. Aschentrup, E. Majkova, U. Kleineberg, U. Heinzmann, A. Klipp, P. Jutzi, A. Anopchenko, M. Jergel, S. Luby, Metal oxide/silicon oxide multilayer with smooth interfaces produced by in situ controlled plasma-enhanced MOCVD, *Thin Solid Films*, Vol. 358, 2000, pp. 90-93.
- [60]. R. L. Thornton, R. D. Burnham, W. Streifer, High reflectivity GaAs-AlGaAs mirrors fabricated by metalorganic chemical vapor deposition, *Applied Physics Letters*, Vol. 45, 1984, pp. 1028-1030.
- [61]. P. B. Corkum, F. Krausz, Attosecond science, *Nature Physics*, Vol. 3, 2007, pp. 381-387.
- [62]. F. Krausz, M. Ivanov, Attosecond physics, *Review of Modern Physics*, Vol. 81, 2009, pp. 163-234.
- [63]. R. López-Martens, K. Varjú, P. Johnsson, J. Mauritsson, Y. Mairesse, P. Salières, M. B. Gaarde, K. J. Schafer, A. Persson, S. Svanberg, C.-G. Wahlström, A. L'Huillier,

- Amplitude and phase control of attosecond light pulses, *Physical Review Letters*, Vol. 94, 2005, 033001.
- [64]. G. Sansone, E. Benedetti, F. Calegari, C. Vozzi, L. Avaldi, R. Flammini, L. Poletto, P. Villoresi, C. Altucci, R. Velotta, S. Stagira, S. De Silvestri, M. Nisoli, Isolated single-cycle attosecond pulses, *Science*, Vol. 314, 2006, pp. 443-446.
 - [65]. T. Rohwer, S. Hellmann, M. Wiesenmayer, C. Sohrt, A. Stange, B. Slomski, A. Carr, Y. Liu, L. M. Avila, M. Kallane, S. Mathias, L. Kipp, K. Rossnagel, M. Bauer, Collapse of long-range charge order tracked by time-resolved photoemission at high momenta, *Nature*, Vol. 471, 2011, pp. 490-493.
 - [66]. A.-S. Morlens, R. López-Martens, O. Boyko, P. Zeitoun, P. Balcou, K. Varjú, E. Gustafsson, T. Remetter, A. L'Huillier, S. Kazamias, J. Gautier, F. Delmotte, M.-F. Ravet, Design and characterization of extreme-ultraviolet broadband mirrors for attosecond science, *Optics Letters*, Vol. 31, 2006, pp. 1558-1560.
 - [67]. Y. A. Uspenskii, V. E. Levashov, A. V. Vinogradov, A. I. Fedorenko, V. V. Kondratenko, Y. P. Pershin, E. N. Zubarev, S. Mrowka, F. Schäfers, Sc-Si normal incidence mirrors for a VUV interval of 35-50 nm, *Nuclear Instruments and Methods in Physics Research Section A*, Vol. 448, 2000, pp. 147-151.
 - [68]. E. Goulielmakis, M. Schultze, M. Hofstetter, V. S. Yakovlev, J. Gagnon, M. Uiberacker, A. L. Aquila, E. M. Gullikson, D. T. Attwood, R. Kienberger, F. Krausz, U. Kleineberg, Single-cycle nonlinear optics, *Science*, Vol. 320, 2008, pp. 1614-1617.
 - [69]. A. Guggenmos, A. Akil, M. Ossiander, M. Schäffer, A. M. Azzeer, G. Boehm, M.-C. Amann, R. Kienberger, M. Schultze, U. Kleineberg, Attosecond photoelectron streaking with enhanced energy resolution for small-bandgap materials, *Optics Letters*, Vol. 41, 2016, pp. 3714-3717.
 - [70]. C. Bourassin-Bouchet, S. de Rossi, F. Delmotte, Multilayer Mirrors for Coherent Extreme-Ultraviolet and Soft X-ray Sources, in *Optical Technologies for Extreme-Ultraviolet and Soft X-ray Coherent Sources* (F. Canova, L. Poletto, Ed.), Springer Series in Optical Sciences, Vol. 197, Springer, 2015.
 - [71]. M. Hofstetter, M. Schultze, M. Fieß, B. Dennhardt, A. Guggenmos, J. Gagnon, V. S. Yakovlev, E. Goulielmakis, R. Kienberger, E. M. Gullikson, F. Krausz, U. Kleineberg, Attosecond dispersion control by extreme ultraviolet multilayer mirrors, *Optics Express*, Vol. 19, 2011, pp. 1767-1776.
 - [72]. Y. Ménesguen, S. de Rossi, E. Meltchakov, F. Delmotte, Aperiodic multilayer mirrors for efficient broadband reflection in the extreme ultraviolet, *Applied Physics A*, Vol. 98, 2010, pp. 305-309.
 - [73]. E. M. Gullikson, J. H. Underwood, P. Batson, V. Nikitin, A soft x-ray/EUV reflectometer based on a laser produced plasma source, *Journal of X-Ray Science and Technology*, Vol. 3, 1992, pp. 283-299.
 - [74]. R. Kienberger, E. Goulielmakis, M. Uiberacker, A. Baltuska, V. Yakovlev, F. Bammer, A. Scrinzi, T. Westerwalbesloh, U. Kleineberg, U. Heinzmann, M. Drescher, F. Krausz, Atomic transient recorder, *Nature*, Vol. 427, 2004, pp. 817-821.
 - [75]. M. Kitzler, N. Milosevic, A. Scrinzi, F. Krausz, T. Brabec, Quantum theory of attosecond XUV pulse measurement by laser dressed photoionization, *Physical Review Letters*, Vol. 88, 2002, 173904.
 - [76]. R. Trebino, K. W. DeLong, D. N. Fittinghoff, J. N. Sweetser, M. A. Krumbügel, B. A. Richman, D. J. Kane, Measuring ultrashort laser pulses in the time-frequency domain using frequency-resolved optical gating, *Review of Scientific Instruments*, Vol. 68, 1997, pp. 3277-3295.
 - [77]. Y. Mairesse, A. de Bohan, L. J. Frasinski, H. Merdji, L. C. Dinu, P. Monchicourt, P. Breger, M. Kovačev, R. Taïeb, B. Carré, H. G. Muller, P. Agostini, P. Salières, Attosecond synchronization of high-harmonic soft x-rays, *Science*, Vol. 302, 2003, pp. 1540-1543.

- [78]. J. Gagnon, E. Goulielmakis, V. S. Yakovlev, The accurate FROG characterization of attosecond pulses from streaking measurements, *Applied Physics B*, Vol. 92, 2008, pp. 25-32.
- [79]. A. Guggenmos, M. Jobst, M. Ossiander, S. Radünz, J. Riemensberger, M. Schäffer, A. Akil, C. Jakubeit, P. Böhm, S. Noever, B. Nickel, R. Kienberger, U. Kleineberg, Chromium/scandium multilayer mirrors for isolated attosecond pulses at 145 eV, *Optics Letters*, Vol. 40, 2015, pp. 2846-2849.
- [80]. A. Guggenmos, R. Rauhut, M. Hofstetter, S. Hertrich, B. Nickel, J. Schmidt, E. M. Gullikson, M. Seibald, W. Schnick, U. Kleineberg, Aperiodic CrSc multilayer mirrors for attosecond water window pulses, *Optics Express*, Vol. 21, 2013, pp. 21728-21740.
- [81]. C. Montcalm, B. T. Sullivan, S. Duguay, M. Ranger, W. Steffens, H. Pépin, M. Chaker, In situ reflectance measurements of soft-x-ray/extreme-ultraviolet Mo/Y multilayer mirrors, *Optics Letters*, Vol. 20, 1995, pp. 1450-1452.
- [82]. R. Soufli, E. Spiller, D. L. Windt, J. C. Robinson, E. M. Gullikson, L. Rodriguez-de Marcos, M. Fernandez-Perea, S. L. Baker, A. L. Aquila, F. J. Dollar, J. A. Méndez, J. I. Larruquert, L. Golub, P. Boerner, In-band and out-of-band reflectance calibrations of the EUV multilayer mirrors of the atmospheric imaging assembly instrument aboard the Solar Dynamics Observatory, *Space Telescopes and Instrumentation 2012: Ultraviolet to Gamma Ray Proceedings of SPIE*, Vol. 8443, 2012, 84433C.
- [83]. S. Bajt, Molybdenum-ruthenium/beryllium multilayer coatings, *Journal of Vacuum Science & Technology A*, Vol. 18, 2000, pp. 557-559.
- [84]. B. Sae-Lao, C. Montcalm, Molybdenum-strontium multilayer mirrors for the 8-12-nm extreme-ultraviolet wavelength region, *Optics Letters*, Vol. 26, 2001, pp. 468-470.
- [85]. C. Montcalm, P. A. Kearney, J. M. Slaughter, B. T. Sullivan, M. Chaker, H. Pépin, C. M. Falco, Survey of Ti-, B-, and Y-based soft x-ray-extreme ultraviolet multilayer mirrors for the 2- to 12-nm wavelength region, *Applied Optics*, Vol. 35, 1996, pp. 5134-5147.
- [86]. J. Gagnon, V. S. Yakovlev, The robustness of attosecond streaking measurements, *Optics Express*, Vol. 17, 2009, pp. 17678-17693.
- [87]. S. Neopl, R. Ernstorfer, E. M. Bothschafter, A. L. Cavalieri, D. Menzel, J. V. Barth, F. Krausz, R. Kienberger, P. Feulner, Attosecond time-resolved photoemission from core and valence states of magnesium, *Physical Review Letters*, Vol. 109, 2012, 087401.
- [88]. A. Aquila, F. Salmassi, E. Gullikson, Metrologies for the phase characterization of attosecond extreme ultraviolet optics, *Optics Letters*, Vol. 33, 2008, pp. 455-457.
- [89]. C. Bourassin-Bouchet, S. de Rossi, J. Wang, E. Meltchakov, A. Giglia, N. Mahne, S. Nannarone, F. Delmotte, Shaping of single-cycle sub-50-attosecond pulses with multilayer mirrors, *New Journal of Physics*, Vol. 14, 2012, 023040.
- [90]. H. G. Muller, Reconstruction of attosecond harmonic beating by interference of two-photon transitions, *Applied Physics B*, Vol. 74, 2002, pp. 17-21.
- [91]. Y. Mairesse, F. Quéré, Frequency-resolved optical gating for complete reconstruction of attosecond bursts, *Physical Review A*, Vol. 71, 2005, 011401.
- [92]. X. Wang, M. Chini, Y. Cheng, Y. Wu, X.-M. Tong, Z. Chang, Subcycle laser control and quantum interferences in attosecond photoabsorption of neon, *Physical Review A*, Vol. 87, 2013, 063413.
- [93]. M. Schultze, E. Bothschafter, A. Sommer, S. Holzner, W. Schweinberger, M. Fieß, M. Hofstetter, R. Kienberger, V. Apalkov, V. Yakovlev, M. I. Stockman, F. Krausz, Controlling dielectrics with the electric field of light, *Nature*, Vol. 493, 2013, pp. 75-78.
- [94]. E. Tea, H. Hamzeh, F. Aniel, Hot carriers relaxation in highly excited polar semiconductors: Hot phonons versus phonon-plasmon coupling, *Journal of Applied Physics*, Vol. 110, 2011, 113108.
- [95]. K. Zhao, Q. Zhang, M. Chini, Y. Wu, X. Wang, Z. Chang, Tailoring a 67 attosecond pulse through advantageous phase-mismatch, *Optics Letters*, Vol. 37, 2012, pp. 3891-3893.

- [96]. T. Popmintchev, M.-C. Chen, D. Popmintchev, P. Arpin, S. Brown, S. Ališauskas, G. Andriukaitis, T. Balciunas, O. D. Mücke, A. Pugzlys, A. Baltuška, B. Shim, S. E. Schrauth, A. Gaeta, C. Hernández-García, L. Plaja, A. Becker, A. Jaron-Becker, M. M. Murnane, H. C. Kapteyn, Bright coherent ultrahigh harmonics in the keV x-ray regime from mid-infrared femtosecond lasers, *Science*, Vol. 336, 2012, pp. 1287-1291.
- [97]. P. H. Bucksbaum, The future of attosecond spectroscopy, *Science*, Vol. 317, 2007, pp. 766-769.
- [98]. S. Haessler, T. Balčiūnas, G. Fan, L. E. Chipperfield, A. Baltuška, Enhanced multi-colour gating for the generation of high-power isolated attosecond pulses, *Scientific Reports*, Vol. 5, 2015, 10084.
- [99]. M. Ossiander, F. Siegrist, V. Shirvanyan, R. Pazourek, A. Sommer, T. Latka, A. Guggenmos, S. Nagele, J. Feist, J. Burgdörfer, R. Kienberger, M. Schultze, Attosecond correlation dynamics, *Nature Physics*, Vol. 13, 2017, pp. 280-285.

Chapter 2

Moiré Effect in 3D Structures

Vladimir Saveljev

2.1. Introduction

The moiré effect is physical effect known from the optics. Involved items are periodic transparent structures (gratings, grids, nets, meshes) with close geometrical characteristics (period, and orientation) viewed through [1]. The moiré effect is a phenomenon of the linear (geometric) optics; the typical period of the gratings is about several mm or cm (at least longer than 1 mm), i.e., much longer than the wavelength of visible light. For the moiré effect, the periodical structure of the layers is important; meshes, nets, and gratings are examples of such layers.

The moiré patterns appear as a series of repeated stripes with their period shorter than the period of either gratings. Because their period differs from the period of gratings, these patterns could look strange, unexpected, sometimes even unpredictable.

The moiré effect was discovered and initially observed under the visible light. According to [2], the first known scientific description of the moiré patterns is credited to Lord Rayleigh, 1874. For an excellent introduction to the moiré effect, refer to [1-5], and references therein. The perfectly illustrated book on the variety of the moiré patterns is published recently [6].

The moiré effect represents an interaction (interference) between superposed periodic transparent layers. Such interaction can be described with using the rays (not waves!) which pass through the gratings. The mathematical representation of such interaction is the multiplication of the transparency/reflectance functions of layers.

The moiré effect in planar flat (2D) structures is investigated most; especially in the cases, when the planar gratings are parallel to each other [7-12]. Another well-developed case is a planar grating combined with a curved surface [4, 13, 14].

Because of the unexpected appearance of the moiré patterns, the moiré effect in the modern digital world is an unwanted guest in digital photographs [15], scans [16] and even in TV news sometimes. For instance, in 3D displays, the minimization of the moiré patterns is necessary [17-19]. At the same time, the practical applications of the moiré effect are not unknown in measurements [2, 5, 14, 20], in alignment [21], in visual security [22-24], in art [25], and even in 3D displays [26].

Besides, the rays of different nature can also demonstrate the moiré effect; examples are the ultraviolet light [27], the infrared light [28, 29], the electron beams [30, 31], the X-rays [32, 33], the ultrasound [34], and the microwaves [35]. Particularly, the moiré effect in planar graphene layers is well known and has been observed under the EM on multiple occasions, for example, in multilayer graphene [31], as well as in graphene on In [36] or on Ru [37]. The typical moiré period in the mentioned EM cases is about several nanometers. In each case, the transparency function should be re-considered correspondingly.

However, non-planar objects (like gridshells) also may show the moiré patterns. Moreover, 2D objects can be transformed into 3D sometimes; see examples shown in Section 2.2. A flat layer alone is not equally rigid in all directions and an external force may deform a flexible mesh, and bent (wrap, wriggle, or fold) it. Resulting is an essentially 3D object.

The observed moiré effect is attributed to layers; In this case, two halves into which a single-walled 3D object can be virtually split, as shown in Fig. 2.1. Since in the wrapped cylinder (as an example), these halves are physically made of the same mesh, the structures (namely, the periods) of both halves are definitely identical. Such object may model a hollow cylindrical nano-cell.

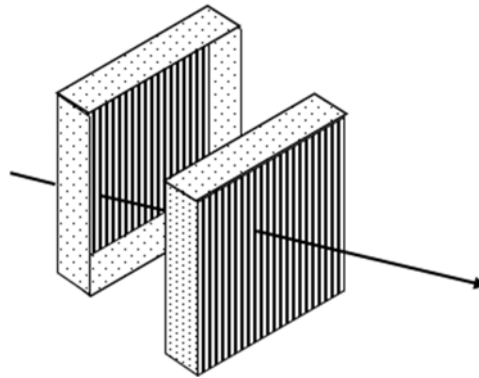


Fig. 2.1. Two logical halves of a 3D object.

After being transformed into a 3D shape, a flat mesh is significantly deformed. But it appears that in some special cases, a planar mesh can be bent without any other deformation except for the bending itself. For example, a metal mesh or a planar piece of paper can be bent into a cylindrical or conical spatial structure along the base and around the axis. As a result, there are no stretches and no wrinkles in such bent mesh at all. In this

case, the local geometric characteristics of the mesh (such as the period and therefore the wavenumber) across the cylindrical surface are the same as in the 2D case. Particularly, the period of the mesh does not change; thus, the wavenumber remains unchanged, and only the direction of the wavevector is changed due to bending (Fig. 2.2).



Fig. 2.2. Direction of the wavevector in planar (blue arrow on the left) and cylindrical (red arrows on the right) meshes.

Among other bent things, the cylindrical nanoparticles (essentially non-planar 3D objects) can be thought as wrapped graphene layers. The moiré effect in non-flat 2D particles can be observed under transmission electron microscope (TEM). For instance, the moiré effect has been observed in nanoparticles and nanotubes; for the moiré patterns in these nano-objects, refer to the famous papers [38-41].

In this chapter, we consider the moiré wavevector as a fundamental unit of the moiré effect. However, physically involved is the projected period, while the transformed wavevector is calculated via the projected period. Therefore firstly, we illustrate this approach to the flat (coplanar) gratings. Then we apply the concept of the transformed wavevector to the spatial (3D) cases. In the most scenarios, it is enough to calculate the period, because the orientation of the patterns is clear as often as not.

Section 2.2 (the moiré effect in the plain coplanar case) is mostly inspired by the papers published by the author previously. At the same time, most of Section 2.4 (2.4.2 and mainly 2.4.3) together with Section 2.5.2 present some new results and ideas.

The chapter is arranged as follows. In Section 2.2, various examples of the moiré effect around us are collected. In Section 2.3, we consider the projected period, the moiré wavevector and the spectral trajectories. Then, the moiré effect in the regular 3D bodies (the parallelepiped, the prism, and the cylinder) is considered in Section 2.4; for the moiré effect in the parallelepiped, we actually consider two parallel planes; for the prism, two crossed planes. Some interesting non-traditional issues of the moiré effect (statistics, lens, regular/irregular patterns) are presented in Section 2.5. Finally, the sections Discussion and Conclusion finalize the chapter.

2.2. Moiré Effect around Us

The moiré wavevector typically represents a new spatial frequency which is different from the wavevectors of either grating. The moiré patterns can be seen literally everywhere around us. Various examples are provided in this section. These photographs are taken by the author. The first series includes dress materials (fabrics), i.e., arbitrary surfaces of meshes and curtains (Figs. 2.3, 2.4).



Fig. 2.3. (a) Boryeong undersea tunnel (Daecheon, Korea);
(b) agricultural setup (Namyangju, Korea).



Fig. 2.4. (a) Yingyuan hotel (Shanghai, China); (b) Boryeong Jeju hotel (Seogwipo, Korea).

The period of the moiré patterns is longer than the periods of the gratings; sometimes the gratings themselves are invisible in the photographs, whereas the moiré patterns are clearly seen, see Figs. 2.3(b), 2.4(a), 2.4(b).

Although the moiré effect is basically originated from fabrics as shown in Figs. 2.3, 2.4, nowadays the moiré effect can be met in the metal too. For example, the moiré patterns are not unknown in architecture [42]. The photographed examples of the next series include buildings (Fig. 2.5), bridges (Fig. 2.6), and structural elements of buildings (Fig. 2.7).



Fig. 2.5. Galleria dept. store (Cheonan, Korea).

Then, the moiré effect in relatively simple regular geometric structures, such as parallelepiped (plain parallel layers), cylinders, and spheres is shown in Figs. 2.8-2.10. In the parallelepiped, we consider the moiré effect generated by two faces which are perpendicular to the axis of the camera. This is the last series of the presented photographs.

Somebody may think that the visible are necessary in the observation of the moiré effect. However, it is important to notice again that even if the structure of layers themselves is unrecognizable from certain (distant) observer position, the moiré effect can be clearly observed, see Figs. 2.3 (b), 2.4 (a), 2.4 (b), 2.6 (c), and 2.8 (a).

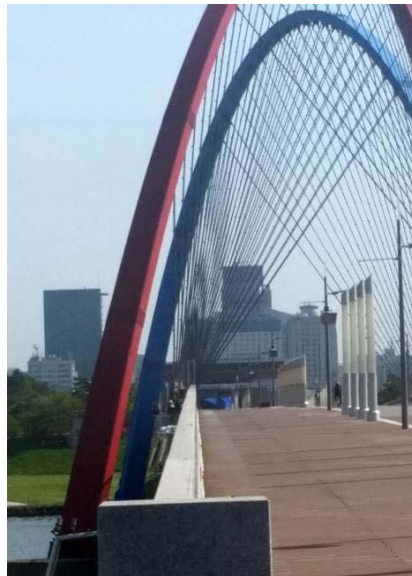
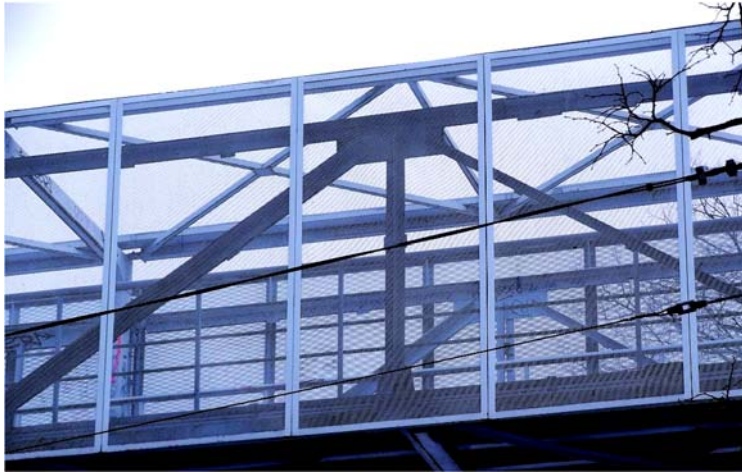


Fig. 2.6. Pedestrian bridges (a) Riga, Latvia; (b) Daejeon, Korea and (c) Namyangju, Korea.

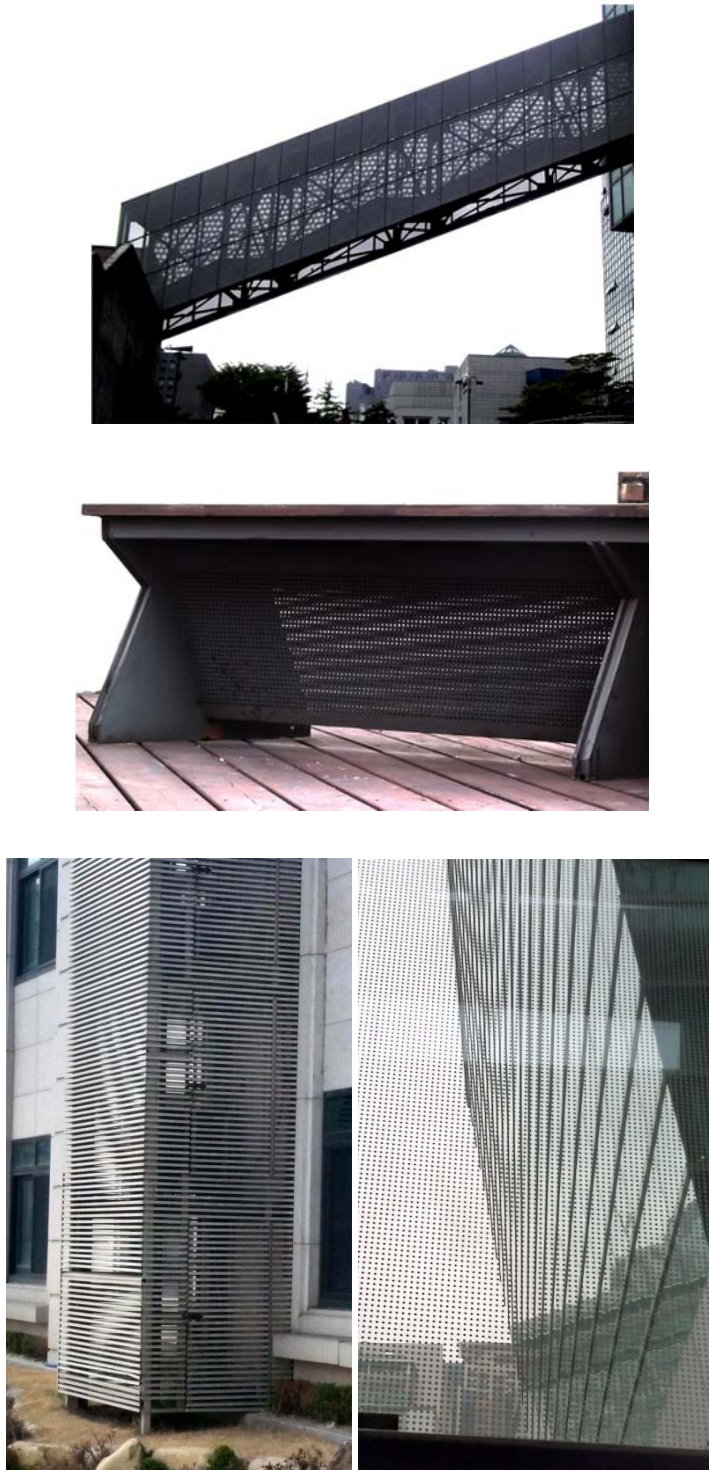


Fig. 2.7. (a) Soongsil University (Seoul, Korea); (b) Seoul National University (Seoul, Korea); (c) Hanyang University (Seoul, Korea) and (d) BEXCO (Busan, Korea).

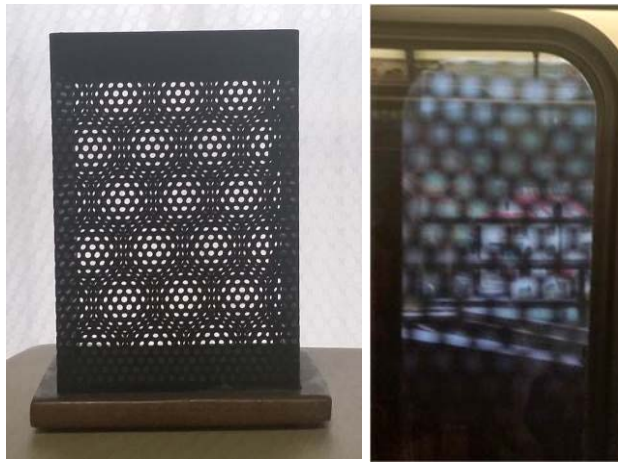


Fig. 2.8. (a) Stationary holder; (b) Everline train (Yongin, Korea).



Fig. 2.9. (a) Stationary holder, and (b) Boryeong Jeju hotel (Seogwipo, Korea).



Fig. 2.10. (a) Convention center (Daejeon, Korea), and (b) Intercity bus stn. (Boryong, Korea).

2.3. Plain Coplanar Case

2.3.1. Projected Mesh

Before everything else, consider a grating in space observed by an off-axis camera. This is a 2D case, when the lines of the gratings are perpendicular to the xz -plane.

With this aim in view, we apply the central projection to the grating. Consider the point A at (x, z) and its projection onto the x -axis. The central projection can be conveniently written with using the homogeneous vectors and matrices [43]. The 4D homogeneous vector of the point A is

$$x_A = (x \ 0 \ z \ 1)^T \quad (2.1)$$

The off-axis camera is located at $(w, 0, L, 1)^T$, the screen at the origin, see Fig. 2.11. Such layout corresponds to the focal distance of the camera equal to L .

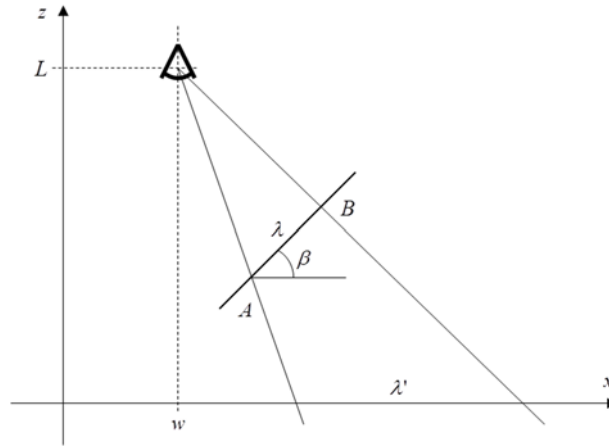


Fig. 2.11. Neighboring points of grating.

Particularly, according to [44] along with other sources, the homogeneous transformation matrix of this case is

$$M = \begin{pmatrix} 1 & 0 & -\frac{w}{L} & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & -\frac{1}{L} & 1 \end{pmatrix}. \quad (2.2)$$

Then the projection of the point A is

$$\begin{pmatrix} 1 & 0 & -\frac{w}{L} & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & -\frac{1}{L} & 1 \end{pmatrix} \begin{pmatrix} x \\ 0 \\ z \\ 1 \end{pmatrix} = \begin{pmatrix} x - w\frac{z}{L} \\ 0 \\ z \\ 1 - \frac{z}{L} \end{pmatrix}. \quad (2.3)$$

The projection transformation in the regular coordinates looks like follows,

$$x_A' = \frac{x - w\frac{z}{L}}{1 - \frac{z}{L}}. \quad (2.4)$$

Similarly, the projection of the neighboring point $B (x+\delta x, z+\delta z)$ is

$$x_B' = \frac{x + \delta x - w\frac{z + \delta z}{L}}{1 - \frac{z + \delta z}{L}}. \quad (2.5)$$

If the length of the displacement vector $(\delta x, \delta z)$ along the grating is one period $\lambda = (\delta x^2 + \delta y^2)^{1/2}$, then the projected period is the difference between the projected coordinates of the points A and B from Eqs. (2.4), (2.5) as follows

$$\lambda' = x_B' - x_A' = \frac{L}{L - z} \frac{(L - z)\delta x - (w - x)\delta z}{(L - z - \delta z)}. \quad (2.6)$$

Denote

$$\begin{aligned} \Delta x &= w - x \\ \Delta z &= L - z \end{aligned} \quad (2.7)$$

Then, the projected period

$$\lambda' = \frac{L}{\Delta z} \frac{\Delta z \delta x - \Delta x \delta z}{\Delta z - \delta z}. \quad (2.8)$$

Let

$$\begin{aligned} \delta x &= \lambda \cos \beta \\ \delta z &= \lambda \sin \beta \end{aligned} \quad (2.9)$$

where λ is the period and β is the inclination angle of the grating at the point A .

Then the projected period is

$$\lambda' = \lambda \frac{L}{\Delta z} \frac{\Delta z \cos \beta + \Delta x \sin \beta}{\Delta z - \lambda \sin \beta} = \lambda L \cos \beta \frac{1 + \frac{\Delta x}{\Delta z} \tan \beta}{\Delta z - \lambda \sin \beta}. \quad (2.10)$$

This formula of the 2D problem can be applied to calculate the projected period of any mesh/net at the point (x, z) with the period λ and the inclination angle β at that point, as it appears in the screen of a displaced (off-axis) camera.

A schematic illustration of the projected period of the inclined grating crossing the origin in the on-axis camera is shown in Fig. 2.12.

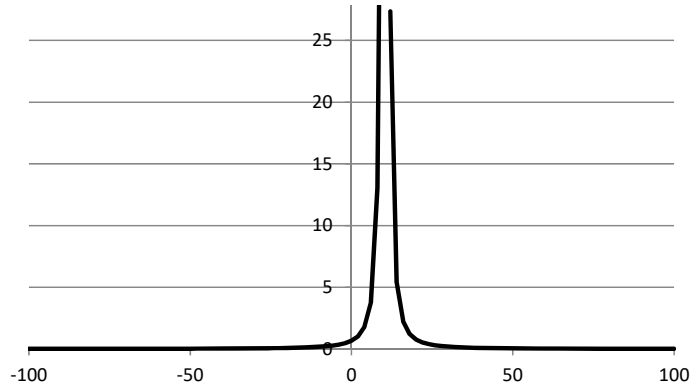


Fig. 2.12. Projected period of an inclined grating as a function of the lateral coordinate.

Particular cases.

a) When $w = 0$, then $\Delta x = -x$ and

$$\lambda' = \lambda \frac{L \cos \beta}{\Delta z} \frac{\Delta z + x \tan \beta}{\Delta z - \lambda \sin \beta}, \quad (2.11)$$

as in [45].

b) When the grating is parallel to xy -plane ($\beta = 0$), $\lambda' = \lambda \frac{L}{\Delta z}$, as known.

c) As far as $\lambda \ll z$, we may neglect the small term in the denominator of Eq. (2.10) and obtain the approximate formulas

$$\lambda' \approx \lambda \frac{L \cos \beta}{(\Delta z)^2} (\Delta z - \Delta x \tan \beta), \quad (2.12)$$

$$\lambda' \approx \lambda \frac{(L \cos \beta - w \sin \beta) \Delta z + \sin \beta (xL - wz)}{(\Delta z)^2}. \quad (2.13)$$

Based on Eq. (2.10), the wavevector (for sake of simplicity, without the coefficient 2π) is

$$k' = \frac{1}{L \cos \beta} \frac{k \Delta z - \sin \beta}{1 - \frac{\Delta x}{\Delta z} \tan \beta} = \frac{1}{L \cos \beta} \frac{k \Delta z - \sin \beta}{1 - \frac{w-x}{\Delta z} \tan \beta} = \frac{1}{L} \frac{\frac{k \Delta z}{\cos \beta} - \tan \beta}{1 + \frac{x}{\Delta z} \tan \beta - \frac{w}{\Delta z} \tan \beta} \quad (2.14)$$

Important is that two points involved in the initiation of the moiré effect (corresponding points) must lie on the same ray from the camera; this ray is defined by

$$\tan \alpha = \frac{x_1}{L - z_1} = \frac{x_2}{L - z_2} = \frac{x_1}{\Delta z_1} = \frac{x_2}{\Delta z_2} = \frac{x}{L}. \quad (2.15)$$

To take this into account, Eq. (2.14) can be rewritten in terms of the angle of sight, which is common for both gratings. Then we get

$$k'_1 = \frac{1}{L} \frac{\frac{k_1 \Delta z_1}{\cos \beta_1} - \tan \beta_1}{1 + \tan \alpha \tan \beta_1 - \frac{w}{\Delta z_1} \tan \beta_1}. \quad (2.16)$$

2.3.2. Moiré Wavevector

In the spatial domain, we deal with the linear quantities, such as the length and the size. When describing the theory of the moiré effect, it is however handier to consider reciprocal (inverse) units, such as inverse length and wavevector. Therefore, the moiré effect can be equivalently described in terms of the wavevector of the moiré patterns. (As mentioned above, we omit the factor 2π from the regular definition of the wave vector.)

For the moiré effect crucial is the layout of the spectral components of gratings, especially their relative configuration which generates the lowest spatial frequency.

This is the spectral domain, where the moiré effect is determined by the difference of the wavevectors obtained as the reciprocal projected periods in two corresponding points of the gratings (as described in Section 2.3.1),

$$k_m = k'_2 - k'_1. \quad (2.17)$$

Because of the symmetry of the Fourier transform, there always exist symmetric (positive and negative) spectral components arranged in pairs. Therefore, the moiré effect can be mathematically represented as a linear combination (with integer coefficients +1 and -1) of moduli of the components which closest to the origin, i.e. fall within the visibility circle [1]. In the simplest case, the combination is subtraction of the moduli of the spectral components.

Let k_1 and k_2 be the coplanar wavevectors and α the angle between them, see Fig. 2.13. Let's find the moiré wavevector k_m and the moiré angle θ , which completely characterize the sinusoidal wave (or the first harmonic of a non-sinusoidal wave).

Based on the geometry, it can be proven that

$$k_m = \sqrt{k_1^2 + k_2^2 - 2k_1k_2 \cos \alpha}, \quad (2.18)$$

$$\tan \theta = \frac{k_2 \sin \alpha}{k_1 - k_2 \cos \alpha}. \quad (2.19)$$

These formulas are not unknown, see, e.g., [1].

Eq. (2.19) represents the orientation of the moiré wavevector (its direction angle)). Together with (2.18), they comprise the full solution (the wavenumber and the angle).

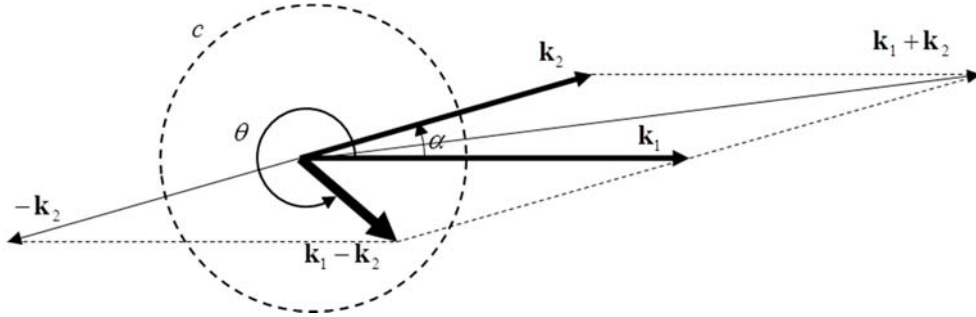


Fig. 2.13. Wavevectors of gratings and the moire wavevector. The visibility circle labeled by c is drawn by the dashed line.

The period is an inverse of Eq. (2.18) as follows,

$$p_m = \frac{1}{k_m} = p_2 \frac{1}{\sqrt{1 + \rho^2 - 2\rho \cos \alpha}}, \quad (2.20)$$

where $\rho = p_2/p_1$ is the ratio of periods.

The orientation Eq. (2.19) can be also rewritten in terms of ρ ,

$$\tan \theta = \frac{\sin \alpha}{\rho - \cos \alpha}. \quad (2.21)$$

The similar formulas can be found in [1, 2, 4, 5]. In this case of the linear proportionality to the period of the grating, Eq. (2.20) can be alternatively re-written as

$$p_m = \mu p_2, \quad (2.22)$$

where

$$\mu = \frac{1}{\sqrt{1 + \rho^2 - 2\rho \cos \alpha}}, \quad (2.23)$$

The expression Eq. (2.22) means some kind of magnification of the period of the grating by the coefficient Eq. (2.23), which means a moiré magnifier. For the concept of the moiré magnifier, refer to [46, 47].

The moiré effect is represented by waves with the longer period. Our assumption is that the patterns with the longer periods are better visible [18]. Correspondingly, the estimation of the visual appearance of moiré patterns is based on the wave numbers. This section is based on [48]. In this section, we mostly consider the moiré effect in the spectral domain.

It was already noted that the moiré wavevector is such a linear combination (sum or difference) of the wavevectors of the gratings. By extension, of the spectral components of gratings which has the lowest spatial frequencies (closest to the origin), and therefore fall within the visibility circle [1] in the Fourier domain, see Fig. 2.13. In the simplest case of close wavevectors, the linear combination which has the lowest modulus (= smallest wavenumber) is the difference of the spectral components of the gratings. Take a look to Fig. 2.13 again and try to choose the correct sign in an alternative case, when the wavevectors of the gratings are k_1 and $-k_2$.

In this section, we consider the functions with sparse and limited spectra. The former term (“sparse”) means that mostly the values of the most spectral components are close to zero (lie below a certain level) and therefore are negligible, except for a few relatively narrow and high peaks. One can say figuratively that the space under the spectral surface is almost “empty”. For example, a sparse matrix is a matrix populated primarily with zeros, which zeroes form a low-level background. The latter term (“limited”) means that instead of the infinite number of spectral peaks routinely required by the Fourier theory, we take into account only few of them, the peaks within certain interval around the origin. In this case, each and every combination of harmonics can be explicitly taken into account.

Besides, the multiplicative model of the superposition is implied, i.e., an interaction of gratings is modeled mathematically by the multiplication. The spectrum of superposed layers can be found from the convolution theorem which states that the Fourier transform

of the product is the convolution of the Fourier transforms of the individual functions [49]. Moreover, for a sparse spectrum, the convolution theorem is equivalent to the vector summation of a few spectral components.

For instance, there are exactly 3 spectral peaks for a sinusoidal grating, including DC and two symmetric peaks representing one wave. For a non-sinusoidal grating, there are $2N+1$ peaks. For instance, a 1D sinusoidal line grating (the intensity profile along certain direction is a sinusoidal function) can be represented by its transparency function $(1 + \cos \vec{k} \cdot \vec{x})/2$. Technically, k may have any sign, plus or minus: these two k 's formally represent the positive and negative components of the Fourier spectrum, which is always symmetric, and in certain sense, is redundant. Therefore, in this section we only consider one component; its counterpart located symmetrically can be easily obtained by the simple inversion of the sign. This makes spectral approach convenient for the limited spectra consisting of a finite number of spectral components (harmonics). As an additional feature, the two-dimensional spectrum can be conveniently represented in the complex plane.

The visible moiré wavevectors lie within the visibility circle [1] which roughly models the human visual system in the spectral domain, see Fig. 2.13, where the circle is denoted as c . Consider the axes u, v of the spectral domain as the real and imaginary axes in the complex plane. The real and imaginary parts of a complex number are u and v , resp. Then, all locations and shapes in the 2D spectrum and in the complex plane remain the same, but the calculations become much simpler and easier, because the complex numbers allows applying many powerful mathematical theorems.

Fig. 2.14 shows the map of spectral peaks of two overlapped 1D sinusoidal gratings (shown by filled triangles) corresponding to Fig. 2.13 and their linear combinations. The linear combinations of the wavevectors (shown by circles) with the smallest wavenumber is the moiré wavevector (filled circles).

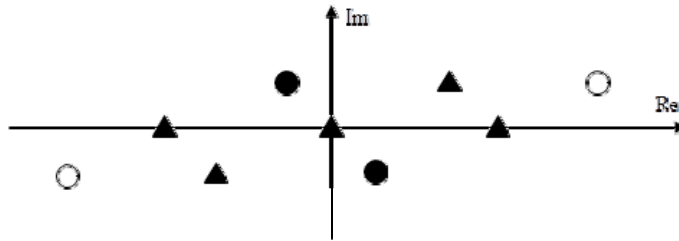


Fig. 2.14. All linear combinations of wavevectors of two overlapped sinusoidal gratings.

According to [48], the general equation of spectral peaks of the overlapped gratings in the complex plane is as follows

$$T = \sum_{n=1}^N p_n k_n e^{i\alpha_n}, \quad (2.24)$$

where N is the number of gratings, and the values of k_n , α_n , p_n , q_n are attributed to the n -th grating ($n = 1, \dots, N$) as follows: the two former values are the basic wavenumber and the rotation angle, while p_n is an integer number between $-q_n$ and $+q_n$.

The particular case of 4 gratings arranged in 2 layers (each layer consists of two orthogonal gratings) [48] is

$$T_{2 \times 2} = k(p_1 \sigma_1 + i p_2) + k \rho(p_3 \sigma_3 + i p_4) e^{i \alpha}, \quad (2.25)$$

where $k_1 = k \sigma_1$, $k_2 = k$, $k_3 = k \sigma_3$, $k_4 = k \rho$, $\alpha_1 = 0$, $\alpha_2 = \pi/2$, $\alpha_3 = \alpha$, $\alpha_4 = \alpha + \pi/2$.

This is a superposition of two rectangular (square) grids, which can be often met in practice, for example in digital 3D displays, where the pixel grid is combined with the lenticular/barrier plate.

2.3.3. Spectral Trajectories

Considered above is a picture of the moiré patterns, which takes place when all involved quantities are fixed and nothing is changed; such picture can be said “static”. However, parameters of the gratings may change sometimes. This changes the spectrum. In the case of a regular (incremental) change of a parameter, a set of correspondingly changed spectra represents a richer picture of the behavior of the patterns. For a sparse spectrum, such picture can be shown in a single picture. For example, when a 1D sinusoidal line grating is rotated (and the running parameter is the angle), its spectrum is also rotated by the same angle. Eight overlapped spectra for a few incrementally changed angles are shown in Fig. 2.15(a). This forms a discrete trajectory. A generalized picture can be schematically drawn by continuous “trajectories” of the peaks in the spectral domain, see Fig. 2.15(b), where these trajectories have the shape of circular arcs.

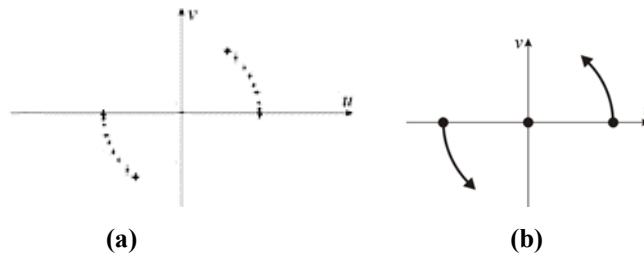


Fig. 2.15. Set of spectra of a rotated grating and the scheme of its spectra. The arrows in (b) show the extensions, when the parameter would keep changing.

A scheme of trajectories Fig. 2.15 (b) shows, where the spectral peaks are located when the angle varies between two limits. We refer such scheme to as spectral trajectory. The spectral trajectories are known in many areas [50, 51]. For the moiré effect, the spectral trajectories were first time proposed in [48].

In the case 2×2 , 4 types of trajectories can be derived from Eqs. (2.3-2.11) depending on the running variable (the angle, the size ratio, and two aspect ratios). These equations look like follows,

$$\begin{aligned} T_{2 \times 2 \alpha}(t) &= (p_1 \sigma_1 + ip_2)k + (p_3 \sigma_3 + ip_4)k \rho e^{i\alpha(t)} \\ T_{2 \times 2 \rho}(t) &= (p_1 \sigma_1 + ip_2)k + (p_3 \sigma_3 + ip_4)k \rho(t) e^{i\alpha} \\ T_{2 \times 2 \sigma_1}(t) &= (p_1 \sigma_1(t) + ip_2)k + (p_3 \sigma_3 + ip_4)k \rho e^{i\alpha} \\ T_{2 \times 2 \sigma_3}(t) &= (p_1 \sigma_1 + ip_2)k + (p_3 \sigma_3(t) + ip_4)k \rho e^{i\alpha} \end{aligned} \quad (2.26)$$

It can be proven that trajectories Eq. (2.26) are either circular arcs or segments of straight lines. The case of running angle $\alpha(t)$ in the first trajectory of Eq. (2.26) includes four subcases depending on the dimensionality of gratings (a line grating is a 1D structure; the square grid is a 2D structure)

$$\begin{aligned} T_{2 \times 2 \alpha 11}(t) &= p_1 \sigma_1 k + p_3 \sigma_3 k \rho e^{i\alpha(t)} \\ T_{2 \times 2 \alpha 12}(t) &= p_1 \sigma_1 k + (p_3 \sigma_3 + ip_4)k \rho e^{i\alpha(t)} \\ T_{2 \times 2 \alpha 21}(t) &= (p_1 \sigma_1 + ip_2)k + \sigma_3 k \rho e^{i\alpha(t)} \\ T_{2 \times 2 \alpha 22}(t) &= (p_1 \sigma_1 + ip_2)k + p_3 \sigma_3 k \rho e^{i\alpha(t)} \end{aligned} \quad (2.27)$$

We may refer the cases of two line gratings Eq. (2.11) and two square grids Eq. (2.22) to as 1D and 2D cases, resp.; while two remaining combinations Eqs. (2.12) and (2.21) can be treated as the intermediate ones.

Among all trajectories, there can be trajectories leaving the visibility circle, approaching, entering and crossing it, as well as the trajectories which are always outside or always inside the visibility circle. Such classification of the trajectories may essentially simplify the analysis.

Examples of the trajectories for the running angle are given in Fig. 2.16. These trajectories were observed experimentally [48].

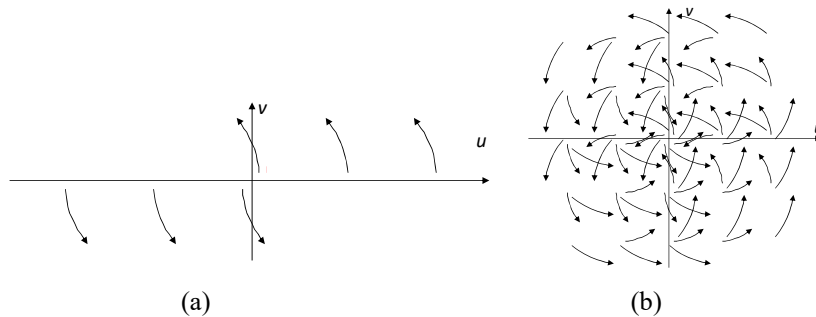


Fig. 2.16. Examples of trajectories Eq. (2.28) for sinusoidal gratings; (a) two line gratings, 0 (b) 2 square grids (running angle $\alpha [5^\circ, 40^\circ]$, $\rho = 1.1$, $\sigma_1 = \sigma_2 = 1$).

Examples of trajectories of the sinusoidal case for other running parameters are shown in Fig. 2.17. Many examples can be found in [52].

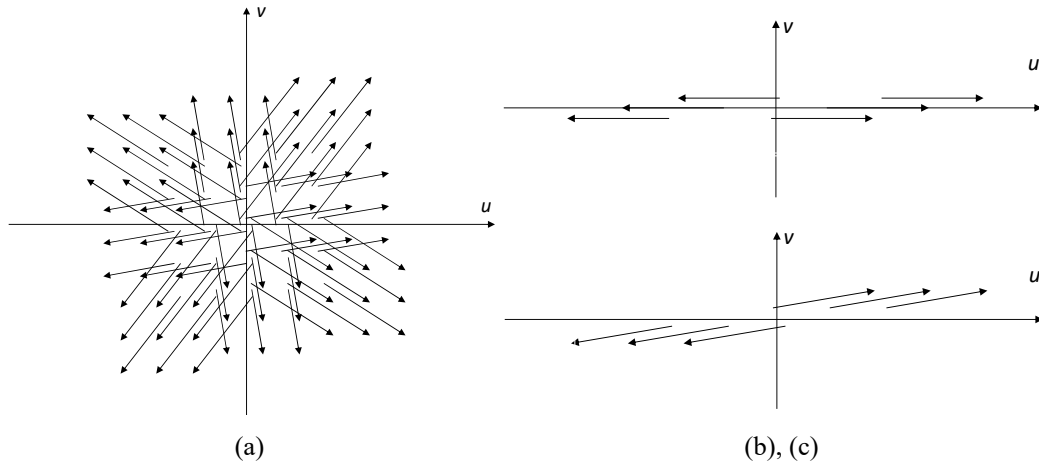


Fig. 2.17. Variety of spectral moiré trajectories (simulation) for two 1D and two 2D gratings:
 (a) varying ρ [1, 3], $\sigma_1 = \sigma_2 = 1.1$, $\alpha = 10^\circ$; (b) varying σ_1 [1, 3], $\rho = \sigma_2 = 1.1$, $\alpha = 10^\circ$;
 (c) varying σ_2 [1, 3], $\rho = \sigma_1 = 1.1$, $\alpha = 10^\circ$.

2.4. Moiré in Regular 3D Bodies

2.4.1. Parallelepiped (Parallel Planes)

Consider inclined gratings and the on-axis camera ($w = 0$, $\beta_1 = \beta_2$) as shown in Fig. 2.18.

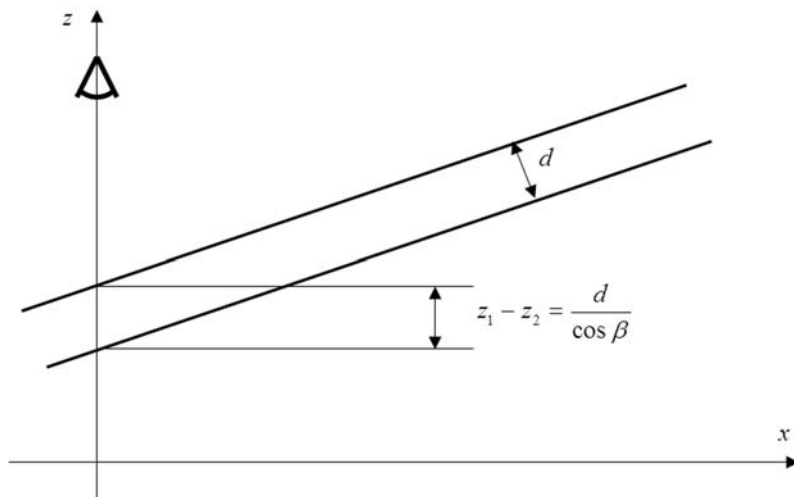


Fig. 2.18. Identical inclined planes.

For the identical gratings, we have from Eqs. (2.16), (2.17),

$$k_m = \frac{k}{L \cos \beta} \frac{\frac{d}{\cos \beta}}{1 + \tan \alpha \tan \beta} = \frac{k}{L \cos^2 \beta} \frac{d}{1 + \tan \alpha \tan \beta}. \quad (2.28)$$

When $\beta = 0$, the familiar formula appears

$$k_m = k \frac{d}{L}. \quad (2.29)$$

For the non-identical gratings parallel to the xy -plane ($\beta = 0$), we have from Eqs. (2.16), (2.17),

$$k_m = \frac{k_2 \Delta z_2 - k_1 \Delta z_1}{L}. \quad (2.30)$$

In the case of $z_1 = 0$, $z_2 = -d$, and $k_1 = \rho k_2$, it becomes

$$k_m = k_2 (s - \rho), \quad (2.31)$$

where $s = l + d/L$.

The wavevector Eq. (2.31) can be re-expressed in terms of k_1 as follows,

$$k_m = \frac{k_1}{\rho} (s - \rho) = k_1 \left(\frac{s}{\rho} - 1 \right), \quad (2.32)$$

or in terms of λ_1 ,

$$\lambda_m = \lambda_1 \frac{\rho}{s - \rho}, \quad (2.33)$$

which is exactly the same as Eq. (2.10) in [44].

The lateral displacement of the camera does not change the period in this case. An illustration is shown in Fig. 2.8. The spectral trajectories can be applied to this non-planar case as well.

In the identical gratings, the moiré mirror effect [44] can be observed. It means that the moiré patterns follow the lateral movements of an observer, as in a mirror.

2.4.2. Prism (Wedge)

The moiré effect in a pyramid observed through two bases does not differ from the moiré effect in the parallel planes (Section 2.4.1). In this section, we consider the pyramid observed through its sides, the inclined planes. Let's consider the symmetric identical gratings crossing the origin and the off-axis camera. Their equation is

$$\begin{aligned} z_1 &= a_1 x, \quad a_1 = \tan \beta_1 \\ z_2 &= a_2 x, \quad a_2 = \tan \beta_2' \end{aligned} \quad (2.34)$$

where $a_2 = -a_1$, $\beta_2 = -\beta_1$.

The ray from the camera goes at the angle α ,

$$\tan \alpha = \frac{x_1 - w}{L - z_1} = \frac{-\Delta x_1}{\Delta z_1} = \frac{x - w}{L}. \quad (2.35)$$

Then, from the general formula for the projected period Eq. (2.10), we have

$$\lambda_1' = \lambda_1 \cos \beta_1 \frac{1 + \tan \alpha \tan \beta_1}{\frac{\Delta z_1}{L} - \frac{\lambda_1}{L} \sin \beta_1}. \quad (2.36)$$

Now we have to express the specific z_1 of this grating using the common variable x which can be applied to both gratings.

This can be done for a known shape. In the current case of the straight line crossing the origin (as shown in Fig. 2.19(b)), we have from Eq. (2.35):

$$Lx - z_1 x + z_1 w = Lx_1. \quad (2.37)$$

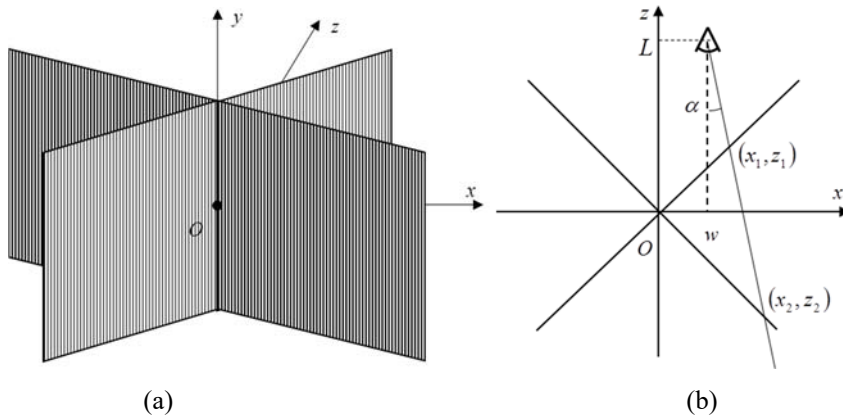


Fig. 2.19. Wedge (a) 3D view, and (b) cross-section.

Then we need the shape of the line, an equation connecting the variables z_1 and x_1 . Substituting Eq. (2.34),

$$x_1 = \frac{Lx}{L + a_1(x - w)}. \quad (2.38)$$

Thus

$$\frac{\Delta z_1}{L} = \frac{1 - \tan \beta_1 \frac{w}{L}}{1 + \tan \alpha \tan \beta_1}. \quad (2.39)$$

Therefore the period,

$$\lambda_1' = \lambda_1 \cos \beta_1 \frac{(1 + \tan \alpha \tan \beta_1)^2}{1 - \tan \beta_1 \frac{w}{L} - \frac{\lambda_1}{L} \sin \beta_1 (1 + \tan \alpha \tan \beta_1)}. \quad (2.40)$$

Then, the wavevector is

$$k' = \frac{1}{\cos \beta_1} \frac{k_1 \left(1 - \tan \beta_1 \frac{w}{L} \right) - \frac{1}{L} \sin \beta_1 (1 + \tan \alpha \tan \beta_1)}{(1 + \tan \alpha \tan \beta_1)^2}. \quad (2.41)$$

For the positive slope β (small angles), Eq. (2.40) represents the raising curve; for the negative slope β it is falling, see Fig. 2.20.

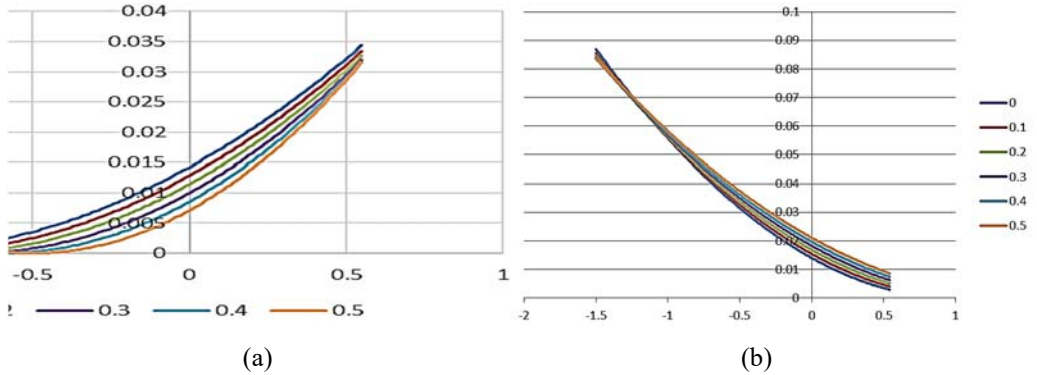


Fig. 2.20. The projected period for (a) positive, and (b) negative slope.

These monotonic curves must cross each other somewhere. At that point, the periods of gratings are equal, and the moiré period is theoretically infinite.

The relation $\tan \alpha = \frac{x-w}{L}$ makes the projected period Eq. (2.40) a function of x ,

$$\lambda_1' = \lambda_1 \cos \beta_1 \frac{\left(1 + \frac{x-w}{L} \tan \beta_1\right)^2}{1 - \tan \beta_1 \frac{w}{L} - \frac{\lambda_1}{L} \sin \beta_1 \left(1 + \frac{x-w}{L} \tan \beta_1\right)}. \quad (2.42)$$

It is zero, when

$$\tan \beta_1 = -\frac{1}{\tan \alpha}. \quad (2.43)$$

It is infinite, when

$$1 - \tan \beta_1 \frac{w}{L} - \frac{\lambda_1}{L} \sin \beta_1 (1 + \tan \alpha \tan \beta_1) = 0. \quad (2.44)$$

The coordinate of the point, where the period is infinite is smaller than the current x . (N.B. In the case of the identical parallel planes orthogonal to the LOS, the coordinate of that point was exactly equal to the current x ; it was called the moiré mirror [44].) Now that this point lies between the origin and the current coordinate.

2.4.3. Cylinder

Let 's take a look to another regular three-dimensional object, a single-layered cylinder made of mesh. In this section, we will calculate the period of the moiré patterns appearing in the screen S at the origin, when the rays pass through the cylinder. This section is mostly based on [45].

Consider the grating with the horizontal wavevector which lies in the xz -plane, and all involved vectors lie in the xz -plane too; therefore, such layout of vectors is planar, as shown in Fig. 2.21. The dotted circular arc shows the bent grating.

For a cylinder, $L = R \cos \psi$, $x = R \sin \psi$, $z = R \cos \psi$, and $\beta = -\psi$. Then, based on the geometry,

$$l_{1,2} = L \cos \alpha \pm \sqrt{L^2 \cos^2 \alpha - L^2 + R^2} = L \cos \alpha \pm \sqrt{R^2 - L^2 \sin^2 \alpha}. \quad (2.45)$$

Two roots Eq. (2.45) represent the intersections of the ray at the angle α with two halves of the cylinder (the lower sign corresponds to the front half, the upper to the rear). The two logical halves are illustrated in Fig. 2.1. Then,

$$\sin \psi_{1,2} = \left(\tilde{L} \cos \alpha \pm \sqrt{1 - \tilde{L}^2 \sin^2 \alpha} \right) \sin \alpha, \quad (2.46)$$

$$\cos \psi_{1,2} = \tilde{L} \sin^2 \alpha \mp \cos \alpha \sqrt{1 - \tilde{L}^2 \sin^2 \alpha}, \quad (2.47)$$

where $\tilde{L} = L/R$ is the relative distance.

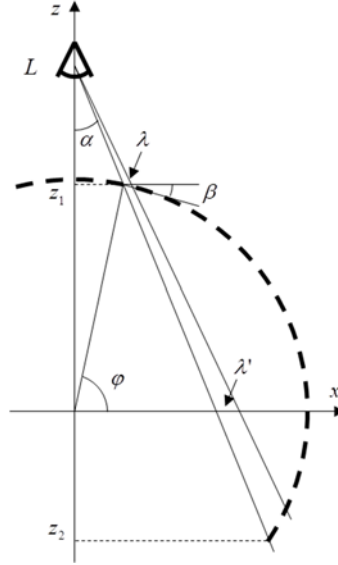


Fig. 2.21. Geometry of the horizontal cross-section of the cylinder.

Then from the general expression Eq. (2.11), we can find the projected period of the grating wrapped around the cylinder. Substituting Eqs. (2.46), (2.47) into Eq. (2.11), we can rewrite it in terms of the observation angle α ,

$$\tilde{\lambda}'_{1,2} = \tilde{\lambda} \frac{\bar{L}}{\cos \alpha} \cdot \frac{1}{(\tilde{L} \mp 1) \cos \alpha - \tilde{\lambda} (\tilde{L} \pm 1) \sin \alpha}, \quad (2.48)$$

where $\tilde{\lambda} = \lambda/R$ is the relative period and the temporary auxiliary variable

$$\tilde{L} = \frac{\bar{L} \cos \alpha}{\sqrt{1 - \bar{L}^2 \sin^2 \alpha}}.$$

The “projected” wavevector is

$$k'_H = k \frac{\cos \alpha}{\bar{L}} \left((\tilde{L} \mp 1) \cos \alpha - \tilde{\lambda} (\tilde{L} \pm 1) \sin \alpha \right). \quad (2.49)$$

Due to cyclic change of the wavevector along the surface of the cylinder, the wavevectors of gratings in two logical halves of the cylinder have the opposite orientations, see Fig. 2.2. Therefore, in this particular case of the moiré effect, the linear combination with the smallest wavenumber is the sum, and instead of the general Eq. (2.17) we add of the wavevectors Eq. (2.49) on the both sides of the cylinder. The resulting moiré wavevector is as follows

$$k_{MH} = 2 \frac{\cos \alpha}{\tilde{L}} (k \cos \alpha + \sin \alpha). \quad (2.50)$$

In bent cylinders, the relative period is always much less than one. In this case, the wavevector can be approximated as

$$k_{MH} \approx 2k \frac{\cos^2 \alpha}{\tilde{L}} \quad (2.51)$$

The corresponding approximate moiré magnification factor can be expressed as follows

$$\mu_H \approx \frac{\tilde{L}}{2 \cos^2 \alpha} \quad (2.52)$$

In the alternative layout, i.e., in the case of the vertical wavevector, the angle β is identically zero, but the definitions $x = R \sin \psi$ and $z = R \cos \psi$ remain the same as before. Then, we get the projected period of the grating from Eq. (2.11) as follows,

$$\lambda'_{1,2} = \lambda \frac{L}{\Delta z_{1,2}} = \lambda \frac{\tilde{L}}{\tilde{L} - \cos \psi_{1,2}} \quad (2.53)$$

Substituting Eqs. (2.47), (2.53) can be rewritten in terms of α , and the projected wavevectors are

$$k'_{1,2} = k \frac{\cos \alpha}{\tilde{L}} \left(\tilde{L} \cos \alpha \pm \sqrt{1 - \tilde{L}^2 \sin^2 \alpha} \right) \quad (2.54)$$

In the current case of the vertical wavevector, the linear combination with the smallest wavenumber is the difference; therefore, the moiré wavevector and the exact moiré factor are as follows,

$$k_{mV}' = k \frac{2 \cos \alpha}{\tilde{L}} \sqrt{1 - \tilde{L}^2 \sin^2 \alpha}, \quad (2.55)$$

$$\mu_V = \frac{1}{\sqrt{1 - \tilde{L}^2 \sin^2 \alpha}} \frac{\tilde{L}}{2 \cos \alpha} \quad (2.56)$$

The graphic illustration of the theoretical dependencies Eqs. (2.52) and (2.56) is given in Fig. 2.22.

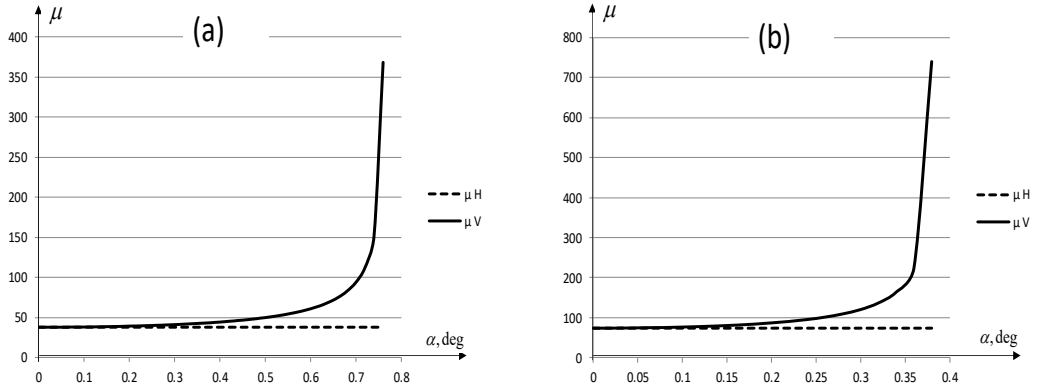


Fig. 2.22. Theoretical moiré factors for $\tilde{L} = 75$ in (a) and $\tilde{L} = 150$ in (b).

Whereas the horizontal component (responsible for the period of the vertical moiré lines) is practically constant, the vertical component (responsible for the period of the horizontal moiré lines) increases near the edge. As a result, the lines in a 2D picture are curved. A photographic illustrations of these cases are shown in Fig 2.9.

2.5. Interesting Moiré Issues

2.5.1. Moiré Statistics

Physical phenomena can be described in terms of probability. For the moiré effect, the probability can be estimated by counting the moiré cases. (For example, in respect to various gratings, we only consider the angles and ignore other particular characteristics.) The relative count of the successful cases (when the moiré patterns are visible) represents the probability for each angle.

This section is based on [53]. Define the rational angles as the angles satisfying the following equation,

$$\tan \alpha = \frac{m}{n}, \text{ m and n integers.} \quad (2.57)$$

The irrational angles do not satisfy Eq. (2.57).

The rational angles can be characterized by pairs of non-negative integer numbers (m, n) . In turn, such pairs represent points in the plane.

The rays with the rational angles Eq. (2.57) repeatedly pass through the nodes, i.e. points with the integer coordinates (m, n) , $(2m, 2n)$, $(3m, 3n)$, etc.; the rays with non-rational angles lie between the rational ones and never touch a single node.

A numerical characteristic how a ray crosses the nodes of the square grid is the distance between the nodes along that ray. The spacing of the nodes can be defined as a reciprocal distance between the nodes, i.e., the number of passed nodes per the length unit as follows,

$$D(x) = \begin{cases} \frac{1}{\sqrt{m^2 + n^2}}, & \text{for rational angles } (x = m/n), \\ 0, & \text{for irrational angles.} \end{cases} \quad (2.58)$$

The spacing as a function of an angle is shown in Fig. 2.23. This is a theoretical function which characterizes the angles in connection to the nodes of the square grid.

For small angles ($m \ll n$), the spacing function can be approximated by the known Thomae's function [54]. The locations of peaks of both functions coincide, although the values of functions (the heights of the peaks) are slightly different. Both mathematical functions are practically close to the experimental probability function shown in Fig. 2.24.

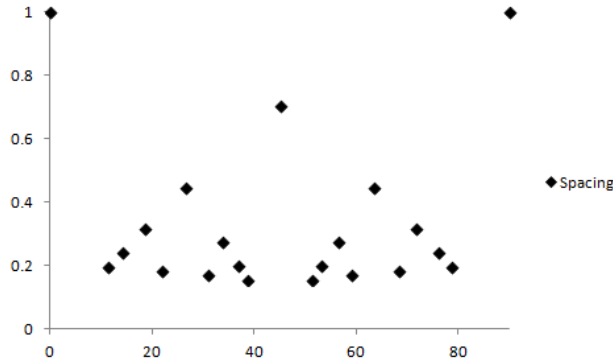


Fig. 2.23. Spacing of nodes (theoretical function) for $m, n \leq 5$.

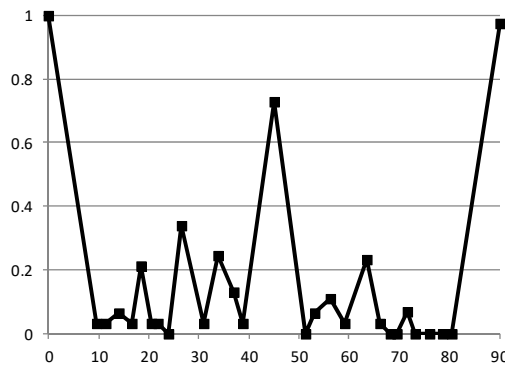


Fig. 2.24. Experimental moiré probability depending on angle.

In the moiré minimization, the statistical approach is helpful in avoiding the angles with the relatively high probability of the moiré effect (the rational angles with $m, n < 5$). Therefore in general, to eliminate the moiré effect in the square grids, the rational angles with small m, n should be avoided; whereas to observe the effect clearly, these rational angles are preferable. This is valid for cells with an arbitrary aspect ratio too. In the rectangular grids, m, n are counts of the rectangles, and therefore, the “rational” angles of this case would be $\tan \alpha = mX / nY$, where X and Y are the sides of the rectangular cell.

2.5.2. Moiré Lens

We already saw in this chapter that in many cases, the moiré period is proportional to the period of the grating and therefore can be expressed with using a proportionality coefficient, the magnification factor. In this section we formally compare the moiré magnification and the lens magnification.

From Eq. (2.33), one can derive the moiré factor for the non-coplanar gratings parallel to the xy plane as follows,

$$\mu^{eo} = \frac{1}{1 - \frac{s}{\rho}} \quad (2.59)$$

(Since the wavelength is always non-negative; therefore implied is the modulus in the denominator of the above expression).

On the other hand, the well-known formula for the magnification factor of the thin lens [55] $M = -d_i/d_o$ (where d_i and d_o are the object and image distances) can be rewritten in the alternative form using the lens equation as follows,

$$M = \frac{d_i}{d_o} = -\frac{1}{1 - \frac{d_o}{f}}, \quad (2.60)$$

where f is the focal length of the lens; the sign of the expression makes the difference between the real lens images and virtual ones.

The formal similarity of both formulas Eqs. (2.59), (2.60) is obvious. In this connection, it is not uninteresting to note that the ratio ρ in Eq. (2.59) could be an equivalent of the focal length of a “moiré lens”, while s could be treated as the distance between the “lens” and the object.

Consequently, the moiré interaction between the non-coplanar gratings with the parallel wavevectors could formally represent a lens, in which the ratio of periods ρ defines the focal length of such “moiré lens”, while the ratio of distances to the first and to the second gratings s defines the effective distance to the object. In other words, the inner structure (periods) of layers defines the focal length, while their layout (the distances) defines the

effective distance, as it is in any physical thin lens made of glass, for which Eq. (2.60) was derived.

2.5.3. Square and Octagon

An interesting unusual phenomenon can be observed, when a grating is rotated. As soon as the moiré patterns are combinational spatial frequencies of spectra of gratings, various geometrical shapes can be generally observed in grids and their spectra. For the identical square gratings which were considered in this section, these are the square and the octahedron. This section is based on [48].

The square can be observed in the spectrum of the superposed identical square grids near the angle 0° , see Fig. 2.25(a). The spectral square corresponds to the square grid in the spatial domain is shown in Fig 2.25(b). This is the strongest and most frequently observed moiré pattern in the square grids.

This square near the origin of the spectral domain consists of 8 spectral peaks which move together (4 in the corners and 4 in the middle of the sides of the square), see Fig. 2.25(a). When the angle between the gratings increases, all these 8 peaks rotate around the origin together in a coordinated (coherent) manner, so as the square is kept within a range of angles.

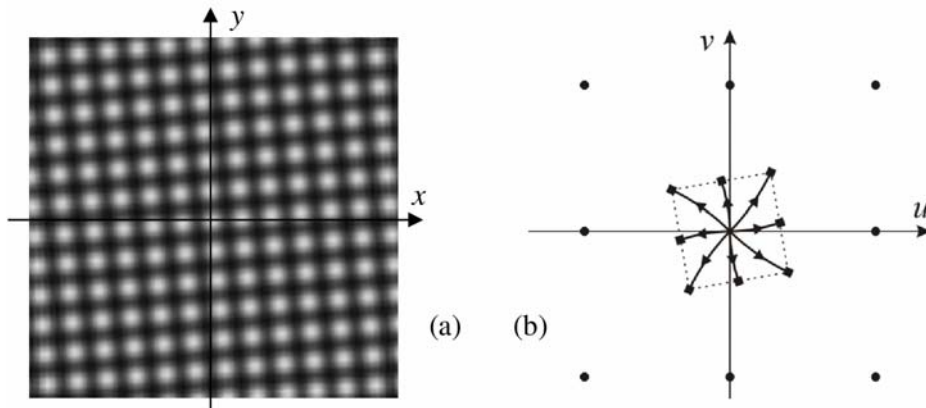


Fig. 2.25. Square moiré patterns: (a) in the spatial domain, and (b) in the spectral domain.

A counterpart of the square, an octagon in the spectrum can be observed when the angle between the gratings is 45° (or within a close neighborhood of this angle), see Fig. 2.26(a). This pattern is not commonly observed, because of a special combination of parameters. The spectral octagon consists of eight spectral peaks at the same distance from the origin of the spectral domain (these eight peaks are regularly distributed by the angle) arranged in 2 quadruplets, see Fig. 2.26. In contrast to the square, the octagonal relation between these two quadruplets is not kept in a wide angular range.

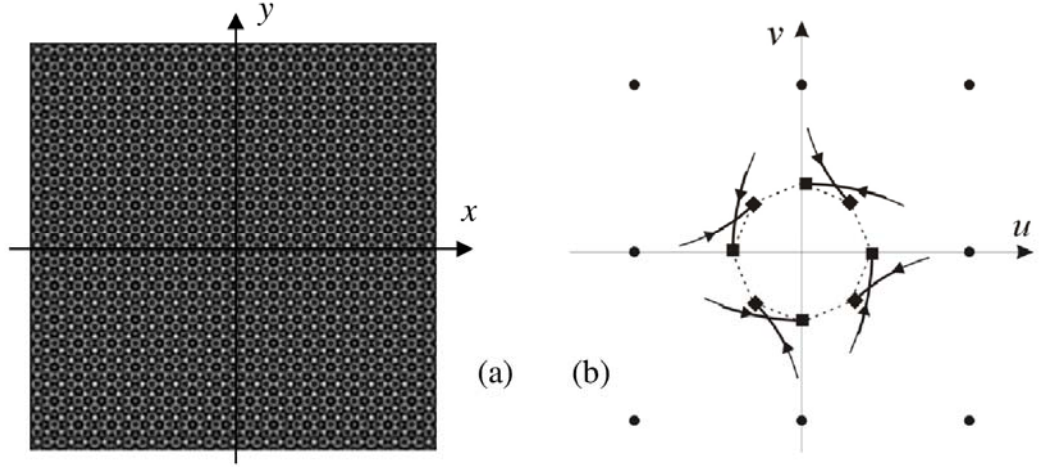


Fig. 2.26. (a) The repeated but non-periodic structure in the spatial domain;
(b) Its spectrum with the octagon inside the visibility circle.

When the angle between the gratings changes, each quadruplet rotates individually (in contrast to the square), so as the regular octagon only occurs at a particular angle 45° . The analytical equations for the spectral square and octagon in the closed form [48] are as follows,

$$\begin{aligned}
 S_1 &= i - i\rho e^{i\alpha}, \\
 S_2 &= (\sigma_1 + i) - (\sigma_3 + i)\rho e^{i\alpha}, \\
 S_3 &= \sigma_1 - \sigma_3\rho e^{i\alpha}, \\
 S_4 &= (\sigma_1 - i) - (\sigma_3 - i)\rho e^{i\alpha}, \\
 S_5 &= -i + i\rho e^{i\alpha}, \\
 S_6 &= -(\sigma_1 + i) + (\sigma_3 + i)\rho e^{i\alpha}, \\
 S_7 &= -\sigma_1 + \sigma_3\rho e^{i\alpha}, \\
 S_8 &= -(\sigma_1 - i) + (\sigma_3 - i)\rho e^{i\alpha},
 \end{aligned} \tag{2.61}$$

for the square and

$$\begin{aligned}
 O_1 &= i + (-1 - i)e^{i\alpha}, \\
 O_2 &= 1 + i - e^{i\alpha}, \\
 O_3 &= i + (-1 + i)e^{i\alpha}, \\
 O_4 &= 1 - i + e^{i\alpha}, \\
 O_5 &= -i + (1 + i)e^{i\alpha}, \\
 O_6 &= -1 - i + ie^{i\alpha}, \\
 O_7 &= -1 + (1 - i)e^{i\alpha}, \\
 O_8 &= -1 + i - ie^{i\alpha},
 \end{aligned} \tag{2.62}$$

for the octahedron.

It is important that in additional contrast to the square, the regular octagon does not correspond to any regular plain layout of regular polygons in the spatial domain, see Fig. 2.26 (b).

Therefore, a principal difference between these square and octagon is as follows. Whereas moiré patterns in gratings of the square case can be periodic, the patterns of the octagonal case cannot. Instead, they comprise octagons combined with other polygons in an irregular manner. This is connected to the topological properties of the Euclidean plane. This phenomenon is somehow similar to the pentagonal organic molecules [56] which also cannot fill the plane regularly.

2.6. Conclusion

In this chapter, the general equation for the transformed wavevector is obtained based on the projected period. The moiré wavevector is calculated as an additive combination of transformed wavevectors.

The following related topics are considered in the planar case: the projected period, the transformed wavevector, and the spectral trajectories of the moiré patterns. Then, this approach is generalized and applied to the moiré effect in the regular spatial structures such as parallelepiped, prism, and cylinder. The statistics, the lens, and the regular/irregular structures are also described. The similarity of the moiré lens to the thin lens may inspire some practically useful devices. The pseudo-random moiré patterns generated by two regular gratings can be used in optics, particularly in the visual security. All these issues together present a complex extendable view to the moiré effect in regular structures.

Results can be applied to the computer simulation and physical modeling of the moirés effect in nanoparticles (plain, cylindrical, and spherical), as well as in 3D displays.

References

- [1]. I. Amidror, Periodic Layers, Vol. 1, The Theory of the Moiré Phenomenon, 2nd ed., *Springer-Verlag*, London, 2009.
- [2]. H. Giger, The moirés, *Computers & Mathematics with Applications*, Vol. 12B, 1986, pp. 329-361.
- [3]. K. Patorsky, M. Kujawinska, Handbook of the Moiré Fringe Technique, *Elsevier*, 1993.
- [4]. K. Creath, J. C. Wyant, Optical Shop Testing, Chapter 16, in Moiré and Fringe Projection Techniques, *John Wiley & Sons*, 1992, pp. 653-685.
- [5]. C. A. Walker, Handbook of Moiré Measurement, *IOP Publishing*, 2004.
- [6]. N. Carsten, Moiré Index, *Gestalten Verlag*, Berlin, 2010.
- [7]. O. Bryngdahl, Characteristics of superposed patterns in optics, *Journal of the Optical Society of America*, Vol. 66, No 2, 1976, pp. 87-94.
- [8]. S. Yokozeki, Theoretical interpretation of the moiré pattern, *Optics Communications*, Vol. 11, No. 4, 1974, pp. 378-381.

- [9]. S. Yokozeki, Y. Kusaka, K. Paturski, Geometric parameters of moiré fringes, *Applied Optics*, Vol. 15, 1976, pp. 2223-2227.
- [10]. C. A. Sciammarella, Basic optical law in the interpretation of the moiré patterns applied to the analysis of strains, Part 1, *Experimental Mechanics*, Vol. 5, No. 5, 1965, pp. 154-160.
- [11]. C. A. Sciammarella, The moiré method - A review, *Experimental Mechanics*, Vol. 22, No. 11, 1982, pp. 418-433.
- [12]. P. A. M. dos Santos, Moiré-like patterns as a spatial beat frequency in photorefractive sinusoidal phase gratings superposition, *Optics Communications*, Vol. 212, No. 4-6, 2002, pp. 211-216.
- [13]. P. S. Theocaris, Moiré topography of curved surfaces, *Experimental Mechanics*, Vol. 7, No. 7, 1967, pp. 289-296.
- [14]. D. Post, B. Han, P. Ifju, High Sensitivity Moiré. Experimental Analysis for Mechanics and Materials, *Springer*, 1994.
- [15]. D. Blatner, C. Chavez, G. Fleishman, S. Roth, Scanning and Halftones, *Peachpit Press*, 2004.
- [16]. Dohnal, Moiré in a scanned image, *Proceedings of SPIE*, Vol. 4016, 1999, pp. 166-170.
- [17]. Y. Kim, G. Park, J.-H. Jung, J. Kim, B. Lee, Color moiré pattern simulation and analysis in three-dimensional integral imaging for finding the moiré-reduced tilted angle of a lens array, *Applied Optics*, Vol. 48, No. 11, 2009, pp. 2178-2187.
- [18]. V. Saveljev, J.-Y. Son, J.-H. Chun, K.-D. Kwack, K.-H. Cha, About a moiré-less condition for non-square grids, *Journal of Display Technology*, Vol. 4, No. 3, 2008, pp. 332-339.
- [19]. L. Kong, G. Jin, T. Wang, Analysis of moiré minimization in autostereoscopic parallax displays, *Optics Express*, Vol. 21, No. 22, 2013, pp. 26068-26079.
- [20]. P. S. Theocaris, S. A. Paipetis, Vibration measurements by the moiré method, *Journal of Physics E: Scientific Instruments*, Vol. 5, No. 3, 1972, pp. 217-219.
- [21]. W. Chen, W. Yan, S. Hu, Y. Yang, S. Zhou, Extended dual-grating alignment method for optical projection lithography, *Applied Optics*, Vol. 49, No. 4, 2010, pp. 708-713.
- [22]. I. Amidror, S. Chosson, R. D. Hersch, Moiré methods for the protection of documents and products, *Journal of Physics: Conference Series*, Vol. 77, 2007, 012001.
- [23]. V. Ostromoukhov, N. Rudaz, I. Amidror, P. Emmel, R. D. Hersch, Anti-counterfeiting features of artistic screening, *Proceedings of SPIE*, Vol. 2951, 2010, pp. 126-133.
- [24]. M. Ragulskis, A. Aleksa, Image hiding based on time averaging moiré, *Optics Communications*, Vol. 282, No. 14, 2009, pp. 2752-2759.
- [25]. G. Oster, Optical art, *Applied Optics*, Vol. 4, No. 11, 1965, pp. 1359-1369.
- [26]. V. Saveljev, S.-K. Kim, Three-dimensional moiré display, *Journal of the Society for Information Display*, Vol. 22, No. 9, 2015, pp. 482-486.
- [27]. J.-H. Park, S.-B. Lee, Evaluation of thermal deformation behavior in electronic package using UV moiré interferometry, in *Proceedings of the 6th International Symposium on Electronics Materials and Packaging (EMAP'05)*, Tokyo, Japan, 2005, pp. 209-214.
- [28]. Y.-Y. Chen, Z. Liu, J. Du, R.-S. Chang, Research on subcutaneous pulse shape measurement by near-infrared moiré technique, *Journal of the Optical Society of Korea*, Vol. 19, No. 2, 2015, pp. 123-129.
- [29]. Y. Hongbing, Y. Xia, Z. Jun, H. Anzhi, Measurement on infrared radiation burning temperature field with moiré deviation tomography, *Microwave and Optical Technology Letters*, Vol. 49, No. 11, 2007, pp. 2761-2763.
- [30]. G. A. Bassett, J. W. Menter, D. W. Pashley, Moiré patterns on electron micrographs, and their application to the study of dislocations in metals, *Proceedings of the Royal Society of London*, Vol. 256, No. 1246, 1958, pp. 345-368.
- [31]. D. L. Miller, K. D. Kubista, G. M. Rutter, M. Ruan, W. A. de Heer, P. N. First, J. A. Stroscio, Structural analysis of multilayer graphene via atomic moiré interferometry, *Physical Review B*, Vol. 81, 2010, 125427.

- [32]. M. Fodchuk, N. D. Raransky, Moiré images simulation of strains in x-ray interferometry, *Journal of Physics D: Applied Physics*, Vol. 36, No. 10A, 2003, p. A55-A59.
- [33]. U. Bonse, W. Grae, G. Materlik, X-ray interferometry and lattice parameter investigation, *Revue de Physique Appliquée (Paris)*, Vol. 11, No. 1, 1976, pp. 83-87.
- [34]. L. M. Fontán, J. L. Fernández, Á. F. Doval, J. L. Meniño, C. Trillo1, J. C. López-Vázquez, Wide-field, low-cost mapping of power ultrasound fields in water by time-average moiré deflectometry, in *Proceedings of the 7th International Workshop on Advanced Optical Imaging and Metrology (Fringe'13)*, 2013, pp. 567-571.
- [35]. D. N. Swingler, A. P. Anderson, Simple microwave holograms and moiré fringes using the 'spinning-dipole' field-perturbation technique, *Electronics Letters*, Vol. 5, No. 14, 1969, pp. 314-316.
- [36]. T. N'Diaye, S. Bleikamp, P. J. Feibelman, T. Michely, Two-dimensional Ir cluster lattice on a graphene moiré on Ir(111), *Physical Review Letters*, Vol. 101, No. 21, 2008, 219904.
- [37]. Y. Pan, H. Zhang, D. Shi, J. Sun, S. Du, F. Liu, H.-J. Gao, Highly ordered, millimeter-scale, continuous, single-crystalline graphene monolayer formed on Ru (0001), *Advanced Materials*, Vol. 21, No. 27, 2009, pp. 2777-2780.
- [38]. M. B. Sadan, L. Houben, A. N. Enyashin, G. Seifert, R. Tenne, Atom by atom: HRTEM insights into inorganic nanotubes and fullerene-like structures, *Proceedings of the National Academy of Sciences*, Vol. 105, 2008, pp. 15643-15648.
- [39]. N. Fukui, Y. Suwa, H. Yoshida, T. Sugai, S. Heike, M. Fujimori, Y. Terada, T. Hashizume, H. Shinohara, Moiré image patterns on double-walled carbon nanotubes observed by scanning tunneling microscopy, *Physical Review B*, Vol. 79, 2009, 125402.
- [40]. X. Gao, C. S. Lao, Y. Ding, Z. L. Wang, Metal/semiconductor core/shell nanodisks and nanotubes, *Advanced Functional Materials*, Vol. 16, No. 1, 2006, pp. 53-62.
- [41]. J. H. Warner, N. P. Young, A. I. Kirkland, G. A. D. Briggs, Resolving strain in carbon nanotubes at the atomic level, *Nature Materials*, Vol. 10, 2011, pp. 958-962.
- [42]. M. Brzezicki, Designer's controlled and randomly generated moiré patterns in architecture, in *Proceedings of the XV Generative Art Conference (GA'11)*, Rome, Italy, 2011, pp. 288-301.
- [43]. E. W. Weisstein, Homogeneous Coordinates,
<http://mathworld.wolfram.com/HomogeneousCoordinates.html>
- [44]. V. Saveljev, S.-K. Kim, Simulation and measurement of moiré patterns at finite distance, *Optics Express*, Vol. 20, No. 3, 2012, pp. 2163-2177.
- [45]. V. Saveljev, H. Lee, J. Kim, Physical model of the moiré effect in nanoparticles, *Journal of the Korean Physical Society*, 2017 (in press).
- [46]. M. C. Hutley, R. Hunt, R. F. Stevens, P. Savandert, Moiré magnifier, *Pure and Applied Optics*, Vol. 3, No. 2, 1994, pp. 133-142.
- [47]. H. Kamal, R. Volkel, J. Alda, Properties of moiré magnifiers, *Optical Engineering*, Vol. 37, 1998, pp. 3007-3014.
- [48]. V. Saveljev, S.-K. Kim, Theoretical estimation of moiré effect using spectral trajectories, *Optics Express*, Vol. 21, No. 2, 2013, pp. 1693-1712.
- [49]. R. N. Bracewell, The Fourier Transform and its Applications, *McGraw-Hill*, 1986.
- [50]. C. Huang, J. R. G. Townshend, X. Zhan, M. Hansen, R. DeFries, R. Solhberg, Developing the spectral trajectories of major land cover change processes, *Proceedings of SPIE*, Vol. 3502, 1998, pp. 155-162.
- [51]. S. Costa-Pereira, P. Maillard, Estimating the age of cerrado regeneration using Landsat TM data, *Canadian Journal of Remote Sensing*, Vol. 36, No. S2, 2010, pp. S243-S256.
- [52]. V. Saveljev, S.-K. Kim, Estimation of moiré patterns using spectral trajectories in the complex plane, *Computer Technology and Application*, Vol. 3, No. 3, 2012, pp. 353-360.

- [53]. V. Saveljev, S.-K. Kim, Probability of the moiré effect in barrier and lenticular autostereoscopic 3D displays, *Optics Express*, Vol. 23, 2015, pp. 25597-25607.
- [54]. S. Abbott, Understanding Analysis, *Springer*, 2001, Chapter 4.1.
- [55]. F. A. Jenkins, H. E. White, Fundamentals of Optics, *McGraw-Hill*, 1981.
- [56]. D. Shechtman, Quasi-periodic materials – a paradigm shift in crystallography, *Nobel Lecture*, Stockholm, Sweden, 8 December 2011.

Chapter 3

Numerical Method for Diffraction by Multilayered Dielectric Gratings Using Scattering Factors

Hideaki Wakabayashi

3.1. Introduction

The wave diffraction by a periodic structure with infinite extent does not take place and only specular reflection appears at a low grazing limit of incidence. This means that an incident plane wave is completely cancelled by the specularly reflected wave. Thus, the total wave field vanishes, and becomes a dark shadow [1]. Nakayama's shadow theory first discussed the diffraction by a perfectly conducting grating, and gave a physical explanation of such a behavior. In the theory, the modified diffracted waves are the solutions for the "primary excitation" defined as the sum of the incident and specularly reflected waves, and are expressed by use of scattering factor being the basic quantity of diffraction amplitude [2, 3].

Many analytical methods have been established for complicated dielectric gratings [4-8]. We have presented the matrix eigenvalues method to study dielectric gratings, which treats the scattering problem as the eigenvalue problem for a coefficient matrix of the first-order differential equations*. Changing the coefficient matrix, our method has been applied to analyses of anisotropic, chiral and various gratings [6-8]. However, a shadow

Hideaki Wakabayashi
Faculty of Computer Science and Systems Engineering, Okayama Prefectural University,
Soja, Okayama, Japan

* Our matrix eigenvalues method is different from the rigorous coupled wave analysis (RCWA) method that M. G. Moharam and T. K. Gaylord presented [9], in that we start with Maxwell's equations, and derive the first-order differential equations directly. The RCWA method is a very popular grating method. The method starts with the second-order differential wave equations, and transforms into the first-order state variable form in order to reduce into the eigenvalue problem. For the case of conical diffraction, rather than attempting to construct and solve the wave equations, it is more straightforward to solve Maxwell's equations directly [8, 10].

phenomenon has been overlooked and has been seldom discussed in their and our methods. By extending the shadow theory to the transmitted region, we applied the theory to our matrix eigenvalues method. We then resolved a singular behavior of low grazing scattering, which corresponds to the fact that the eigenvalues degenerate by zero and the coefficient matrix cannot be diagonalized. It was also demonstrated that the primary excitation enables us to avoid the problem of eigenvalues degeneracy and that of matrix diagonalization [11]. In the numerical analysis of multilayered dielectric gratings, the eigenvalues related to the waves have the possibility of approaching to zero and degenerating in the middle regions. We thought the eigenvalues degeneracy corresponds to a phenomenon that the incoming wave grazing along the boundary is cancelled by the reflected wave with the same amplitude and reverse phase. Incorporating the concepts of the primary excitation and manipulation by the shadow theory, we newly composed transformation and propagation matrices, and obtained a new description of electromagnetic fields by the product of these matrices. Due to overflows on numerical computations, we applied the manipulation to only modes related by degenerating eigenvalues for a uniform region and eigenvalues close to zero for a periodic region in the formulation [12, 13]. However, the manipulation should be applied to all modes in all regions, frankly.

These works [2, 3, 11-13] on the shadow theory were all restricted to two dimensional scattering problem, and only TE and TM cases were discussed. Since a shadow must be a common phenomenon regardless of the incident polarization, a discussion for cases of conical mounting and circular polarization are needed to be done.

This chapter targets formulations of applying the shadow theory to all regions of a multilayered dielectric grating, placed in conical mounting. We resolves the problem of eigenvalues degeneracy and that of matrix diagonalization for any complex angle of oblique incidence. Since our formulations affiliate for TE and TM components, the reflection of TE and TM incident components occurs with the reflection coefficient -1 at a low grazing limit. Bearing this in mind, we propose an oblique primary excitation represented by the sum of each component and its specularly reflected wave. Then, we give a new transformation matrix for conical mounting of the shadow theory. Setting the phase reference positions of the waves so as to avoid overflows on numerical computations, we newly form an improved propagation matrix over the former matrix [12]. In terms of these matrices, we can apply the shadow theory to all modes on computations, and obtain a new description of three dimensional fields even for the cases that eigenvalues degenerate in any region. Additionally, defining scattering factors for TE and TM components, we newly obtain a representation of diffraction efficiencies for linearly and circularly polarized propagating and evanescent incidence.

Section 3.6 will be devoted to the study of scattering of linearly and circularly polarized incidence by a dielectric grating with asymmetric triangular profiles, placed in conical mounting [14]. Numerical results demonstrate that, the total field vanishes and becomes a shadow regardless of the incident polarization at a low grazing limit. We find that, in a circularly polarized incidence case, only specular reflection of the same circular polarization as the incident wave occurs at a low grazing limit. Section 3.7 will study for

the three dimensional scattering by a multilayered dielectric grating [15]. Numerical results show that, the eigenvalues obtained numerically degenerate in a periodic region, and using the shadow theory, calculations can be carried out without the problem of eigenvalues degeneracy. Then, we specify that the eigenvalues degeneracy corresponds to a phenomenon that the incoming wave and the reflected wave with the same amplitude and reverse phase graze along the boundary, and are cancelled out in the middle region. We newly find that, when the eigenvalues degenerate by zero, all field components in Cartesian coordinate decay proportionally to the position in the middle region.

In this chapter, time dependence $e^{j\omega t}$ is assumed and suppressed. The space variables $\mathbf{r} = (x, y, z)$ are normalized by the wave number k_0 such that $k_0 \mathbf{r} \rightarrow \mathbf{r}$ in the coordinate system.

3.2. Matrix Eigenvalues Method

In this section, we describe the matrix eigenvalues method by use of non-degenerate eigenvalues for the scattering problem of a multilayered lossless dielectric (magnetic) grating, placed in conical mounting.

From a viewpoint of formulations, we consider a generalized structure which consists of incident, multilayered and transmitted regions, as shown in Fig. 3.1. The structure is placed in conical mounting and is illuminated by a plane wave of wavelength λ with a polar angle θ measured from the $-x$ direction, azimuthal angle ϕ between the y axis and the incident plane and a polarization angle γ . The relative permittivity and permeability in the regions 0 and N are denoted by (ϵ_a, μ_a) and (ϵ_s, μ_s) , respectively. In the multilayered regions, for a periodic case of n layers, the permittivity and permeability profiles are described by a function of position z having periodicity Λ , and for the residual uniform case, the permittivity and permeability are constants.

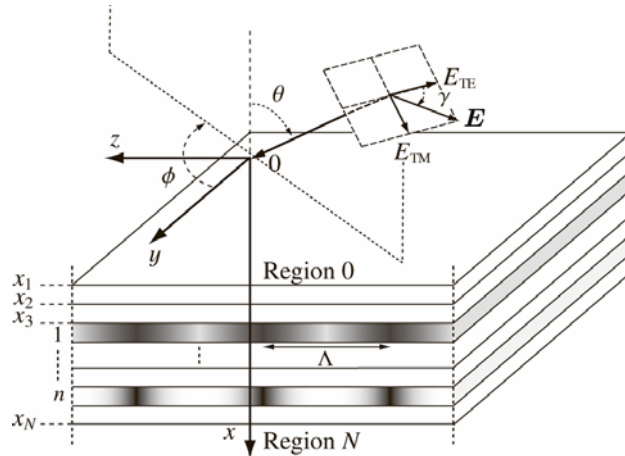


Fig. 3.1. A multilayered periodic structure in conical mounting.

Using the normalized space variables, we start with Maxwell's equations in dimensionless form as

$$\overline{\text{curl}}\sqrt{Y_0}\mathbf{E} = -j\mu(z)\sqrt{Z_0}\mathbf{H}, \quad \overline{\text{curl}}\sqrt{Z_0}\mathbf{H} = j\varepsilon(z)\sqrt{Y_0}\mathbf{E}, \quad (3.1)$$

where $\overline{\text{curl}} = \text{rot} / k_0$ and $Z_0 = 1 / Y_0$. Z_0 is the wave impedance in free space. Since the structure is periodic, the electromagnetic fields Ψ_ℓ ($\ell = x, y, z$) are expressed by the infinite sum of the spatial harmonics. Making truncation of $(2M+1)$ expansion coefficients $\psi_{\ell m}$, the fields can be approximated by

$$\Psi_\ell(x, y, z) = \sum_{m=-M}^M \psi_{\ell m}(x) \exp\{-j(q_0 y + s_m z)\}, \quad (3.2)$$

$$\begin{aligned} s_m &= s_0 + mn_K, & n_K &= \lambda / \Lambda, \\ s_0 &= \sqrt{\varepsilon_a \mu_a} \sin \theta \sin \phi, & q_0 &= \sqrt{\varepsilon_a \mu_a} \sin \theta \cos \phi. \end{aligned} \quad (3.3)$$

The relative permittivity $\varepsilon(z)$, the relative permeability $\mu(z)$ and their inverses can be expanded in Fourier series of the truncated order $N_f (= 2M)$ as

$$\zeta(z) = \sum_{m=-N_f}^{N_f} \tilde{\zeta}_m \exp\left(jm \frac{\lambda}{\Lambda} z\right), \quad \tilde{\zeta}_m = \frac{1}{\Lambda} \int \zeta(z) \exp\left\{-jm \left(\frac{\lambda}{\Lambda}\right) z\right\} dz, \quad (3.4)$$

where $\zeta = \varepsilon, \mu, 1/\varepsilon, 1/\mu$. The Fourier expansions of $1/\varepsilon(z)$ and $1/\mu(z)$ are introduced by the concept of the inverse rule [16, 17]. Making column vectors composed of the expansion coefficients $\psi_{\ell m} (= e_{\ell m}, h_{\ell m})$ in Eq. (3.2) as

$$\psi_\ell(x) = [\psi_{\ell-M}(x) \cdots \psi_{\ell 0}(x) \cdots \psi_{\ell M}(x)]^T, \quad (3.5)$$

and arranging Maxwell's equations through the inverse rule, we obtain the following coupled-wave equations concerning the x , y and z field components directly,

$$\frac{d}{dx} \begin{bmatrix} \mathbf{e}_y(x) \\ \mathbf{e}_z(x) \\ \mathbf{h}_y(x) \\ \mathbf{h}_z(x) \end{bmatrix} = j[\mathbf{C}] \begin{bmatrix} \mathbf{e}_y(x) \\ \mathbf{e}_z(x) \\ \mathbf{h}_y(x) \\ \mathbf{h}_z(x) \end{bmatrix}, \quad \begin{bmatrix} \mathbf{e}_x(x) \\ \mathbf{h}_x(x) \end{bmatrix} = [\mathbf{N}] \begin{bmatrix} \mathbf{e}_y(x) \\ \mathbf{e}_z(x) \\ \mathbf{h}_y(x) \\ \mathbf{h}_z(x) \end{bmatrix}, \quad (3.6)$$

$$[\mathbf{C}] = \begin{bmatrix} [0] & [0] - [q][\varepsilon]^{-1}[s] & [q][\varepsilon]^{-1}[q] - [1/\mu]^{-1} \\ [0] & [0] & [\mu] - [s][\varepsilon]^{-1}[s] & [1/\varepsilon]^{-1} - [q][\mu]^{-1}[q] \\ [q][\mu]^{-1}[s] & [1/\varepsilon]^{-1} - [q][\mu]^{-1}[q] & [0] & [s][\varepsilon]^{-1}[q] \\ [s][\mu]^{-1}[s] - [\varepsilon] & -[s][\mu]^{-1}[q] & [0] & [0] \end{bmatrix}, \quad (3.7)$$

$$[\mathbf{N}] = \begin{bmatrix} [0] & [0] & [\varepsilon]^{-1}[s] & -[\varepsilon]^{-1}[q] \\ -[\mu]^{-1}[s] & [\mu]^{-1}[q] & [0] & [0] \end{bmatrix}, \quad (3.8)$$

where

$$[s] = [\delta_{mn} s_m], \quad [q] = q_0 [\delta_{mn}],$$

$$[\varepsilon] = [\tilde{\varepsilon}_{n-m}], \quad [\mu] = [\tilde{\mu}_{n-m}], \quad [1/\varepsilon] = \left[\widetilde{(1/\varepsilon)}_{n-m} \right], \quad [1/\mu] = \left[\widetilde{(1/\mu)}_{n-m} \right], \quad (3.9)$$

and δ_{mn} stands for the Kronecker's delta. The solutions of the first order differential equation in Eq. (3.6) are reduced to the matrix eigenvalue problem of the coefficient matrix $[\mathbf{C}]$. Using $(2M+1)$ -dimensional mode amplitude vectors ${}^1\mathbf{g}^{\pm}$, ${}^2\mathbf{g}^{\pm}$ and $4(2M+1) \times 4(2M+1)$ -dimensional transformation matrix $[\mathbf{T}]$, we can express the expansion vectors of electromagnetic fields as

$$\begin{bmatrix} \mathbf{e}_y(x) \\ \mathbf{e}_z(x) \\ \mathbf{h}_y(x) \\ \mathbf{h}_z(x) \end{bmatrix} = [\mathbf{T}] \begin{bmatrix} {}^1\mathbf{g}^+(x) \\ {}^2\mathbf{g}^+(x) \\ {}^1\mathbf{g}^-(x) \\ {}^2\mathbf{g}^-(x) \end{bmatrix}, \quad (3.10)$$

where the symbol " \pm " denotes propagation directions along the x direction. The distinctions of ${}^1\mathbf{g}$ and ${}^2\mathbf{g}$ are not identified. When $[\mathbf{C}]$ has $4(2M+1)$ non-degenerate eigenvalues and 4 independent eigenvectors, using the eigenvectors matrix $[\mathbf{T}]$, we can diagonalize $[\mathbf{C}]$ by similar transformation as follows:

$$[\mathbf{T}]^{-1}[\mathbf{C}][\mathbf{T}] = \begin{bmatrix} [\delta_{mn} {}^1\kappa_m^+] & [0] & [0] & [0] \\ [0] & [\delta_{mn} {}^2\kappa_m^+] & [0] & [0] \\ [0] & [0] & [\delta_{mn} {}^1\kappa_m^-] & [0] \\ [0] & [0] & [0] & [\delta_{mn} {}^2\kappa_m^-] \end{bmatrix}, \quad (3.11)$$

which holds for $m, n = -M, \dots, 0, \dots, M$. Thus, we can solve the differential Eq. (3.6). The solution is

$$\begin{bmatrix} \mathbf{e}_y(x) \\ \mathbf{e}_z(x) \\ \mathbf{h}_y(x) \\ \mathbf{h}_z(x) \end{bmatrix} = [\mathbf{T}][\mathbf{P}(x)] \begin{bmatrix} {}^1\mathbf{g}^+(\dot{x}) \\ {}^2\mathbf{g}^+(\dot{x}) \\ {}^1\mathbf{g}^-(\ddot{x}) \\ {}^2\mathbf{g}^-(\ddot{x}) \end{bmatrix}, \quad (3.12)$$

where $[\mathbf{P}(x)]$ is a propagation matrix, and is expressed by

$$[\mathbf{P}(x)] = \begin{bmatrix} \begin{bmatrix} \delta_{mn} e^{j^1\kappa_m^+(x-\dot{x})} \end{bmatrix} & [0] & [0] & [0] \\ [0] & \begin{bmatrix} \delta_{mn} e^{j^2\kappa_m^+(x-\dot{x})} \end{bmatrix} & [0] & [0] \\ [0] & [0] & \begin{bmatrix} \delta_{mn} e^{j^1\kappa_m^-(x-\ddot{x})} \end{bmatrix} & [0] \\ [0] & [0] & [0] & \begin{bmatrix} \delta_{mn} e^{j^1\kappa_m^-(x-\ddot{x})} \end{bmatrix} \end{bmatrix}. \quad (3.13)$$

Here, the phase reference positions stand for \dot{x} and \ddot{x} . In our previous works [11-14], we set the same positions. We changes the positions corresponding to the propagation directions, and avoids overflows on computations.

In uniform regions, we can express electromagnetic fields by a superposition of TE and TM waves. Therefore, we can denote the eigenvalues ${}^{1(2)}\kappa_m^\pm$ by ${}^{E(M)}\kappa_m^\pm$ and the mode amplitudes vectors ${}^{1(2)}\mathbf{g}^\pm$ by ${}^{E(M)}\mathbf{g}^\pm$. Here, the superscripts E and M refer to TE and TM waves, respectively. In periodic regions, since the eigenvalues ${}^{1(2)}\kappa_m^\pm$ are numerically obtained, we cannot distinguish between ${}^1\mathbf{g}^\pm$ and ${}^2\mathbf{g}^\pm$ so far.

3.3. Application of Shadow Theory to All Regions

In this section, we pay attention to the middle regions of a multilayered dielectric grating. Applying the concept of the shadow theory to the middle regions in addition to the cases of the incident and transmitted regions which we reported in Ref. [12], we present a new description of the three dimensional electromagnetic fields in all regions without the problem of eigenvalues degeneracy.

3.3.1. In the Case of Uniform Regions

In any uniform region, since the permittivity and permeability are constants, all submatrices of the coefficient matrix $[\mathbf{C}]$ can be expressed by diagonal matrices. The

spatial harmonic waves have no coupling, and are plane waves. Therefore, we consider only $[C_m]$ corresponding to the m th mode. The m th order eigenvalues ${}^E \kappa_m^\pm = {}^M \kappa_m^\pm = \mp \xi_m$ can be given analytically as

$$\xi_m = \begin{cases} \sqrt{\varepsilon\mu - q_0^2 - s_m^2} & (\varepsilon\mu \geq q_0^2 + s_m^2) \\ j\sqrt{q_0^2 + s_m^2 - \varepsilon\mu} & (\varepsilon\mu < q_0^2 + s_m^2) \end{cases}. \quad (3.14)$$

The eigenvectors matrix $[T_m]$ can be given by

$$[T_m] = \begin{bmatrix} {}^E \mathbf{t}_m^+ & {}^M \mathbf{t}_m^+ & {}^E \mathbf{t}_m^- & {}^M \mathbf{t}_m^- \end{bmatrix} \begin{bmatrix} -\bar{s}_m \sqrt{\mu} & -\xi_m \bar{q}_0 / \sqrt{\varepsilon} & -\bar{s}_m \sqrt{\mu} & \xi_m \bar{q}_0 / \sqrt{\varepsilon} \bar{q}_0 \sqrt{\mu} \\ \bar{q}_0 \sqrt{\mu} & \xi_m \bar{s}_m / \sqrt{\varepsilon} & \bar{q}_0 \sqrt{\mu} & -\xi_m \bar{s}_m / \sqrt{\varepsilon} \\ -\xi_m \bar{q}_0 / \sqrt{\mu} & \bar{s}_m \sqrt{\varepsilon} & \xi_m \bar{q}_0 / \sqrt{\mu} & \bar{s}_m \sqrt{\varepsilon} \\ -\xi_m \bar{s}_m / \sqrt{\mu} & -\bar{q}_0 \sqrt{\varepsilon} & \xi_m \bar{s}_m / \sqrt{\mu} & -\bar{q}_0 \sqrt{\varepsilon} \end{bmatrix}, \quad (3.15)$$

$$\bar{s}_m = \begin{cases} \frac{s_m}{\sqrt{q_0^2 + s_m^2}} \\ \sin \phi \end{cases}, \quad \bar{q}_0 = \begin{cases} \frac{q_0}{\sqrt{q_0^2 + s_m^2}} (m \neq 0) \\ \cos \phi (m = 0) \end{cases}. \quad (3.16)$$

We employ the eigenvectors ${}^E \mathbf{t}^\pm$ and ${}^M \mathbf{t}^\pm$ normalized such that the magnitudes of the electric fields $\sqrt{|e_x|^2 + |e_y|^2 + |e_z|^2}$ in TE and TM waves are the same. When the eigenvalues do not degenerate, by $[T_m]$, $[P_m(x)]$ and ${}^{E(M)} g_m^\pm$, the m th mode electromagnetic fields can be expressed as

$$\begin{bmatrix} e_{ym}(x) \\ e_{zm}(x) \\ h_{ym}(x) \\ h_{zm}(x) \end{bmatrix} = [T_m][P_m(x)] \begin{bmatrix} {}^E g_m^+(\dot{x}) \\ {}^M g_m^+(\dot{x}) \\ {}^E g_m^-(\ddot{x}) \\ {}^M g_m^-(\ddot{x}) \end{bmatrix}, \quad (3.17)$$

where

$$[P_m(x)] = \begin{bmatrix} e^{-j\xi_m(x-\dot{x})} & 0 & 0 & 0 \\ 0 & e^{-j\xi_m(x-\dot{x})} & 0 & 0 \\ 0 & 0 & e^{j\xi_m(x-\ddot{x})} & 0 \\ 0 & 0 & 0 & e^{j\xi_m(x-\ddot{x})} \end{bmatrix}. \quad (3.18)$$

When the eigenvalues ξ_m degenerate by zero, since Eq. (3.15) has linearly dependent eigenvectors, we obtain an inexact description in Eq. (3.17) based on the conventional approach. At this situation, we cannot diagonalize the coefficient matrix. Getting eigenvectors in extended eigenspace by using Hamilton-Cayley theorem, we can construct the transformation matrix of the coefficient matrix into Jordan normal form [12]. However, rather than transforming into Jordan normal form to resolve the problem of the matrix diagonalization, it is convenient to apply the shadow theory to our formulations, as described below. Incorporating the concepts of the primary excitation, and manipulating by the shadow theory, Eq. (3.17) yields

$$\begin{aligned}
 & \begin{bmatrix} e_{ym}(x) \\ e_{zm}(x) \\ h_{ym}(x) \\ h_{zm}(x) \end{bmatrix} = [T_m] [P_m(x)] \begin{bmatrix} {}^E g_m^+(\dot{x}) \\ {}^M g_m^+(\dot{x}) \\ {}^E g_m^-(\ddot{x}) \\ {}^M g_m^-(\ddot{x}) \end{bmatrix} \\
 & = {}^E \mathbf{t}_m^+ e^{-j\xi_m(x-\bar{x})} {}^E g_m^+(\dot{x}) - {}^E \mathbf{t}_m^- e^{j\xi_m(x-\bar{x})} {}^E g_m^+(\dot{x}) + {}^E \mathbf{t}_m^- e^{j\xi_m(x-\bar{x})} \left\{ {}^E g_m^+(\dot{x}) + {}^E g_m^-(\ddot{x}) \right\} \\
 & \quad + {}^M \mathbf{t}_m^+ e^{-j\xi_m(x-\bar{x})} {}^M g_m^+(\dot{x}) - {}^M \mathbf{t}_m^- e^{j\xi_m(x-\bar{x})} {}^M g_m^+(\dot{x}) + {}^M \mathbf{t}_m^- e^{j\xi_m(x-\bar{x})} \left\{ {}^M g_m^+(\dot{x}) + {}^M g_m^-(\ddot{x}) \right\} \\
 & = \begin{bmatrix} -j\bar{s}_m\sqrt{\mu} \frac{e^{-j\xi_m(x-\bar{x})} - e^{j\xi_m(x-\bar{x})}}{2j\xi_m} & -\frac{\bar{q}_0}{\sqrt{\varepsilon}} \frac{e^{-j\xi_m(x-\bar{x})} + e^{j\xi_m(x-\bar{x})}}{2} & -\bar{s}_m\sqrt{\mu} e^{j\xi_m(x-\bar{x})} & \xi_m \frac{\bar{q}_0}{\sqrt{\varepsilon}} e^{j\xi_m(x-\bar{x})} \\ j\bar{q}_0\sqrt{\mu} \frac{e^{-j\xi_m(x-\bar{x})} - e^{j\xi_m(x-\bar{x})}}{2j\xi_m} & -\frac{\bar{s}_m}{\sqrt{\varepsilon}} \frac{e^{-j\xi_m(x-\bar{x})} + e^{j\xi_m(x-\bar{x})}}{2} & \bar{q}_0\sqrt{\mu} e^{j\xi_m(x-\bar{x})} & \xi_m \frac{\bar{s}_m}{\sqrt{\varepsilon}} e^{j\xi_m(x-\bar{x})} \\ -\frac{\bar{q}_0}{\sqrt{\mu}} \frac{e^{-j\xi_m(x-\bar{x})} + e^{j\xi_m(x-\bar{x})}}{2} & j\bar{s}_m\sqrt{\varepsilon} \frac{e^{-j\xi_m(x-\bar{x})} - e^{j\xi_m(x-\bar{x})}}{2j\xi_m} & \xi_m \frac{\bar{q}_0}{\sqrt{\mu}} e^{j\xi_m(x-\bar{x})} & \bar{s}_m\sqrt{\varepsilon} e^{j\xi_m(x-\bar{x})} \\ -\frac{\bar{s}_m}{\sqrt{\mu}} \frac{e^{-j\xi_m(x-\bar{x})} + e^{j\xi_m(x-\bar{x})}}{2} & -j\bar{q}_0\sqrt{\varepsilon} \frac{e^{-j\xi_m(x-\bar{x})} - e^{j\xi_m(x-\bar{x})}}{2j\xi_m} & \xi_m \frac{\bar{s}_m}{\sqrt{\mu}} e^{j\xi_m(x-\bar{x})} & -\bar{q}_0\sqrt{\varepsilon} e^{j\xi_m(x-\bar{x})} \end{bmatrix} \\
 & \quad \times \begin{bmatrix} 2\xi_m {}^E g_m^+(\dot{x}) \\ 2\xi_m {}^M g_m^+(\dot{x}) \\ {}^E g_m^+(\dot{x}) + {}^E g_m^-(\ddot{x}) \\ {}^M g_m^+(\dot{x}) + {}^M g_m^-(\ddot{x}) \end{bmatrix} = [\bar{T}_m] [\bar{P}_m(x)] \begin{bmatrix} 2\xi_m {}^E g_m^+(\dot{x}) \\ 2\xi_m {}^M g_m^+(\dot{x}) \\ {}^E g_m^+(\dot{x}) + {}^E g_m^-(\ddot{x}) \\ {}^M g_m^+(\dot{x}) + {}^M g_m^-(\ddot{x}) \end{bmatrix} \\
 & = [\bar{T}_m] [\bar{P}_m(x)] \begin{bmatrix} {}^E M_m^\oplus \\ {}^M M_m^\oplus \\ {}^E M_m^- \\ {}^M M_m^- \end{bmatrix}, \tag{3.19}
 \end{aligned}$$

where

$$[\bar{T}_m] = \begin{bmatrix} 0 & -\bar{q}_0 / \sqrt{\varepsilon} & -\bar{s}_m \sqrt{\mu} & \xi_m \bar{q}_0 / \sqrt{\varepsilon} \\ 0 & -\bar{s}_m / \sqrt{\varepsilon} & \bar{q}_0 / \sqrt{\mu} & \xi_m \bar{s}_m / \sqrt{\varepsilon} \\ -\bar{q}_0 \sqrt{\mu} & 0 & \xi_m \bar{q}_0 / \sqrt{\mu} & \bar{s}_m / \sqrt{\varepsilon} \\ -\bar{s}_m / \sqrt{\mu} & 0 & \xi_m \bar{s}_m / \sqrt{\mu} & -\bar{q}_0 \sqrt{\varepsilon} \end{bmatrix}, \quad (3.20)$$

$$[\bar{P}_m(x)] = \begin{bmatrix} e^{-j\xi_m(x-\dot{x})} & 0 & 0 & 0 \\ 0 & e^{-j\xi_m(x-\ddot{x})} & 0 & 0 \\ \frac{\sin \xi_m \left(x - \frac{\dot{x} + \ddot{x}}{2} \right)}{j\xi_m} e^{j\xi_m \frac{\dot{x}-\ddot{x}}{2}} & 0 & e^{j\xi_m(x-\ddot{x})} & 0 \\ 0 & \frac{\sin \xi_m \left(x - \frac{\dot{x} + \ddot{x}}{2} \right)}{j\xi_m} e^{j\xi_m \frac{\dot{x}-\ddot{x}}{2}} & 0 & e^{j\xi_m(x-\ddot{x})} \end{bmatrix}, \quad (3.21)$$

$${}^{\text{E(M)}} M_m^{\oplus} = 2\xi_m {}^{\text{E(M)}} g_m^+, \quad {}^{\text{E(M)}} M_m^- = {}^{\text{E(M)}} g_m^+ + {}^{\text{E(M)}} g_m^-. \quad (3.22)$$

Here, ${}^{\text{E(M)}} M_m^{\oplus}$ and ${}^{\text{E(M)}} M_m^-$ are the m th order modified diffraction amplitudes. Since the symbol "+" denoting propagation direction has no physical meaning, we use \oplus . At

$\xi_m = 0$, $\lim_{\xi_m \rightarrow 0} \frac{\sin \xi_m \left(x - \frac{\dot{x} + \ddot{x}}{2} \right)}{j\xi_m} e^{j\xi_m \frac{\dot{x}-\ddot{x}}{2}} = -j \left(x - \frac{\dot{x} + \ddot{x}}{2} \right)$. We newly introduce $[\bar{T}_m]$

and $[\bar{P}_m(x)]$ which denote transformation and propagation matrices of the shadow theory, respectively. The first and second columns of $[\bar{T}_m]$ agree with eigenvectors obtained in extended eigenspace. The third and fourth columns are usual eigenvectors. Even when the eigenvalues degenerate, and in a normal case that the eigenvalues do not degenerate, we can express a new description (3.19) of electromagnetic fields. In our previous work [12], we applied the manipulation by the shadow theory to only modes related by degenerating eigenvalues in uniform regions. In this section, using an improved matrix (3.21) with a modification of the phase reference positions, we can apply the manipulation to all modes related by all eigenvalues on numerical computations.

Making $4(2M+1) \times 4(2M+1)$ -dimensional matrices $[\bar{\mathbf{T}}]$ and $[\bar{\mathbf{P}}(x)]$ including $(2M+1)$ spatial harmonics as

$$[\bar{\mathbf{T}}] = \begin{bmatrix} [0] & [\bar{T}_{12,m} \delta_{mn}] & [T_{13,m} \delta_{mn}] & [T_{14,m} \delta_{mn}] \\ [0] & [\bar{T}_{22,m} \delta_{mn}] & [T_{23,m} \delta_{mn}] & [T_{24,m} \delta_{mn}] \\ [\bar{T}_{31,m} \delta_{mn}] & [0] & [T_{33,m} \delta_{mn}] & [T_{34,m} \delta_{mn}] \\ [\bar{T}_{41,m} \delta_{mn}] & [0] & [T_{43,m} \delta_{mn}] & [T_{44,m} \delta_{mn}] \end{bmatrix}, \quad (3.23)$$

$$[\bar{\mathbf{P}}(x)] = \begin{bmatrix} [P_{11,m} \delta_{mn}] & [0] & [0] & [0] \\ [0] & [P_{22,m} \delta_{mn}] & [0] & [0] \\ [\bar{P}_{31,m} \delta_{mn}] & [0] & [P_{33,m} \delta_{mn}] & [0] \\ [0] & [\bar{P}_{42,m} \delta_{mn}] & [0] & [P_{44,m} \delta_{mn}] \end{bmatrix}, \quad (3.24)$$

where the subscripts (ij, m) denote the i th row and j th column element of the m th order matrices $[T_m]$ or $[\bar{T}_m]$ and $[P_m]$ or $[\bar{P}_m]$, the electromagnetic fields (3.19) can be rewritten by

$$\begin{bmatrix} \mathbf{e}(x) \\ \mathbf{h}(x) \end{bmatrix} = [\bar{\mathbf{T}}][\bar{\mathbf{P}}(x)] \begin{bmatrix} \mathbf{M}^{\oplus} \\ \mathbf{M}^{-} \end{bmatrix}, \quad (3.25)$$

using $2(2M + 1)$ -dimensional vectors \mathbf{e} , \mathbf{h} , \mathbf{M}^{\oplus} and \mathbf{M}^{-} as

$$\mathbf{e} = [\mathbf{e}_y \quad \mathbf{e}_z]^T, \quad \mathbf{h} = [\mathbf{h}_y \quad \mathbf{h}_z]^T, \quad \mathbf{M}^{\oplus} = [\mathbf{M}^{\oplus} \quad \mathbf{M}^{\oplus}]^T, \quad \mathbf{M}^{-} = [\mathbf{M}^{-} \quad \mathbf{M}^{-}]^T. \quad (3.26)$$

3.3.2. In the Case of Periodic Regions

We note that, in periodic regions, the matrix eigenvalues of the coefficient matrix $[\mathbf{C}]$ are numerically calculated by computers, and include eigenvalues having the possibility of degenerating by zero. In this section, we numerically introduce a new transformation matrix $[\bar{\mathbf{T}}]$ and an improved matrix $[\bar{\mathbf{P}}(x)]$ for periodic regions, and present a new description of electromagnetic fields including the cases that eigenvalues degenerate.

We rewrite the coefficient matrix $[\mathbf{C}]$ in the differential Eq. (3.6) and the eigenvector matrix $[\mathbf{T}]$ as

$$[\mathbf{C}] = \begin{bmatrix} [0] & [\mathbf{C}_1] \\ [\mathbf{C}_2] & [0] \end{bmatrix}, \quad [\mathbf{T}] = \begin{bmatrix} [\mathbf{T}_{11}] & [\mathbf{T}_{12}] \\ [\mathbf{T}_{21}] & [\mathbf{T}_{22}] \end{bmatrix}, \quad (3.27)$$

which hold for $k, l = 1, \dots, 2(2M + 1)$. Let us consider the matrix product $[[C_2][C_1]]$ of the half-dimensional submatrices. Note that, the eigenvalue is given by $\kappa_k^2 (= \kappa_{k'}^2)$, and the diagonalization matrix is given by $[\mathbf{T}_{21}] (= [\mathbf{T}_{22}])$ [12]. Therefore, the eigenvalues of $[\mathbf{C}]$ is given by $[\{\kappa_k\}\{\kappa_{k'}\}] = [\{-\xi_k\}\{\xi_k\}]$. Putting $[\mathbf{T}_{21}] = [\Phi]$, $[\mathbf{T}]$ can be written as

$$[\mathbf{T}] = \begin{bmatrix} -[\mathbf{C}_2]^{-1}[\Phi][-\xi_k \delta_{kl}] & [\mathbf{C}_2]^{-1}[\Phi][\xi_k \delta_{kl}] \\ [\Phi] & [\Phi] \end{bmatrix}, \quad (3.28)$$

which means that the eigenvalues calculation can be performed under the half-dimension. Transforming Eq. (3.28) so as not to calculate the inverse matrix, we can obtain

$$[\mathbf{T}] = \begin{bmatrix} [\mathbf{C}_1][\Phi] & [\mathbf{C}_1][\Phi] \\ -[\Phi][\xi_k \delta_{kl}] & [\Phi][\xi_k \delta_{kl}] \end{bmatrix}. \quad (3.29)$$

Denoting $2(2M + 1)$ -dimensional vectors by $\mathbf{e} = [\mathbf{e}_y \ \mathbf{e}_z]^T$, $\mathbf{h} = [\mathbf{h}_y \ \mathbf{h}_z]^T$ and $\mathbf{g}^+ = [\mathbf{g}^+ \ \mathbf{g}^+]^T$, the electromagnetic fields (3.12) by the conventional approach can be expressed by

$$\begin{bmatrix} \mathbf{e}(x) \\ \mathbf{h}(x) \end{bmatrix} = [\mathbf{T}][\mathbf{P}(x)] \begin{bmatrix} \mathbf{g}^+(\dot{x}) \\ \mathbf{g}^-(\ddot{x}) \end{bmatrix}, \quad (3.30)$$

where $[\mathbf{P}(x)]$ is a propagation matrix composed of the $4(2M + 1)$ eigenvalues $\{-\xi_k \ \xi_k\}$, and is given by

$$[\mathbf{P}(x)] = \begin{bmatrix} \begin{bmatrix} \delta_{kl} e^{-j\xi_k(x-\dot{x})} \\ 0 \end{bmatrix} & [0] \\ [0] & \begin{bmatrix} \delta_{kl} e^{j\xi_k(x-\ddot{x})} \end{bmatrix} \end{bmatrix}. \quad (3.31)$$

Although the dimension numbers are different, the electromagnetic fields Eq. (3.30) has the same form as the two dimensional case [7, 12].

Applying the manipulation by the shadow theory to Eq. (3.30), we obtain

$$\begin{aligned}
 \begin{bmatrix} e(x) \\ h(x) \end{bmatrix} &= [T] [P(x)] \begin{bmatrix} g^+(\dot{x}) \\ g^-(\ddot{x}) \end{bmatrix} \\
 &= \begin{bmatrix} [C_1][\Phi] \\ -[\Phi][\xi_k \delta_{kl}] \end{bmatrix} [\delta_{kl} e^{-j\xi_k(x-\dot{x})}] g^+(\dot{x}) - \begin{bmatrix} [C_1][\Phi] \\ [\Phi][\xi_k \delta_{kl}] \end{bmatrix} [\delta_{kl} e^{j\xi_k(x-\ddot{x})}] g^-(\ddot{x}) \\
 &\quad + \begin{bmatrix} [C_1][\Phi] \\ [\Phi][\xi_k \delta_{kl}] \end{bmatrix} [\delta_{kl} e^{j\xi_k(x-\ddot{x})}] \{g^+(\dot{x}) + g^-(\ddot{x})\} \\
 &= [\bar{T}] [\bar{P}(x)] \begin{bmatrix} 2[\xi_k \delta_{kl}] g^+(\dot{x}) \\ g^+(\dot{x}) + g^-(\ddot{x}) \end{bmatrix} = [\bar{T}] [\bar{P}(x)] \begin{bmatrix} M^{\oplus} \\ M^- \end{bmatrix}.
 \end{aligned} \tag{3.32}$$

Here, a transformation matrix $[\bar{T}]$ and a propagation matrix $[\bar{P}]$ are numerically given by

$$[\bar{T}] = \begin{bmatrix} [0] & [C_1][\Phi] \\ -[\Phi] & [\Phi][\xi_k \delta_{kl}] \end{bmatrix}, \tag{3.33}$$

$$[\bar{P}(x)] = \begin{bmatrix} \begin{bmatrix} \delta_{kl} e^{-j\xi_k(x-\dot{x})} & [0] \\ \delta_{kl} \frac{\sin \xi_k \left(x - \frac{\dot{x} + \ddot{x}}{2} \right)}{j\xi_k} e^{j\xi_k \frac{\dot{x} - \ddot{x}}{2}} & [\delta_{kl} e^{j\xi_k(x-\ddot{x})}] \end{bmatrix} \end{bmatrix}. \tag{3.34}$$

The manipulation by the shadow theory was applied to only modes related by eigenvalues close to zero in periodic regions [12]. The newly-obtained transformation and propagation matrices in this section enable us to apply the manipulation to all modes. Therefore, applying the shadow theory, we can carry out calculations without the problem of the matrix diagonalization.

3.4. An Excitation Source and Boundary Conditions

Putting $m = 0$ in Eq. (3.19), the 0th mode electromagnetic fields in the region 0 are given by

$$\begin{bmatrix} e_{y0}^a(x) \\ e_{z0}^a(x) \\ h_{y0}^a(x) \\ h_{z0}^a(x) \end{bmatrix} = \begin{bmatrix} j \bar{s}_0 \sqrt{\mu_a} \frac{\sin(\xi_0^a x)}{\xi_0^a} & -\frac{\bar{q}_0}{\sqrt{\epsilon_a}} \cos(\xi_0^a x) & -\bar{s}_0 \sqrt{\mu_a} e^{j\xi_0^a x} & \xi_0^a \frac{\bar{q}_0}{\sqrt{\epsilon_a}} e^{j\xi_0^a x} \\ -j \bar{q}_0 \sqrt{\mu_a} \frac{\sin(\xi_0^a x)}{\xi_0^a} & -\frac{\bar{s}_0}{\sqrt{\epsilon_a}} \cos(\xi_0^a x) & \bar{q}_0 \sqrt{\mu_a} e^{j\xi_0^a x} & \xi_0^a \frac{\bar{s}_0}{\sqrt{\epsilon_a}} e^{j\xi_0^a x} \\ -\frac{\bar{q}_0}{\sqrt{\mu_a}} \cos(\xi_0^a x) & -j \bar{s}_0 \sqrt{\epsilon_a} \frac{\sin(\xi_0^a x)}{\xi_0^a} & \xi_0^a \frac{\bar{q}_0}{\sqrt{\mu_a}} e^{j\xi_0^a x} & \bar{s}_0 \sqrt{\epsilon_a} e^{j\xi_0^a x} \\ -\frac{\bar{s}_0}{\sqrt{\mu_a}} \cos(\xi_0^a x) & j \bar{q}_0 \sqrt{\epsilon_a} \frac{\sin(\xi_0^a x)}{\xi_0^a} & \xi_0^a \frac{\bar{s}_0}{\sqrt{\mu_a}} e^{j\xi_0^a x} & -\bar{q}_0 \sqrt{\epsilon_a} e^{j\xi_0^a x} \end{bmatrix} \begin{bmatrix} E M_0^{\oplus} \\ M M_0^{\oplus} \\ E M_0^- \\ M M_0^- \end{bmatrix}, \tag{3.35}$$

where ξ_0^a is the 0th order eigenvalue in the region 0. At $x = 0$, the primary excitation as an excitation source is written as

$$\begin{bmatrix} e_{y0}^a(x) \\ e_{z0}^a(x) \\ h_{y0}^a(x) \\ h_{z0}^a(x) \end{bmatrix} = \begin{bmatrix} 0 & -\bar{q}_0 / \sqrt{\varepsilon_a} \\ 0 & -\bar{s}_0 / \sqrt{\varepsilon_a} \\ -\bar{q}_0 / \sqrt{\mu_a} & 0 \\ -\bar{s}_0 / \sqrt{\mu_a} & 0 \end{bmatrix} \begin{bmatrix} {}^E M_0^{a\oplus} \\ {}^M M_0^{a\oplus} \end{bmatrix}. \quad (3.36)$$

Taking an amplitudes vector of the modified diffraction amplitudes,

$$\mathbf{M}_0^\oplus = \begin{bmatrix} \mathbf{E} & \mathbf{M}_0^\oplus \\ \mathbf{M}_0^\oplus & \mathbf{M} \end{bmatrix}^T = \begin{bmatrix} 0 \cdots 0 & 2\xi_0^a \mathbf{E} \mathbf{g}_0^{a+} & 0 \cdots 0 & 2\xi_0^a \mathbf{M} \mathbf{g}_0^{a+} & 0 \cdots 0 \end{bmatrix}^T, \quad (3.37)$$

and denoting the matrix product of $\begin{bmatrix} \bar{\mathbf{T}} \end{bmatrix}$ and $\begin{bmatrix} \bar{\mathbf{P}}(\mathbf{x}) \end{bmatrix}$ of the regions $(n-1)$ and n by

$$[\mathbf{A}_n] = [\bar{\mathbf{T}}_{n-1}] [\bar{\mathbf{P}}_{n-1}(x_n)], \quad [\mathbf{B}_n] = [\bar{\mathbf{T}}_n] [\bar{\mathbf{P}}_n(x_{n+1})], \quad (3.38)$$

we obtain a set of equations for the boundary conditions at $x = x_1, x_2, \dots, x_N$:

$$\begin{aligned}
& \left[\mathbf{A}_1 \right] \begin{bmatrix} \mathbf{0} \\ \mathbf{M}_0^- \end{bmatrix} - \left[\mathbf{B}_1 \right] \begin{bmatrix} \mathbf{M}_1^\oplus \\ \mathbf{M}_1^- \end{bmatrix} = \begin{bmatrix} \left[\delta_{m0} \frac{\bar{q}_0}{\sqrt{\varepsilon_a}} 2 \xi_0^{a \text{ M}} g_0^{a+} \right] \\ \left[\delta_{m0} \frac{\bar{s}_0}{\sqrt{\varepsilon_a}} 2 \xi_0^{a \text{ M}} g_0^{a+} \right] \\ \left[\delta_{m0} \frac{\bar{q}_0}{\sqrt{\mu_a}} 2 \xi_0^{a \text{ E}} g_0^{a+} \right] \\ \left[\delta_{m0} \frac{\bar{s}_0}{\sqrt{\mu_a}} 2 \xi_0^{a \text{ E}} g_0^{a+} \right] \end{bmatrix}, \\
& \left[\mathbf{A}_2 \right] \begin{bmatrix} \mathbf{M}_1^\oplus \\ \mathbf{M}_1^- \end{bmatrix} - \left[\mathbf{B}_2 \right] \begin{bmatrix} \mathbf{M}_2^\oplus \\ \mathbf{M}_2^- \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \\
& \dots \dots \dots, \\
& \left[\mathbf{A}_N \right] \begin{bmatrix} \mathbf{M}_{N-1}^\oplus \\ \mathbf{M}_{N-1}^- \end{bmatrix} - \left[\mathbf{B}_N \right] \begin{bmatrix} \mathbf{M}_N^+ \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}.
\end{aligned}
\tag{3.39}$$

Since we do not need to apply the shadow theory in the region N , we denote the unknown by \mathbf{M}_N^+ . Here, we introduce scattering factors \mathbf{S}^+ and \mathbf{S}^- having the relation with the modified diffraction amplitudes \mathbf{M}^+ and \mathbf{M}^- of

$$\mathbf{M}^{\oplus} = 2 \begin{bmatrix} \xi_0^a & \delta_{kl} \end{bmatrix} \mathbf{S}^{\oplus}, \quad \mathbf{M}^{-} = 2 \begin{bmatrix} \xi_0^a & \delta_{kl} \end{bmatrix} \mathbf{S}^{-}. \quad (3.40)$$

From Eqs. (3.19) or (3.32), by use of the scattering factors, we obtain a new description of electromagnetic fields

$$\begin{bmatrix} \mathbf{e}(x) \\ \mathbf{h}(x) \end{bmatrix} = \begin{bmatrix} \bar{\mathbf{T}} \\ \bar{\mathbf{P}}(x) \end{bmatrix} \begin{bmatrix} 2 \begin{bmatrix} \xi_0^a & \delta_{kl} \end{bmatrix} \mathbf{S}^{\oplus} \\ 2 \begin{bmatrix} \xi_0^a & \delta_{kl} \end{bmatrix} \mathbf{S}^{-} \end{bmatrix}. \quad (3.41)$$

We normalize Eq. (3.39) by $2\xi_0^a$ to obtain the solutions of \mathbf{S}_0^{-} and \mathbf{S}_N^{+} which are called the scattering factors in the original shadow theory. On the other hand, $(\mathbf{S}_1^{\oplus}, \mathbf{S}_1^{-})$, $(\mathbf{S}_2^{\oplus}, \mathbf{S}_2^{-})$, \dots , $(\mathbf{S}_{N-1}^{\oplus}, \mathbf{S}_{N-1}^{-})$ are extended scattering factors to the middle regions. Determining the scattering factors, the electromagnetic fields in all regions can be obtained.

In uniform regions, the m th order diffraction amplitudes in TE and TM waves are given by

$${}^{\text{E(M)}}g_{n,m}^{+}(x_n) = \frac{2\xi_0^a {}^{\text{E(M)}}S_{n,m}^{\oplus}}{2\xi_{n,m}}, \quad (3.42)$$

$${}^{\text{E(M)}}g_{n,m}^{-}(x_n) = \left(2\xi_0^a {}^{\text{E(M)}}S_{n,m}^{-} - {}^{\text{E(M)}}g_{n,m}^{+} \right) e^{j\xi_{n,m}(x_n - x_{n+1})}, \quad (3.43)$$

where $\xi_{n,m}$ is the m th order eigenvalues in region n .

3.5. Diffraction Efficiencies and Optical Theorem

The average power flow in the x direction is expressed by

$$\begin{aligned} & \frac{1}{\Lambda} \int_{-\Lambda/2}^{\Lambda/2} \left\{ \mathbf{E}_a(x_1, y, z) \times \mathbf{H}_a^*(x_1, y, z) \right\}_x dz \\ & - \frac{1}{\Lambda} \int_{-\Lambda/2}^{\Lambda/2} \left\{ \mathbf{E}_s(x_N, y, z) \times \mathbf{H}_s^*(x_N, y, z) \right\}_x dz = 0, \end{aligned} \quad (3.44)$$

where $(\mathbf{E}_a, \mathbf{H}_a)$ and $(\mathbf{E}_s, \mathbf{H}_s)$ are the electromagnetic fields in the incident and transmitted regions, respectively. Substituting the description of the electromagnetic fields by use of the diffraction amplitudes into Eq. (3.44) and taking out the real part, we obtain

$$\begin{aligned}
 & \text{Re}[\xi_0^a] |^E g_0^{a+}|^2 + 2 \text{Im}[\xi_0^a] \text{Im} \left[\left({}^E g_0^{a+} \right)^* {}^E g_0^{a-} \right] \\
 & + \text{Re}[\xi_0^a] |^M g_0^{a+}|^2 + 2 \text{Im}[\xi_0^a] \text{Im} \left[\left({}^M g_0^{a+} \right)^* {}^M g_0^{a-} \right] - \sum_m \text{Re}[\xi_m^a] |^E g_m^{a-}|^2 \quad (3.45) \\
 & - \sum_m \text{Re}[\xi_m^a] |^M g_m^{a-}|^2 = \sum_m \text{Re}[\xi_m^s] |^E g_m^{s+}|^2 + \sum_m \text{Re}[\xi_m^s] |^M g_m^{s+}|^2.
 \end{aligned}$$

Therefore, the m th order reflected diffraction efficiencies ${}^E \eta_m^r$, ${}^M \eta_m^r$ for TE and TM waves and the transmitted diffraction efficiencies ${}^E \eta_m^t$, ${}^M \eta_m^t$ are given by

$${}^{E(M)} \eta_m^r = \frac{\text{Re}[\xi_m^a] |^{E(M)} g_m^{a-}|^2}{P_{\text{in}}}, \quad {}^{E(M)} \eta_m^t = \frac{\text{Re}[\xi_m^s] |^{E(M)} g_m^{s+}|^2}{P_{\text{in}}}, \quad (3.46)$$

where

$$\begin{aligned}
 P_{\text{in}} = & \text{Re}[\xi_0^a] |^E g_0^{a+}|^2 + 2 \text{Im}[\xi_0^a] \text{Im} \left[\left({}^E g_0^{a+} \right)^* {}^E g_0^{a-} \right] \\
 & + \text{Re}[\xi_0^a] |^M g_0^{a+}|^2 + 2 \text{Im}[\xi_0^a] \text{Im} \left[\left({}^M g_0^{a+} \right)^* {}^M g_0^{a-} \right]. \quad (3.47)
 \end{aligned}$$

Since the denominators of Eq. (3.46) become zero at a low grazing limit ($\xi_0^a \rightarrow 0$), we cannot express the discontinuity of diffraction efficiencies. From the relations of

$$\begin{aligned}
 {}^{E(M)} g_m^{a+} + {}^{E(M)} g_m^{a-} &= \delta_{m0} {}^{E(M)} g_0^{a+} + {}^{E(M)} g_m^{a-} = 2 \xi_0^a {}^{E(M)} S_m^{a-} \text{ and} \\
 {}^{E(M)} g_m^{s+} + {}^{E(M)} g_m^{s-} &= {}^{E(M)} g_m^{s+} = 2 \xi_0^a {}^{E(M)} S_m^{s+}, \text{ Eq. (3.45) becomes}
 \end{aligned}$$

$$\begin{aligned}
 & \text{Re}[\xi_0^a] |^E S_0^{a\oplus}|^2 + 2 \text{Im}[\xi_0^a] \text{Im} \left[\left({}^E S_0^{a\oplus} \right)^* \left(2 \xi_0^a {}^E S_0^{a-} - {}^E S_0^{a\oplus} \right) \right] \\
 & + \text{Re}[\xi_0^a] |^M S_0^{a\oplus}|^2 + 2 \text{Im}[\xi_0^a] \text{Im} \left[\left({}^M S_0^{a\oplus} \right)^* \left(2 \xi_0^a {}^M S_0^{a-} - {}^M S_0^{a\oplus} \right) \right] \\
 & = \sum_m \text{Re}[\xi_m^a] |2 \xi_0^a {}^E S_m^{a-} - \delta_{m0} {}^E S_0^{a\oplus}|^2 + \sum_m \text{Re}[\xi_m^a] |2 \xi_0^a {}^M S_m^{a-} - \delta_{m0} {}^M S_0^{a\oplus}|^2 \\
 & + \sum_m \text{Re}[\xi_m^s] |2 \xi_0^a {}^E S_m^{s+}|^2 + \sum_m \text{Re}[\xi_m^s] |2 \xi_0^a {}^M S_m^{s+}|^2. \quad (3.48)
 \end{aligned}$$

Here, ξ_m^s is the m th order eigenvalue in the substrate region, and ${}^{E(M)} S_0^{a\oplus} = {}^{E(M)} g_0^{a+}$ in the region 0. Using Eq. (3.48), we obtain new formulas of the reflected and transmitted diffraction efficiencies ${}^{E,M,R,L} \eta_m^r$ and ${}^{E,M,R,L} \eta_m^t$ for TE, TM, RC and LC waves as

- For a propagating incident wave (when ξ_0^a is real)

- Linearly polarized incidence case

$${}^{E(M)}\eta_m^r = \begin{cases} \frac{2\xi_0^{a-} S_0^{a-} - {}^{E(M)}S_0^{a\oplus}}{P_{in}'} m = 0 \\ \frac{4\xi_0^a \text{Re}[\xi_m^a] |{}^{E(M)}S_m^{a-}|^2}{P_{in}'} m \neq 0 \end{cases}, \quad {}^{E(M)}\eta_m^t = \frac{4\xi_0^a \text{Re}[\xi_m^s] |{}^{E(M)}S_m^{s+}|^2}{P_{in}'}, \quad (3.49)$$

- Circularly polarized incidence case

$${}^{R(L)}\eta_m^r = \begin{cases} \frac{2\xi_0^a S_0^{a-} - {}^E S_0^{a\oplus} \pm j(2\xi_0^M S_0^{a-} - {}^M S_0^{a\oplus})^2}{2P_{in}'} (m=0) \\ \frac{4\xi_0^a \text{Re}[\xi_m^a] |{}^E S_m^{a-} \pm j {}^M S_m^{a-}|^2}{2P_{in}'} (m \neq 0) \end{cases},$$

$${}^{R(L)}\eta_m^t = \frac{4\xi_0^a \text{Re}[\xi_m^s] |{}^E S_m^{s+} \mp j {}^M S_m^{s+}|^2}{2P_{in}'}, \quad (3.50)$$

where

$$P_{in}' = |{}^E S_0^{a\oplus}|^2 + |{}^M S_0^{a\oplus}|^2. \quad (3.51)$$

- For an evanescent incident wave (when ξ_0^a is imaginary)

- Linearly polarized incidence case

$${}^{E(M)}\eta_m^r = \begin{cases} 0 (m=0) \\ \frac{\text{Re}[\xi_m^a] |{}^{E(M)}S_m^{a-}|^2}{P_{in}''} (m \neq 0) \end{cases}, \quad {}^{E(M)}\eta_m^t = \frac{\text{Re}[\xi_m^s] |{}^{E(M)}S_m^{s+}|^2}{P_{in}''}, \quad (3.52)$$

- Circularly polarized incidence case

$${}^{R(L)}\eta_m^r = \begin{cases} 0 m = 0 \\ \frac{\text{Re}[\xi_m^a] |{}^E S_m^{a-} \pm j {}^M S_m^{a-}|^2}{2P_{in}''} m \neq 0 \end{cases}, \quad {}^{R(L)}\eta_m^t = \frac{\text{Re}[\xi_m^s] |{}^E S_m^{s+} \mp j {}^M S_m^{s+}|^2}{2P_{in}''} \quad (3.53)$$

where

$$P_{in}'' = \text{Re}\left[({}^E S_0^{s\oplus})^* {}^E S_0^{a-}\right] + \text{Re}\left[({}^M S_0^{a\oplus})^* {}^M S_0^{a-}\right], \quad (3.54)$$

which support that by use of the scattering factors, we can express the discontinuity of diffraction efficiencies even at a low grazing limit. ($\xi_0^a \rightarrow 0$). Here, a linearly polarized incident case is described by $(^E S_0^{a\oplus}, ^M S_0^{a\oplus}) = (\cos \gamma, \sin \gamma)$, and a circularly polarized incident case is described by $(^E S_0^{a\oplus}, ^M S_0^{a\oplus}) = (1/\sqrt{2}, \mp j/\sqrt{2})$. The signs “ \mp ” refer to right-circularly (RC) and left-circularly (LC) polarized waves.

Eq. (3.48) can be arranged to

$$\begin{aligned} \operatorname{Re} \left[\left(^E S_0^{a\oplus} \right)^* ^E S_0^{a-} \right] + \operatorname{Re} \left[\left(^M S_0^{a\oplus} \right)^* ^M S_0^{a-} \right] &= \sum_m \operatorname{Re} \left[\xi_m^a \right] \left| ^E S_m^{a-} \right|^2 \\ &+ \sum_m \operatorname{Re} \left[\xi_m^a \right] \left| ^M S_m^{a-} \right|^2 + \sum_m \operatorname{Re} \left[\xi_m^s \right] \left| ^E S_m^{s+} \right|^2 + \sum_m \operatorname{Re} \left[\xi_m^s \right] \left| ^M S_m^{s+} \right|^2, \end{aligned} \quad (3.55)$$

where the left-hand side is the loss of the reflection amplitudes, and the right-hand side is the total energy of diffraction. This should be understood as an optical theorem in the grating by the scattering factor. From Eq. (3.55), letting the modified diffraction efficiencies by

- Linearly polarized incidence case

$$^{E(M)} \overline{\eta}_m^r = \frac{\operatorname{Re} [\xi_m^a] \left| ^{E(M)} S_m^{a-} \right|^2}{P_{\text{in}}''}, \quad ^{E(M)} \overline{\eta}_m^t = \frac{\operatorname{Re} [\xi_m^s] \left| ^{E(M)} S_m^{s+} \right|^2}{P_{\text{in}}''}. \quad (3.56)$$

- Circularly polarized incidence case

$$^{R(L)} \overline{\eta}_m^r = \frac{\operatorname{Re} [\xi_m^a] \left| ^E S_m^{a-} \pm j ^M S_m^{a-} \right|^2}{2 P_{\text{in}}''}, \quad ^{R(L)} \overline{\eta}_m^t = \frac{\operatorname{Re} [\xi_m^s] \left| ^E S_m^{s+} \mp j ^M S_m^{s+} \right|^2}{2 P_{\text{in}}''}. \quad (3.57)$$

Eqs. (3.52), (3.53) and (3.56), (3.57) indicate that, the diffraction efficiency for an evanescent incidence and the optical theorem are based on the same conception. For the case of a lossless dielectric grating, the relations of energy conservation are given as follows:

$$\begin{aligned} \sum_m \left(^{E(R)} \eta_m^r + ^{M(L)} \eta_m^r + ^{E(R)} \eta_m^t + ^{M(L)} \eta_m^t \right) &= 1, \\ \sum_m \left(^{E(R)} \overline{\eta}_m^r + ^{M(L)} \overline{\eta}_m^r + ^{E(R)} \overline{\eta}_m^t + ^{M(L)} \overline{\eta}_m^t \right) &= 1, \end{aligned} \quad (3.58)$$

which are used in checking the accuracy of calculations.

3.6. Analysis of Low Grazing Scattering by Dielectric Gratings

In this section, we consider a dielectric grating with asymmetric triangular profiles, as shown in Fig. 3.2. Calculations are performed under the conditions that $\varepsilon_a = \mu_a = \mu_s = 1.0$, $\varepsilon_s = 3.0$, $\Lambda / \lambda = 1.25$, $d / \lambda = 0.4$, $a / \Lambda = 0.9$ and $b / \Lambda = 0.1$, which we treated in the two dimensional case [11]. The spatial harmonic expansion terms are truncated to $2M + 1 = 51$, the number of partitioned layers is chosen as $N - 1 = 10$, and the azimuthal angle of incidence is $\phi = 110^\circ$. We confirmed numerically that the energy error is less than 10^{-10} in all calculations.

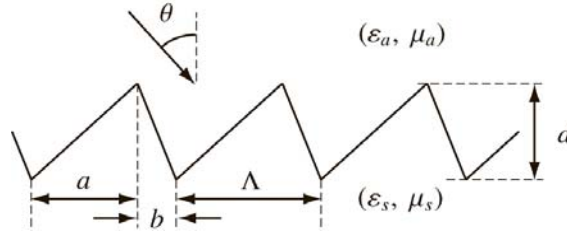


Fig. 3.2. The z - x cross-sectional diagram of an asymmetric triangular profiled grating.

We illustrate the reflected diffraction efficiencies against $\sin \theta$ for the linearly polarized incidence with a polarization angle $\gamma = 45^\circ$, as show in Fig. 3.3. Figs. (a) and (b) show the efficiencies of TE and TM waves, respectively. In the range of $|\sin \theta| < 1$, the results show the characteristics for a usual incidence. In the range of $|\sin \theta| > 1$, the results show for an evanescent wave incidence decaying into the x direction. The results for $|\sin \theta| = 1$ display the diffraction efficiencies at a low grazing limit. We found that when an incident angle becomes low grazing, no diffraction take place and the 0th reflected efficiencies of TE and TM waves become the intensities of TE and TM components in the incident wave, respectively.

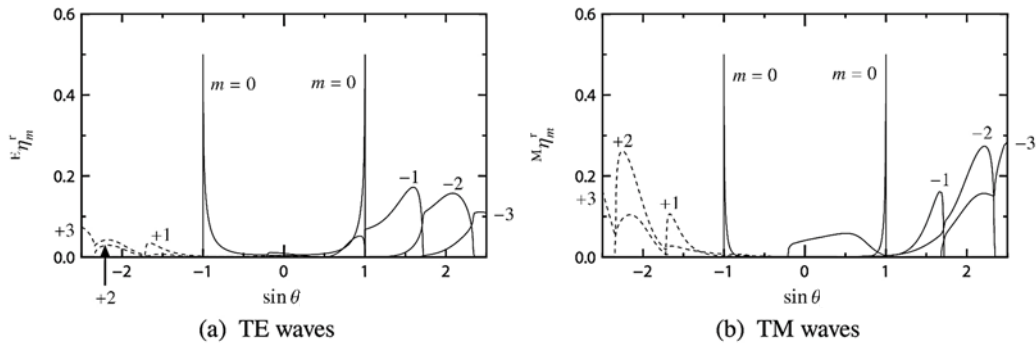


Fig. 3.3. The reflected diffraction efficiencies against $\sin \theta$ for linearly polarized incidence with $\gamma = 45^\circ$.

In Fig. 3.4, we plot the reflected diffraction efficiencies against $\sin\theta$ for RC incidence. From this figure, when $|\sin\theta|$ is small, the efficiencies of LC waves become larger. However, at $|\sin\theta| = 1$, the 0th order reflected efficiency of RC wave becomes unity, and any other efficiencies vanish. For a circularly polarized incidence case, when an incident angle is small, the reflection has a large component of a reverse circular polarization to the incident wave. When an incident angle approaches to a low grazing limit, only specular reflection of the same circular polarization occurs.

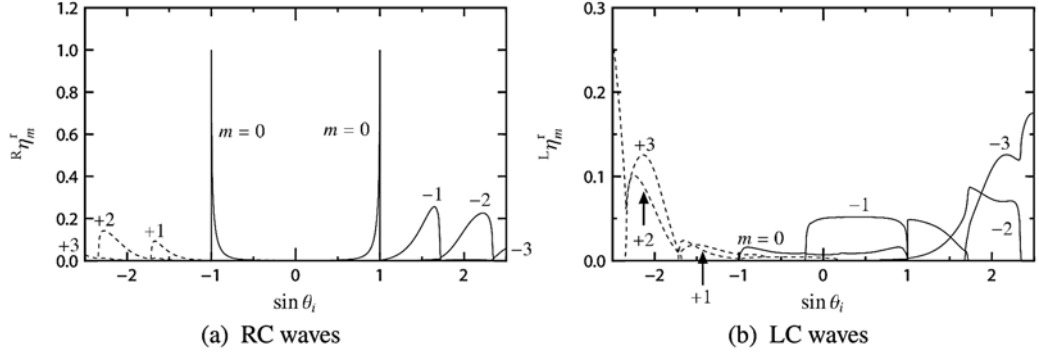


Fig. 3.4. The reflected diffraction efficiencies against $\sin\theta$ for RC incidence.

In order to investigate a singular behavior around a low grazing limit of circularly polarized incidence, we illustrate the phases and amplitudes of the reflected diffraction amplitudes ${}^{E(M)}g_m^{a-} (= 2\xi_0^a {}^{E(M)}S_m^{a-} - \delta_{m0} {}^{E(M)}g_0^{a+})$ against a grazing angle $(90 - \theta)^\circ$ from 0.00001° to 100° for RC incidence in Fig. 3.5(a) shows that when $(90 - \theta) \leq 1^\circ$, the reflection phases $\text{Arg}({}^E g_0^{a-})$ and $\text{Arg}({}^M g_0^{a-})$ are almost equal to 180° and 90° , respectively. Comparing with the incident phases $\text{Arg}({}^E g_0^{a+}) = 0^\circ$ and $\text{Arg}({}^M g_0^{a+}) = -90^\circ$, the incident and the specularly reflected waves in TE and TM wave components have the phase difference 180° when an incident angle becomes low grazing. Fig. 3.5(b) shows that when $(90 - \theta) \leq 1^\circ$, the amplitude $|{}^{E(M)}g_0^{a-}|$ is almost $1/\sqrt{2} (= |{}^{E(M)}g_0^{a+}|)$, and the amplitude $|{}^{E(M)}g_m^{a-}|$ ($m \neq 0$) decreases proportionally to $\xi_0^a = \sqrt{\varepsilon_a \mu_a} \sin(90 - \theta) \approx (90 - \theta)$ [3]. These facts of Figs. 3.5(a) and (b) support the characteristics at $|\sin\theta| = 1$ in Fig. 3.4. We can see that also in a circularly polarized incidence case like a linearly polarized incidence case, the incident region becomes a shadow at a low grazing limit, since no diffraction take place and only specular reflection occurs.

3.7. Analysis of Eigenvalues Degeneracy in the Middle Region

In the following calculations, putting $N=3$ and $n=1$ in Fig. 3.1, we consider a 4 layered dielectric grating having a periodic structure in region 2, as shown in Fig. 3.6. This is the structure which we treated in Ref. [12], and in this section, we set an azimuthal angle ϕ . h_1 is the thickness of the region 1. h_2 , Λ , W and ε_2 are the thickness, periodicity, width and relative permittivity of rectangular profiles in the region 2, respectively. $\varepsilon_0 (= \varepsilon_a$ in Fig. 3.1), ε_1 and $\varepsilon_3 (= \varepsilon_s$ in Fig. 3.1) are the relative permittivities of the incident, middle and substrate regions, respectively. The relative permeability is set to be $\mu = 1$ in all regions. Calculations are performed under $h_1 / \lambda = 1.0$, $h_2 / \lambda = 0.5$, $\Lambda / \lambda = 1.2$, $W / \Lambda = 0.6$, $\varepsilon_0 = \varepsilon_2 = \varepsilon_3 = 2.25$ and $\varepsilon_1 = 1.0$. The number of spatial harmonics is truncated to $2M+1=51$, and we confirmed numerically that the energy conservation error is less than 10^{-10} in all calculations. Under $\varepsilon_0 > \varepsilon_1$, there exists an incident angle $\sin^{-1} \sqrt{\varepsilon_1 / \varepsilon_0} = \theta_c^1$ where the 0th order eigenvalues $\xi_{1,0}$ in the middle region 1 degenerate by zero.

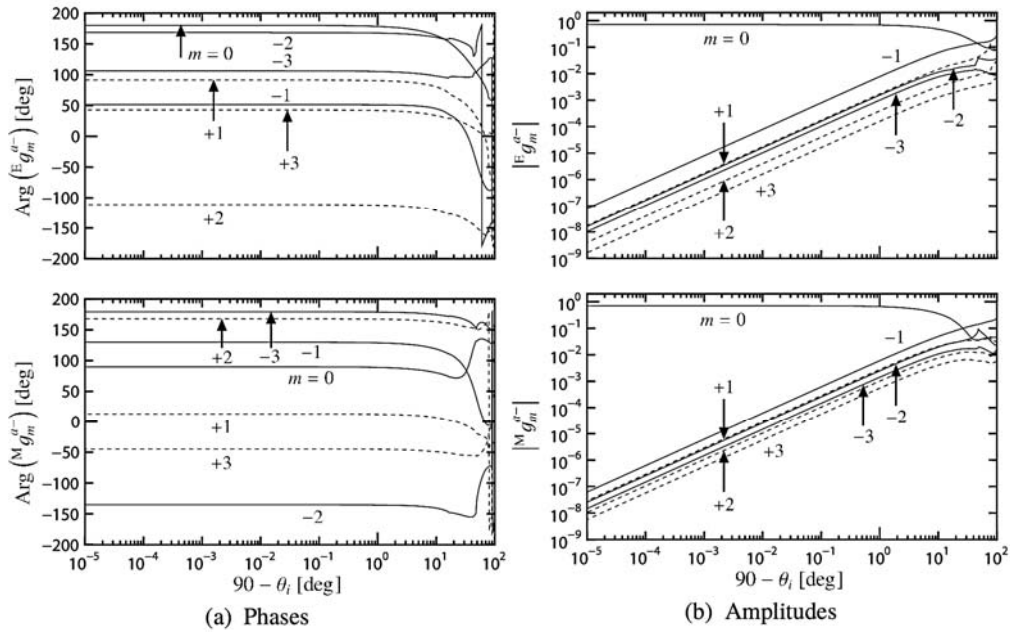


Fig. 3.5. Phases and amplitudes of the reflected diffraction amplitudes against $(90 - \theta)$ [deg] for RC incidence.

¹ We call this angle a critical angle. A critical angle is originally stated in the reflection problem of two lossless isotropic media separated by a planar interface.

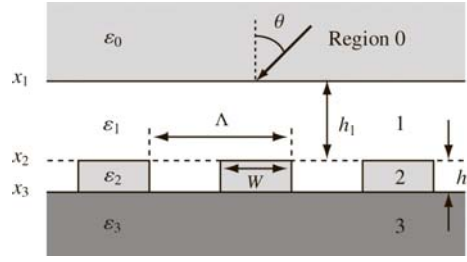


Fig. 3.6. Analytical model of a multilayered periodic structure.

First of all, we consider TM wave analysis in the two dimensional problem [12]. Setting $\phi = 90^\circ$ and $\gamma = 90^\circ$, we plot the m th order diffraction efficiencies against $s_0 = \sqrt{\varepsilon_0} \sin \theta$ in Figs. 3.7 (a) and (b) show the cases of reflected and transmitted waves, respectively. In these figures, $s_0 > 1.5$ means an evanescent wave incidence, and $s_0 = 1.5$ means a low grazing limit of incidence. We see that the diffraction takes place even for an evanescent wave incidence. In the range of $0 \leq s_0 \leq 2$, due to the 0th or high order eigenvalues degeneracy, the values of open circles "○" for 9 points of s_0 cannot be calculated by the conventional formulations. That is to say, the values of "○" are calculated by the shadow theory. In Ref. [12], we only formulated that discriminating eigenvalues closeto zero in periodic regions, the manipulation of the shadow theory was applied to only modes related by the eigenvalues. In this numerical calculation, we have for the first time discovered eigenvalues degenerating by zero in a periodic region. At 4 points of s_0 ($=0.368513294893446, 0.464820015880559, 1.201846803189671, 1.298153308870118$), the eigenvalues in the periodic region 2 degenerate numerically. At the remaining 5 points of s_0 which are analytically obtained, the eigenvalues in the uniform regions 0, 1 and/or 3 degenerate by zero. We gives formulations of applying the manipulation of the shadow theory to all modes in all regions, and actually shows that the values can be calculated even for the cases that the eigenvalues degenerate in any region. Accordingly, it turns out that, using the present formulations, stable numerical computations are carried out without noting the eigenvalues degeneracy. The degeneracy of eigenvalues in periodic regions for the conical mounting case however have not been found, although the investigations are ongoing.

Secondly, we assume $\phi = 110^\circ$ and $\gamma = 45^\circ$. In order to investigate a behavior on the eigenvalues degeneracy in the middle region, by Eqs. (3.42) and (3.43), we calculate the 0th mode diffraction amplitudes ${}^{\text{E(M)}}g_{1,0}^+(x_1)$ and ${}^{\text{E(M)}}g_{1,0}^-(x_1)$ in the region 1. Fig. 3.8 illustrates the phase differences and amplitude ratios of the diffraction amplitudes for $|\theta_c - \theta|$ from 0.000001° to 10° . When an incident angle θ approaches to θ_c , the phase differences are almost 180° , and the amplitude ratios are almost 1 in TE and TM waves. At a critical angle, in the region 1, the incoming TE and TM waves grazing along the boundary are cancelled by the reflected wave with the same amplitude and reverse phase.

This fact suggests that the manipulation in Eqs. (3.19) and (3.32) based on the primary excitation is applicable to the middle regions.

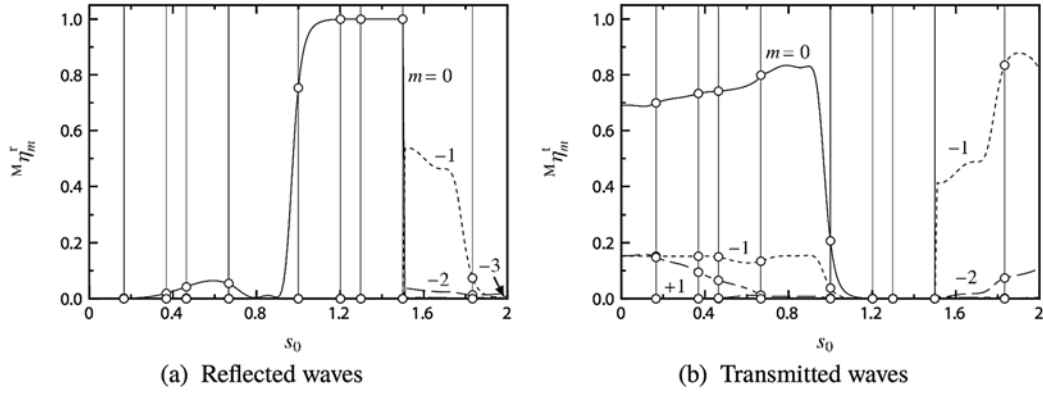


Fig. 3.7. The m th diffraction efficiencies against s_0 .

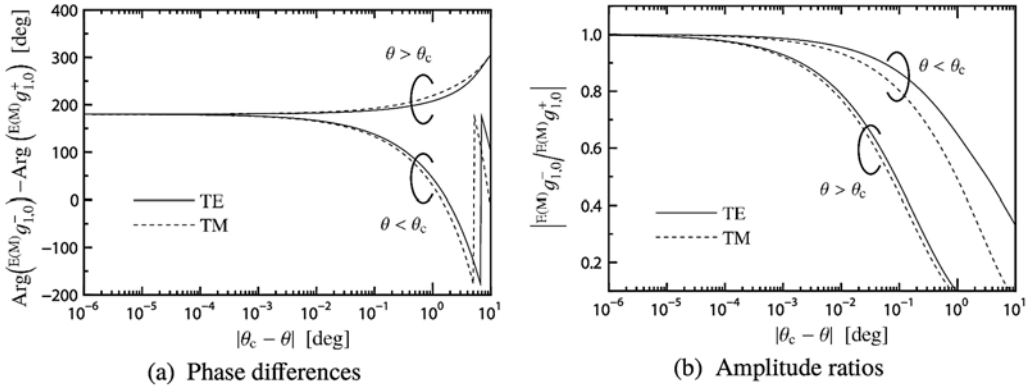


Fig. 3.8. The 0th diffracted amplitudes in the middle region 1.

We calculate the 0th mode field distributions around $\theta = \theta_c$ in Fig. 3.9(a) shows the electric fields $|e_{y0}|$, $|e_{z0}|$ and flux density $|d_{x0}|$ against x/λ . Fig. 3.9(b) shows the magnetic fields $|h_{y0}|$, $|h_{z0}|$ and flux density $|b_{x0}|$. We take $x_1 = 0$. The distributions in the periodic region ($1 \leq x/\lambda \leq 1.5$) are estimated values which are calculated by extracting the elements corresponding to the 0th mode. These figures indicate that we can obtain the distributions using a new description of the fields by the shadow theory even whether eigenvalues degenerate or not. In the two dimensional problem [12], at a critical angle of TE or TM incidence, the z components of the magnetic or electric fields become almost constants independent of x in the middle region, whereas Fig. 3.9 shows that the z components of the fields in addition to the y components decrease almost linearly with x

. Eqs. (3.19), (3.20) and (3.21)¹ supports that the fields are in proportion to x when the m th order eigenvalues degenerate by zero. The x components of the flux densities also decrease almost linearly with x . In a general case that the incident plane is not perpendicular to the grating grooves, all components of the fields in Cartesian coordinate decay proportionally to the position when the eigenvalues degenerate by zero in the middle region. In the substrate region ($x/\lambda \geq 1.5$), a forward wave propagates by tunneling effects even at a critical angle of incidence.

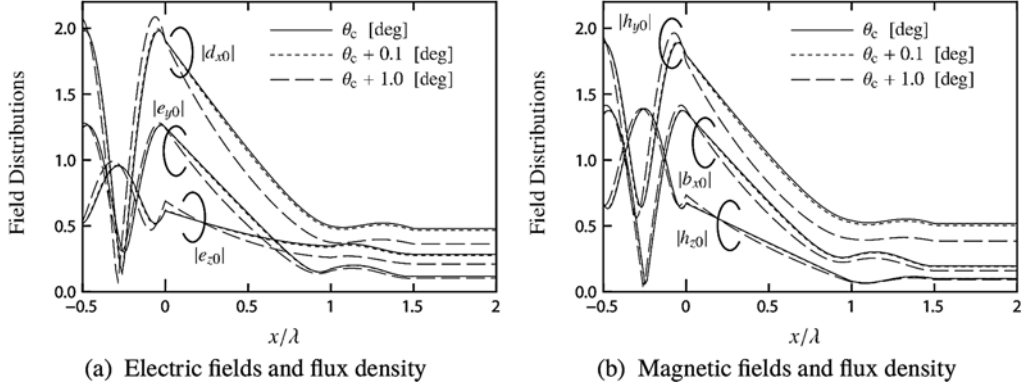


Fig. 3.9. The 0th mode fields distributions.

3.8. Conclusions

This chapter has proposed formulations of applying the shadow theory to the three dimensional scattering problem of multilayered dielectric gratings in a general case.

By newly introducing the oblique primary excitation, the scattering factors and the diffraction efficiencies through our analytical method by use of the matrix eigenvalues, we have resolved the problem of eigenvalues degeneracy and that of matrix diagonalization for any complex angle of oblique incidence. In the matrix eigenvalues method, we have applied the manipulation by the shadow theory to all modes of the electromagnetic fields in all regions. Getting new transformation and improved matrices of the shadow theory, we have presented a new description of the electromagnetic fields using the product of these matrices and scattering factors.

¹ Putting $\dot{x} = x_0 = 0$, $\ddot{x} = x_1$ and $\xi_m = 0$, the m th mode electromagnetic fields in the region 1 can be written by

$$\begin{bmatrix} e_{ym}(x) \\ e_{zm}(x) \\ h_{ym}(x) \\ h_{zm}(x) \end{bmatrix} = \begin{bmatrix} j\bar{s}_m \left(x - \frac{x_1}{2}\right) & -\bar{q}_0/\sqrt{\varepsilon_1} & -\bar{s}_m & 0 \\ -j\bar{q}_0 \left(x - \frac{x_1}{2}\right) & -\bar{s}_m/\sqrt{\varepsilon_1} & \bar{q}_0 & 0 \\ -\bar{q}_0 & -j\bar{s}_m\sqrt{\varepsilon_1} \left(x - \frac{x_1}{2}\right) & 0 & \bar{s}_m\sqrt{\varepsilon_1} \\ -\bar{s}_m & j\bar{q}_0\sqrt{\varepsilon_1} \left(x - \frac{x_1}{2}\right) & 0 & -\bar{q}_0\sqrt{\varepsilon_1} \end{bmatrix} \begin{bmatrix} E_m^\oplus \\ M_m^\oplus \\ E_m^- \\ M_m^- \end{bmatrix}$$

Numerical results have been given for dielectric gratings with asymmetric triangular profiles, placed in conical mounting. From the results, when an incident angle becomes low grazing, only specular reflection of the TE and TM wave components occurs, and in a circularly polarized incidence case, only specular reflection of the same circular polarization as the incident wave occurs. We conclude that, at a low grazing limit, the diffraction disappears and only reflection occurs with reflection coefficient -1 , regardless of the incident polarization. Additionally, numerical results have demonstrated that, we can perform stable computations without noting the eigenvalues degeneracy and the problem of the matrix diagonalization in any region. Then, we have specified that the eigenvalues degeneracy in the middle region corresponds to a phenomenon that the incoming wave and its reflected wave with the same amplitude and reverse phase graze along the boundary, and are cancelled out. We newly found that, when the eigenvalues degenerate by zero, all components of the fields in Cartesian coordinate decay proportionally to the position in the middle region.

In the future, we will formulate and introduce a transformation matrix and scattering factors for RC and LC waves directly [18].

References

- [1]. M. I. Charnotskii, Wave scattering by periodic surface at low grazing angles: single grazing mode, *Progress in Electromagnetic Research*, PIER 26, 2000, pp. 1-41.
- [2]. J. Nakayama, Shadow theory of diffraction gratings, *IEICE Transactions on Electronics*, Vol. E92-C, No. 1, 2009, pp. 17-24.
- [3]. J. Nakayama, Y. Tamura, Shadow theory of diffraction grating: reciprocity, symmetry and average filter, *IEICE Transactions on Electronics*, Vol. E97-C, No. 10, 2014, pp. 1036-1040.
- [4]. R. Petit, Electromagnetic Theory of Gratings, *Springer-Verlag*, Berlin, Heidelberg, New York, 1980.
- [5]. R. Ozaki, T. Yamasaki, T. Hinata, Scattering of electromagnetic waves by dielectric gratings with dielectric rectangular cylinders sandwiched between two multilayers, *IEEJ Transactions on Fundamentals and Materials*, Vol. 129, No. 10, 2009, pp. 718-724.
- [6]. K. Rokushima, J. Yamakita, Analysis of anisotropic dielectric gratings, *Journal of the Optical Society of America*, Vol. 73, No. 7, 1983, pp. 901-908.
- [7]. H. Wakabayashi, J. Yamakita, K. Matsumoto, M. Asai, Analysis of infinitely thin dielectric gratings with surface relief, *IEICE Transactions on Electronics*, Vol. J82-C-I, No. 11, 1999, pp. 175-183 (in Japanese).
- [8]. H. Wakabayashi, J. Yamakita, Analysis of thickness-profiled gratings for oblique incidence using approximate modeling by plane gratings with surface resistance, *Radio Science*, Vol. 44, No. 5, 2009, RS5009.
- [9]. M. G. Moharam, T. K. Gaylord, Rigorous coupled-wave analysis of planar-grating diffraction, *Journal of the Optical Society of America*, Vol. 71, No. 7, 1981, pp. 811-818.
- [10]. M. G. Moharam, T. K. Gaylord, Three-dimensional vector coupled-wave analysis of planar-grating diffraction, *Journal of the Optical Society of America*, Vol. 73, No. 9, 1983, pp. 1105-1112.
- [11]. H. Wakabayashi, M. Asai, K. Matsumoto, J. Yamakita, On a scattering field expression by dielectric diffraction gratings using shadow theory, *IEICE Transactions on Electronics*, Vol. J93-C, No. 3, 2010, pp. 81-90 (in Japanese).

- [12]. H. Wakabayashi, K. Matsumoto, M. Asai, J. Yamakita, Numerical methods of multilayered dielectric gratings by application of shadow theory to middle regions, *IEICE Transactions on Electronics*, Vol. E95-C, No. 1, 2012, pp. 44-52.
- [13]. H. Wakabayashi, M. Asai, K. Matsumoto, J. Yamakita, Numerical methods for composite dielectric gratings embedded with conducting strips using scattering factors, *IEICE Transactions on Electronics*, Vol. E96-C, No. 1, 2013, pp. 19-27.
- [14]. H. Wakabayashi, M. Asai, K. Matsumoto, J. Yamakita, Analysis of low grazing scattering by dielectric gratings in conical mounting using scattering factors, *IEICE Transactions on Electronics*, Vol. E97-C, No. 1, 2014, pp. 50-57.
- [15]. H. Wakabayashi, M. Asai, K. Matsumoto, J. Yamakita, Numerical method of applying shadow theory to all regions of multilayered dielectric gratings in conical mounting, *Journal of the Optical Society of America A*, Vol. 33, No. 11, 2016, pp. 2188-2197.
- [16]. L. Li, Use of Fourier series in the analysis of discontinuous periodic structures, *Journal of the Optical Society of America A*, Vol. 13, No. 9, 1996, pp. 870-876.
- [17]. T. Okada, H. Wakabayashi, H. Inai, Improved convergence in the analysis of dielectric gratings with permittivity profile, *IEICE Transactions on Electronics*, Vol. J97-C, No. 5, 2014, pp. 235-238 (in Japanese).
- [18]. H. Wakabayashi, M. Asai, J. Yamakita, A scattering field expression by dielectric gratings using circularly polarized waves, *IEICE Transactions on Electronics*, Vol. E98-C, No. 2, 2015, pp. 162-165.

Chapter 4

An Alternative Model of the Spatial Light Intensity Distribution

David Kaljun and Janez Žerovnik

4.1. Introduction

The standard model of the spatial light intensity distribution [16, 6] assumes that the best fit parameters of the model are defined via the usual 2-norm. In other words, the error of approximation is given by the average error. However, in particular when considering the light distribution of commercial LED lenses, it may be very undesirable to have most of the difference between the desired and designed light distribution concentrated in one (or very few) directions. This may mean that although the RMS error (Root Mean Square error, to be defined below) is relatively small, the LED product's design may not fit the users' needs or expectations. To avoid such disappointments when using automated or semi-automated design tools, minimizing the maximal error should be used. This leads to minimization of error of a new fitness function that does not correspond to minimization of a differentiable function. Consequently, some standard methods for solving the appropriate optimization problem, i.e. to find good or best fit approximations, cannot be applied directly. In this chapter, we propose an alternative model of the spatial light intensity distribution that is based on the ∞ -norm, giving rise to consideration of MAX error. We provide some evidence that the new model may in practical applications provide more reliable results. We emphasize that the large deviation in one (or very few) directions may be overseen when the standard RMS error is used, which may at least be inconvenient if not causing severe difficulties when testing (or even selling) the product.

In the chapter, we first give some motivation for the problem that is considered. In Section 4.3, mathematical model is defined formally. In Section 4.4, the main contribution of this work, experimental comparison of the standard model and the new alternative model is given. Within the section, the discrete algorithms used previously on the standard model are recalled, supporting the choice of one heuristics that is used in comparison experiment. Subsection 4.4.2 gives the outline of the two versions of the heuristics, one

for the standard, and one for the alternative model. Then, the experimental setup is given, and finally, in Subsection 4.4.4, the results are explained and discussed. Conclusions are drawn in Section 4.5.

4.2. Motivation and Related Previous Work

The LED (Light emitting diode) industry has been evolving rapidly in the past several years. The fast pace of research and development in the field had some expected impact. One of the results is a massive use and implementation of LED elements in all kind of luminaires. While some of these luminaires are designed for ambient illumination the majority are technical luminaires that have to conform not only to electrical and mechanical safety regulations but also to regulations that define and restrict the photometry of a certain luminaire. This means that the photometry of a luminaire has to be defined prior to the production. In order to do that efficiently and with minimal errors the design engineer must virtually test the luminaires performance. Tools that can be used (Speos, LighTools, TracePro) [17, 13, 19] do exist and they offer a vast repository of sub-modules to develop and design custom lenses, reflectors, light guides, etc. These universal tools however do not completely exploit the luminaire design possibilities that were introduced by the transition from conventional light source technologies to LED. One of the possibilities is to have an expert or intelligent system which would be capable of suggesting a secondary lens combination that would result in a user defined end photometry. This is the main goal of a bigger study that incorporates the research that is presented here. For more details see [4]. In other words, the system would take some stock secondary LED lenses from different manufactures, place them on a defined LED array and search for the optimal combination of the lenses so that the resulting photometry would be as close as possible to the user defined one. This method could enable the luminaire designer to custom design the light engine to a specific area of illumination, while keeping the mechanical and electrical parts of a luminaire untouched. This would in turn provide a customer with a tailored solution that would guarantee maximum efficiency, lower prices, fewer light pollution and the possibility to individualize the illumination effect while maintaining a consistent visual appearance of the luminaries. There are several optimization tasks related to development of the above idea.

The method mentioned above seems natural and straightforward, but at a closer look, we observe some fundamental problems related to realization of the main idea. Namely, both the spatial light distribution of LED lenses and the desired illumination are given in the standard data formats [1, 12], that are just long unstructured lists of data. In particular, when the aim is to construct a lighting system that provides the desired illumination of the environment, it is necessary or at least very convenient to have the data in some more structured format. It is known that the spatial light distribution of some LED lenses can be approximated by a sum of a small number of certain basis functions [16]. Provided the approximation is sufficiently good, it may be possible to provide designs combining several lenses with controlled error rate. This naturally opens several research avenues. For example, it is important to have error free or at least very good approximations of the basic lenses, and to have methods that are stable in the sense that they are not too sensitive to the noise in the presentation of basic elements. In this chapter we consider only the first

above mentioned task, approximation of the unstructured spatial light distribution data. [6, 16]. The problem that is defined formally in the next section is motivated by the following. The data describing the properties of the lenses and/or of the desired light distribution is nowadays usually given in some standard format files that correspond to the measured (or desired) values at a number of points in space. This results in relatively large data files of unstructured data. Clearly, if the data can be well enough approximated c.f. as a linear combination of certain basis functions, this may enable faster computations using less computer storage. In previous studies, the approximation problem was defined in terms of 2-norm, in other words, the aim was minimizing the least squares error, also called the RMS error. Several standard metaheuristics were adopted and tested on this problem in the past. In particular, it is known that parameters of the standard model can be computed efficiently, with small RMS error in acceptable computation time by various optimization algorithms including versions of the local search, genetic algorithms, and a combination called hybrid algorithm (Genetic + local search). Small RMS error here means less than 5 % [16], as the noise at measurement by current technology is at best 6-7 %. The approximations obtained by the above algorithms were usually much lower, i.e. below 5 %, and even better on artificial dataset [10]. However, on the other hand note that the human perception is much more limited, it is known that average human eye does not detect less than 20 % relative difference in luminous intensity. This is due to the complicated relations between perceived brightness, surface reflections and luminous flux [4, 20]. Application of the standard Newton's method proved that good initial solutions provided by discrete optimization were substantially improved, and on the other hand, the convergence of the Newton's method has been assured by the initial solutions of good quality [7].

For some special cases including LED lenses with symmetric light distribution, it is possible to find reasonably good approximations fast (8 minutes' runtime on an Intel Core I7-4790K CPU @ 4 GHz, the code is written in C++ and is not fully optimized). [6, 10] Sufficiently good approximation here means 2-5 % RMS error (to be defined later) for target light distribution, taking into account expected noise in measurement using current technology. More recent experiments showed that sufficiently good approximations can be obtained by some basic optimization algorithms, including local search algorithms and genetic algorithms [5-7, 9, 10]. However, when using predefined lenses to design a luminaire that closely approximates a desired light distribution, it may be essential that the approximation error is much lower. The same task can also be seen as solving a problem of data compression, replacing a long unstructured data file with a much shorter one, in this case a sequence of parameters. It makes sense to aim at 0 % approximation when considering the data compression task. As the functions to be approximated are smooth, it is natural to try to improve the basic discrete optimization methods with continuous optimization techniques, c.f. Newton's method [3, 7]). When the goal is to find as good approximations as possible that would enable further manipulation with the lenses at controlled error rate, it was shown that separating the linear and nonlinear parameters may dramatically improve the speed of the approximation algorithms [2, 9].

4.3. Mathematical Model

In the model used here [16], the Luminous intensity $I(\varphi; \mathbf{a}, \mathbf{b}, \mathbf{c})$ at the polar angle of φ is given in the form

$$I(\varphi; \mathbf{a}, \mathbf{b}, \mathbf{c}) = I_{max} \sum_{k=1}^K a_k \cos^{c_k}(\varphi - b_k), \quad (4.1)$$

where K is the number of functions to sum and a_k, b_k, c_k are the function coefficients that we search for. For brevity, coefficients are written as vectors $\mathbf{a} = (a_1, a_2, \dots, a_k)$, $\mathbf{b} = (b_1, b_2, \dots, b_k)$, and $\mathbf{c} = (c_1, c_2, \dots, c_k)$. It has been shown that the model is appropriate. More precisely, Moreno and Sun [16] observed that the far field radiation pattern of a LED (photometry) can be approximated with a sum of three functions and it was shown in [6] that the same holds for LED with attached secondary optics (see Fig 4.1) that provide symmetric radiation patterns. We extended the usability of the principles to asymmetric lenses which have different patterns on different C-planes (see Fig 4.2). Basically, we approximate every C-plane separately [4] where the model parameters of the previous C-plane as a starting point for the next C-plane are used, which slightly speeds up of the algorithm.

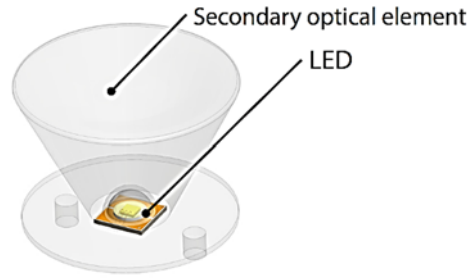


Fig. 4.1. LED with attached secondary lens.

Here we must note two restrictions on the model. The first restriction emerges from the LEDs physical design. The LED cannot emit any light to the back side which is the upper hemisphere in our case. That is why all intermediate values that are calculated at the combined angle $(\varphi - b_k)$ greater than 90 equals zero. The second restriction deals with the slightly unusual description of the light distribution in standard files such as Elumdat (file extension. ldt) [LDT] and Iesna (.ies) [IES]. These files present measured candela values per angle j on so called C planes. One C plane is only one half of the corresponding cross-section and does not describe the other half. But from a physical point of view we need to consider the impact from the other half of the cross-section.

In the studies [5-11], the goodness of fit is given by an expression derived from the one presented by Moreno and Sun [16]:

$$RMS(\mathbf{a}, \mathbf{b}, \mathbf{c}) = \sqrt{\frac{1}{N} \sum_{i=1}^N [I_m(\varphi_i) - I(\varphi_i, \mathbf{a}, \mathbf{b}, \mathbf{c})]^2}, \quad (4.2)$$

where RMS represents the error of the approximation, N the number of measured points in the input data, $I_m(\varphi_i)$ the measured Luminous intensity value at the polar angle φ from the input data, and $I(\varphi_i, \mathbf{a}, \mathbf{b}, \mathbf{c})$ the calculated Luminous intensity value at the given polar angle φ . The sum of squares of errors is the usual definition of approximation error, and is often the first choice in practical approximation. As the least squares method is very popular, it is theoretically well studied. Among other benefits, the Eq. (4.2) allows usage of continuous optimization methods, such as the Newton's method. Minimization of RMS is thus an obvious optimization goal.

As mentioned in the introduction, the fitness function defined by Eq. (4.2) may not be the best choice. Namely, when aiming to build a lighting system that delivers the light to the environment, large deviation in emitting light in only one direction may be prohibitive. In spite of low RMS error, the quality of such solution may be seen as very poor by the users, when too large deviation occurs in one or a few directions. Alternative definition of approximation quality is to minimize the maximal error. As already mentioned before, the new idea and the main contribution of this chapter is to show that the model in which approximation quality is measured in terms of maximal error may provide much better results. We later show several examples of typical phenomena that can occur, which show the potential benefits of using the alternative definition of approximation errors.

Formally, the alternative evaluation function used here is written as

$$MAXe(\mathbf{a}, \mathbf{b}, \mathbf{c}) = \max_i [I_m(\varphi_i) - I(\varphi_i, \mathbf{a}, \mathbf{b}, \mathbf{c})], \quad (4.3)$$

where $MAXe$ represents the maximum error, all other variables are as defined in Eq. (4.2). For easier comparison of the solution quality, the results are provided as relative RMS error and $MAXe$ error that are defined by Eqs. (4.4) and (4.5).

$$RMS_p(\mathbf{a}, \mathbf{b}, \mathbf{c}) = \frac{100 \cdot N \cdot RMS(\mathbf{a}, \mathbf{b}, \mathbf{c})}{\sum_{i=1}^N [I_m(\varphi_i)]} [\%], \quad (4.4)$$

$$MAXe_p(\mathbf{a}, \mathbf{b}, \mathbf{c}) = \frac{100 \cdot N \cdot MAXe(\mathbf{a}, \mathbf{b}, \mathbf{c})}{I_{max}} [\%]. \quad (4.5)$$

For easier evaluation and comparison between the discrete input data and the functional represented result data, we developed a custom application. The application consists of several modules such as the project preparation module, calculation module and result analysis module. The most interesting is the results analysis module which exhibits a simple and intuitive graphical user interface with two text input forms which are extended with the use of the system browser so that the file selection process is point and click rather than writing the absolute file path. The user can load one or two photometric files either in discrete or functional form. The results are then shown in the main window. The window supports overlaid objects so the user can visually assess the differences of the models (photometry). In Fig. 4.4 the loaded files are presented in polar graph form and in Fig. 4.3 the models are rendered as point clouds for the original in white and result in red color. One can also overlay surfaces and make it transparent or opaque depending on the current needs.

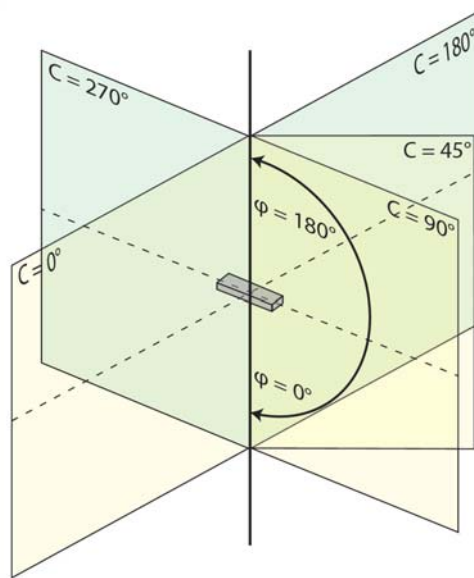


Fig. 4.2. C-planes according to standard. C-planes angles: 0 - 360 —F angles: 0 to 90.

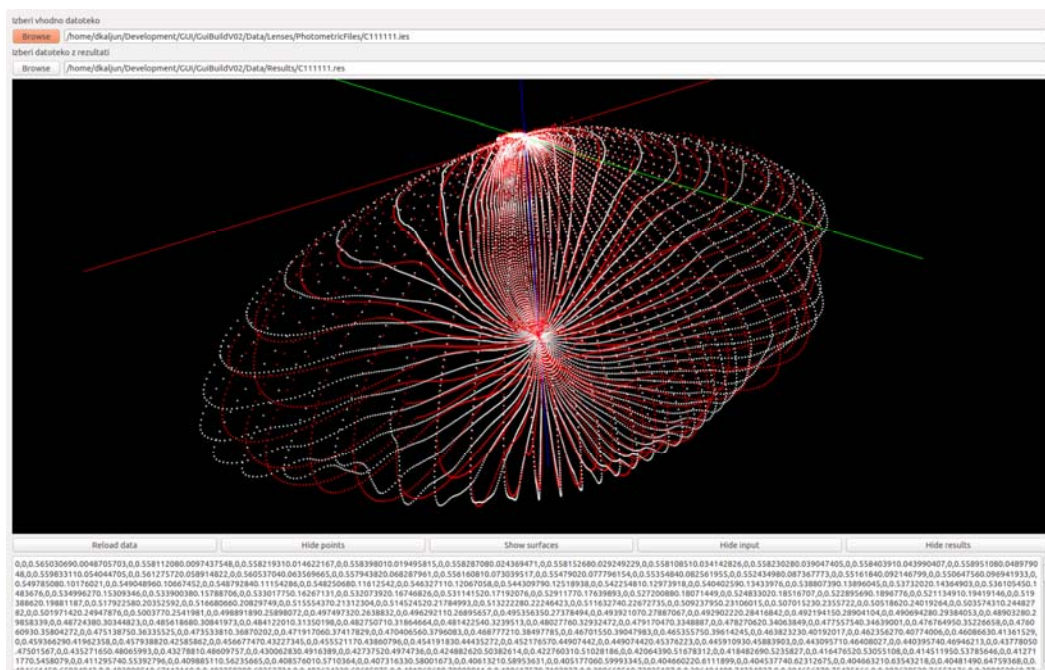


Fig. 4.3. Sample of modeled spatial intensity distribution in 3D view.

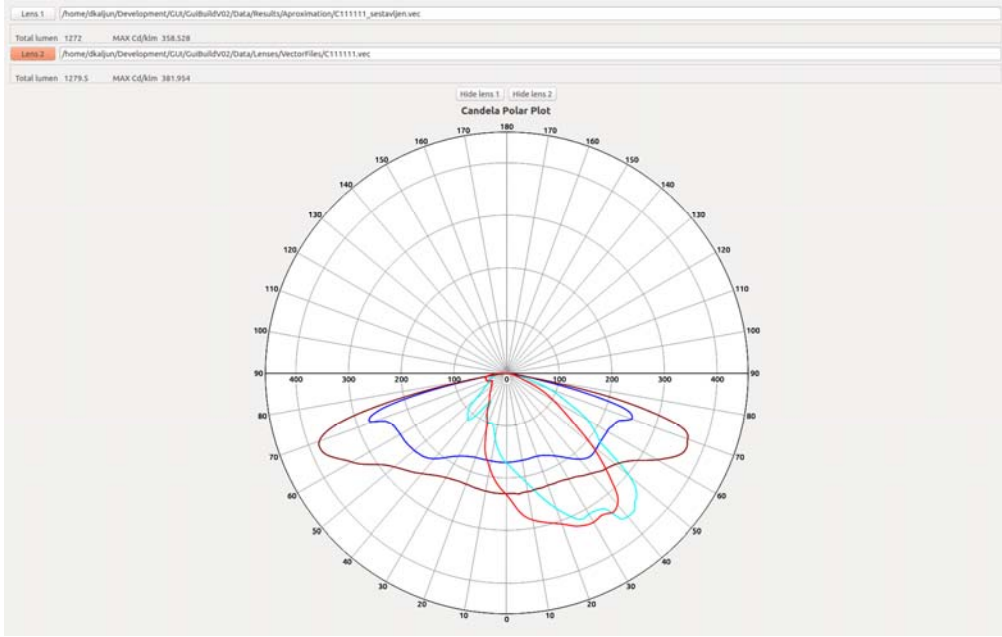


Fig. 4.4. Sample of modeled spatial intensity distribution in polar plot view.

4.4. Experiment – a Comparison of Two Error Measures

In the continuation, we provide some experimental evidence supporting our claim that the fitness function based on the maximum norm is more suitable for the application at hand. The experiment also shows that the model using alternative fitness function allows application of standard discrete optimization heuristics, in particular those that were used in the standard model.

We first briefly recall the outline of the algorithms that were used in previous work to obtain good approximations in terms of RMS error. In the experiments [5, 10, 11], the algorithm IF was found to be competitive, and we therefore chose it as the algorithm in the comparison test. Another convenient property of algorithm IF is its simplicity. In particular, it is very easy to adapt this algorithm to minimize the nonstandard error. Later we refer to the two versions of the algorithm as IF-RMS and IF-MAX.

In the rest of this section we first recall the algorithms used in previous studies. Then the two algorithms used in the experiment here are defined and finally, we present and discuss the results of the experiment.

We ran the experiment on a set of six lenses. As they are not rotational symmetric this leads to $6 \times 25 = 150$ instances of the basic optimization problem (on a single C-plane). The same algorithm is used, with two different goal (or, fitness) functions. As the focus, here is the comparison of the models, more precisely the fitness functions, we use only one algorithm, which is conceptually among the simplest, and at the same time has

provided solutions of very good quality in previous experiments [10]. As the version with standard fitness function has been used in several previous experiments, we ran several experiments to tune the parameters of the algorithm for the new fitness function. However, as the tuning period was much shorter in the second case it is likely that more tuning would result in better behavior of the new version.

4.4.1. The Algorithms Used in Previous Studies

To fully understand the drive and what lead us to the development of algorithms with hybrid evaluation functions which are showcased here, it may be helpful to present the algorithms that were developed prior to this. First, the steepest descend algorithm was used for optimizing the photometry description of a LED with mounted secondary optics [6]. Our original implementation worked, and the algorithm was able to provide usable results for all tested lenses from the Ledil catalogue [12] that we have chosen, however the performance was far from ideal. With search times ranging from 30 to 45 minutes per C-panel this just was not a feasible solution. That is why we moved forward in exploring different heuristic and metaheuristic methods. We adopted a version of iterative improvement algorithm, and some genetic algorithms including a hybrid between both of them. The results which are published in [10, 11], show that the most promising algorithm is the iterative improvement, provided the correct tuning. With the addition of multi-start with different initial search parameters and Newton's method end optimization we were able to lower the initial 30 minute per C-Panel to more manageable 4 seconds, which result in 8 minutes for a lens with 120 C-panels. A substantial speedup of optimization process was achieved after adoption of the idea of separation of linear and nonlinear parameters [9].

Below we describe in some more detail the algorithms used in previous studies. We focus on the discrete optimization algorithms because with a minimal change they can be used for approximation with alternative error functions, in particular with the alternative error function proposed here. Unfortunately, it is not that easy to use the Newton's method if at all possible as the alternative error function is not differentiable.

In the rest of this subsection we provide in some more detail the descriptions of the algorithms. Six different algorithms based on standard metaheuristics were designed (i.e. adapted) and tested on the standard model (using the RMS error). Three of them are well known standard algorithms for local search type heuristics, two are genetic algorithms which will be explained in detail bellow and one is a random sampling algorithm that we use to exclude any possible trivial solution.

The **steepest descent (SD)** algorithm begins with the initialization of the initial function parameter values that are ($a = [0, 1]$, $b = [-90, 90]$ and $c = [0, 100]$) with discrete sets ($a = [0, 0.001, 0.002, \dots, 1]$, $b = [-90, -89.9, -89.8, \dots, 90]$ and $c = [0, 1, 2, \dots, 100]$).

Next it initializes the search step values which are for $d_a = 0.01$, for $d_b = 1$ and for $d_c = I_{max}/10$ giving the 512 neighbors of the initial solution: $(a_1 \pm d_a, b_1 \pm d_b, c_1 \pm d_c, a_2 \pm d_a, b_2 \pm d_b, c_2 \pm d_c, a_3 \pm d_a, b_3 \pm d_b, c_3 \pm d_c)$.

If there is a neighbor with better RMS value, the search moves to the neighbor with minimal RMS value (if there are more minimal neighbors, all are chosen with the same probability). If none of the 512 is better than the current solution a new set of neighboring solutions are generated, this time with a double step. It repeats for ten steps and if there still is no better solution it breaks the search, multiplies the step value with 0.9, so the step is finer, and begins the search from the beginning about the current solution. The algorithm stops when the number of generated solutions reaches T_{max} .

The **iterative improvement with fixed neighborhood (IF)** algorithm initializes the same neighborhood as **SD**. Instead of considering all 512 neighbors at once, the algorithm generates a neighbor randomly, and moves to that neighbor if its RMS value is better than the current RMS value, regardless of the unchecked ones. If no better neighbor is found after 1000 trials, it is assumed that no better neighbor exists. As above, the algorithm changes to a new neighborhood, this time with a double step. It goes for ten steps and if there is still no better solution, it breaks the search, multiplies the step value with 0.9, so the step is finer and begins search from the start in the neighborhood of the current solution. The algorithm stops when the number of generated solutions reaches T_{max} .

The **iterative improvement with a variable neighborhood (IR)** algorithm begins as the previous two algorithms. It initializes the same initial function parameter values but a different neighborhood which has the search step value within a range, rather than a static fixed value. The ranges are for ($d_a = [0, 0.001, 0.002, \dots, 0.1]$, $d_b = [-9, -8.9, -8.8, \dots, 9]$ and $d_c = [-10, -9, -8, \dots, 10]$). It begins generating solutions, using the step range around the initial solution and calculating their RMS error.

As soon as it generates a better solution, it stops, shifts the focus on that solution, resets the step range to the initial value, and continues the search in the neighborhood of the new best solution. If after four hundred thousand generated solutions no better solution is found, the step range gets doubled, and the search continues in the current neighborhood with a larger neighborhood. The stopping condition is the same as before.

The **genetic algorithms** used in the experiment and described below do mimic the evolutionary behavior but may differ a bit from the standards. Especially we would like to note that the algorithms were rewritten opposing ones that were presented in [5].

The standard genetic algorithm addressed in the results as **SGA**, uses three genetic operators: selection, cross-breeding and mutation. The selection [3, 14, 18] operator works as a kind of a filter where fitter individuals in a population get to have higher weights as the less fit. This is then transmitted to the cross-breeding operator in the way that the individuals with higher weights are more likely to be chosen as parents as opposed to the others. The cross-breeding or crossover operator [3, 18] is where a population is created by generating new solutions.

These are created by randomly combining and crossing parameters from two randomly chosen parent solutions from the current population. The crossing is done via cross point so that every parent pair produces a pair of children. The cross point is chosen randomly and the children are generated in the flowing sequence $C_1 = [P_1^{bCP}, CP, P_2^{aCP}]$ and

$C_2 = [P_2^{bCP}, CP, P_1^{aCP}]$, where C_n is the child being generated, CP is the cross-point parameter, P_n^{aCP} are all of the parent's parameters that are after the CP and P_n^{bCP} are all of the parent's parameters that are before the CP. The last operator in every generation is the self-adapting mutation operator which finalizes the individuals in the new population. The mutation operates in the following manner: in the randomly chosen individual, a random number of parameters are chosen to be changed (mutated) which is done by adding a randomly chosen value for $da_1 = da_2 = da_3 = [-0.01, -0.009, -0.008, \dots, 0.01]$, for $db_1 = db_2 = db_3 = [-0.25, -0.24, -0.23, \dots, 0.25]$ and $dc_1 = dc_2 = dc_3 = [-2.5, -2.4, -2.3, \dots, 2.5]$ to the current parameter value.

The whole algorithm begins with the generation and calculation of the initial population (the zero population). Next it sorts the population entities from the fittest to the least fit and applies weights to them. After the sorting process, the algorithm generates with the crossover operator the next generation, which is then submitted to mutation with the adaptive mutation operator. When the new generation is fully formed the algorithm begins the process from the point of selection. It continues to do so until the last generation is finalized. The number of generations to be generated is calculated as the quotient of the maximal number of iterations minus the population size and the population size $N_G = (T_{max} - N_P) / N_P$.

The genetic algorithms gave us some encouraging results but the preliminary tests in [10] showed that a hybrid genetic algorithm should provide even better ones. That is why we developed an adapted standard genetic algorithm, which is named the hybrid genetic algorithm and addressed as **HGA**. The hybrid genetic algorithm works in the same way as the standard one but with an extra operator before the crossover. It starts with generating the initial solution and sorts the entities in the current solution from the fittest to least fit. Then instead of directly cross breeding the new generation it first runs the iterative improvement with fixed neighborhood algorithm on 10 best entities of the current generation which in turn get locally optimized (enhanced) for a number of iterations. After that the HGA follows the same path as the standard genetic algorithm does. For the number of generations to be executed on HGA algorithm, the formula is a bit more complicated, because it has to include the iterations of the local search. The formula can be written as $N_G = (T_{max} - N_P) / (N_P + 10 \times N_{iter})$.

4.4.2. The Algorithms IF-RMS and IF-MAX

As the IF algorithm proved to be the most successful and scalable algorithm of the group developed and described above we decided that the algorithm used here is a discrete optimization algorithm, we replace the interval ranges of the parameters ($a = [0, 1]$, $b = [-90, 90]$ and $c = [0, 100]$) with discrete sets ($a = [0, 0.001, 0.002, \dots, 1]$, $b = [-90, -89.9, -89.8, \dots, 90]$ and $c = [0, 1, 2, \dots, 100]$). The iterative improvement (**IF**) heuristic algorithm begins with the initialization of the initial function parameter values that are $a_1 = a_2 = a_3 = 0.5$, $b_1 = b_2 = b_3 = 0$, and $c_1 = c_2 = c_3 = 1$. Next it initializes the search step values which are for $d_a = 0.01$, for $d_b = 1$ and for $d_c = I_{max}/10$ giving the 512 neighbors of the initial solution: $(a_1 \pm d_a, b_1 \pm d_b, c_1 \pm d_c, a_2 \pm d_a, b_2 \pm d_b, c_2 \pm$

$d_c, a_3 \pm d_a, b_3 \pm d_b, c_3 \pm d_c$). The algorithm generates a neighbour randomly, and moves to that neighbour if its fitness function value is better than the current fitness value, regardless of the unchecked ones. If no better neighbor is found after 1000 trials, it is assumed that no better neighbor exists. If this happens, a new set of neighbors is generated, this time with a double step. It repeats for ten steps and if there still is no better solution it breaks the search, multiplies the step value with 0.9, so the step is finer, and begins the search from the beginning in the neighborhood of the current solution. The algorithm stops when the number of generated solutions reaches T_{max} . We call the versions using the fitness functions *RMS* (4.4) and *MAXe* (4.5) respectively as IF-RMS algorithm and IF-MAX algorithm.

Below we give results of an experiment to compare the results of the two algorithm. The results are closely examined and used to support our claim that the MAX error should be used for achieving more reliable practical results.

We wish to only note here that the idea of two approximation algorithms based on different error measures may be of independent interest. When the optimal approximation i.e. zero error solutions coincide, the algorithms may be combined in a way analogous to the idea of variable neighborhood search [15] in discrete optimization. For an initial study and discussion of the idea see [8].

4.4.3. Setup

We have set-up an experiment with six real lenses. Each lens has 25 measured C-planes which means that the overall count of individual C-planes is 150. On each C-plane of each lens we ran a multi-start version of the IF algorithm, such that each short run is allowed four hundred thousand iterations, $T_{max} = 400000$. Such a number of individual test instances is sufficient for a comparison of the influence of the two fitness functions used.

4.4.4. Results

We summarize the results obtained by the two algorithms in Table 4.1 and Table 4.2.

- First compare the approximation error in terms of the fitness functions used in each of the optimization algorithms (columns 2-4 of Table 4.1 and 5-7 of Table 4.2). From this comparison, the standard algorithm seemingly clearly beats the new version, but note that the measures of quality used here are not the same!
- Comparison of the results using the same measure is seen in Table 4.1 (comparing RMS error of the final results of both algorithms) and Table 4.2 (comparing MAX error of the final results of both algorithms). Clearly, IF-RMS outperforms IF-MAX in measure used in Table 4.1, while the superiority of IF-MAX over IF-RMS in MAX error is not that substantial. However, comparing the columns 4 and 7 in Table 4.2 shows that the worst results of IF-MAX are considerably worse than that of IF-MAX.

- On the positive side, we can conclude from the information given in Table 4.1 that both algorithms achieved low approximation errors on five of the six lenses. An exception is the lens CA12087 where the average error is acceptable (6 % and 8 %) but the maximal error is over 14 % for RMS and over 21 % for the MAXe. The conclusion based on the cumulative results from Table 4.1 may be that both algorithms can usually be expected to provide acceptable approximations of comparable quality.
- Table 4.2 on the other hand shows that solutions of both algorithms do have large MAX error in although they are seemingly very good, in terms of the RMS error. Below (Figs. 4.5-4.10) we look closely at the results at some random C-planes selected by hand to show various phenomena that were observed.

Table 4.1. Algorithm comparison - RMS error.

Lens	RMS algorithm RMS error			MAX algorithm RMS error		
	Best	Average	Worst	Best	Average	Worst
C10818	2.1021	3.3288	4.8276	2.6412	5.2852	8.6834
C10949	1.2023	2.1319	2.9317	1.9277	2.9424	4.5883
CA11416	1.5784	2.6409	4.5131	2.3757	3.353	5.2343
CA11426	1.9361	3.5252	6.4964	2.7196	4.8998	8.5563
CA12050	1.5788	2.8181	4.5389	1.9821	3.7464	6.7437
CA12087	2.149	5.4651	14.1541	2.854	7.6331	21.6723

Table 4.2. Algorithm comparison - MAX error.

Lens	RMS algorithm MAX error			MAX algorithm MAX error		
	Best	Average	Worst	Best	Average	Worst
C10818	5.428	8.7847	12.9902	5.0801	9.2589	14.2687
C10949	3.0165	5.6989	8.7168	3.1546	5.067	7.1241
CA11416	4.1378	7.3093	13.3517	3.974	6.0391	9.6082
CA11426	5.5003	9.8205	22.8541	5.6641	8.8768	15.089
CA12050	3.9828	7.6927	14.8069	3.5539	6.6315	11.1721
CA12087	6.0859	19.4859	64.3265	5.3171	13.2861	37.4918

Typically, the average error is lower for IF-RMS solutions (red) than for the IF-MAX solutions. However, looking at the maximal error, in most cases the situation is opposite. Usually, the numbers are close, such as in Fig. 4.5, while in some cases, the maximal error of the IF-RMS solution is significantly higher such as in Fig. 4.7, Fig. 4.8 and Fig. 4.9. This is the phenomena that may result in a practically bad solution, which may be refused by users of the product, despite low RMS approximation error. We wish to emphasize that a seemingly excellent result of approximation of the IF-RMS algorithm may hide a prohibitive large error in one direction. This cannot happen with IF-MAX, and in this sense the results of the new algorithm are more reliable. For completeness, we provide an example of a C-plane where both algorithms behave poorly (Fig. 4.10), but note that also

in this case the maximal error of IF-RMS is about 47 % which is considerably more than 26 % of IF-MAX solution while the average error 7 % of IF-RMS is only about half of the IF-MAX error 13 %.

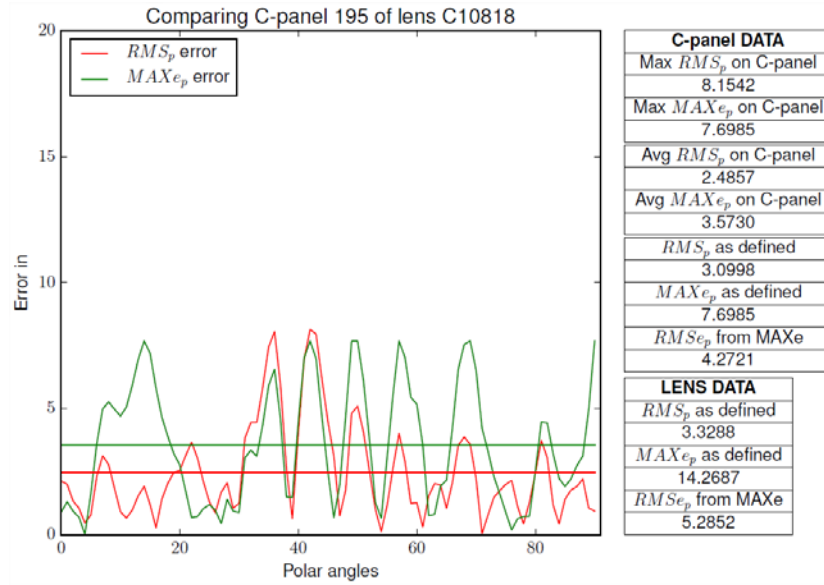


Fig. 4.5. MAXe and RMS error comparison on C-panel 195 of C10818.

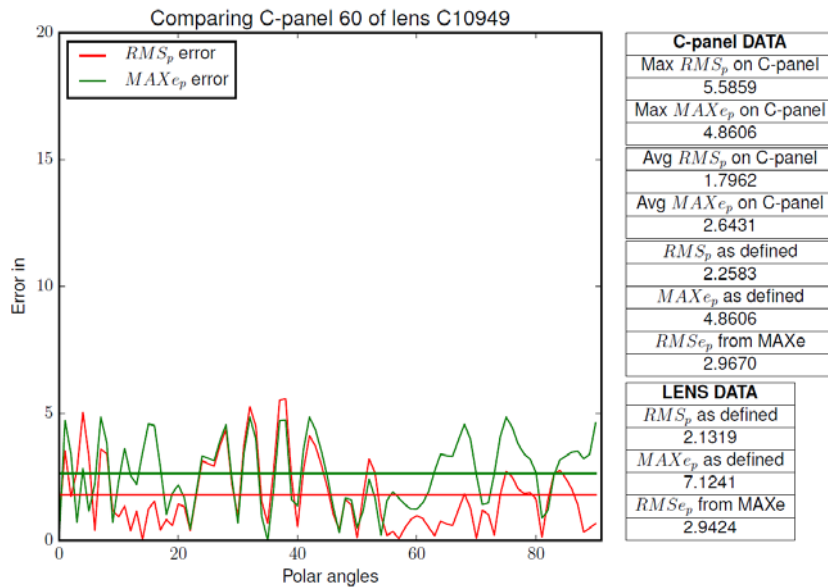


Fig. 4.6. MAXe and RMS error comparison on C-panel 60 of C10949.

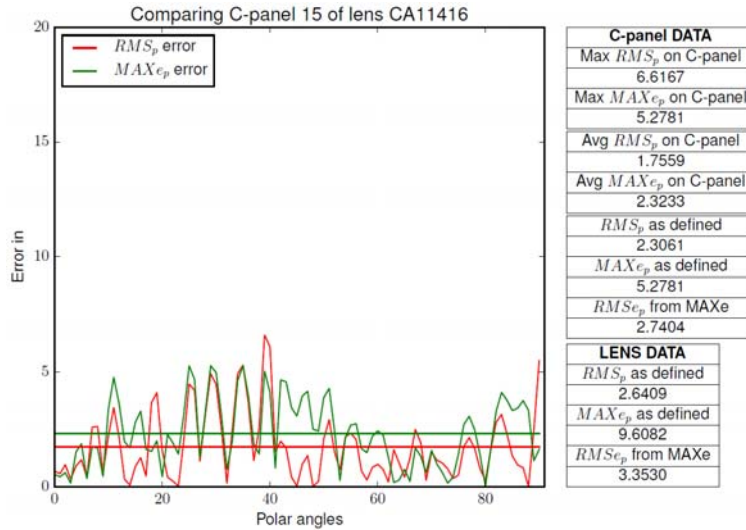


Fig. 4.7. MAXe and RMS error comparison on C-panel 15 of C11416.

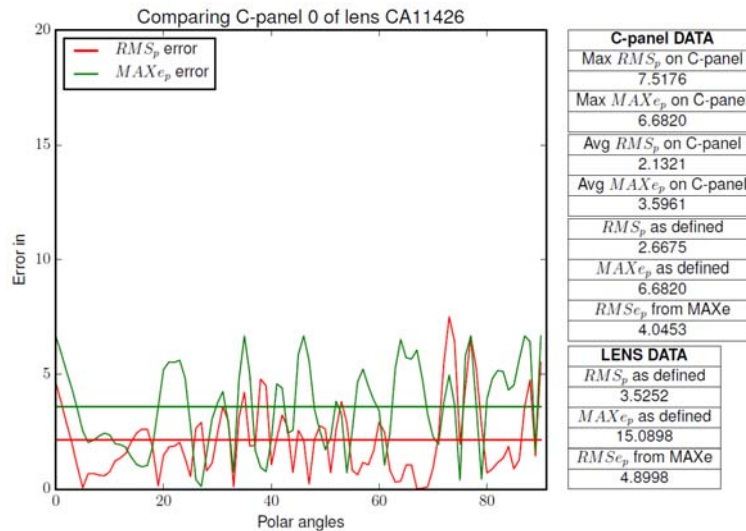


Fig. 4.8. MAXe and RMS error comparison on C-panel 0 of CA11426.

4.5. Conclusion

Here we proposed the use of a different fitness function for the field radiation pattern model of a LED with attached secondary optics. The new fitness function is based on the maximum error, compared to the average error of the RMS function. We show that the best fit parameters with the new fitness function (MAXe) can be efficiently computed by existing discrete optimization algorithms. While the solutions obtained by the optimization algorithm based on the new fitness function are of slightly lower quality

measured in the standard RMS measure, they are of better quality in the sense of reliability and consistency over all C-panels. Namely, low MAXe error guarantees that the solution is free of large deviations from expected values in all directions. In conclusion, we believe that the proposed fitness function should be considered as an important addition to the standard RMS fitness function, and will provide appropriate results where the RMS function fails to do so. With that in mind the overall approximation quality will be increased.

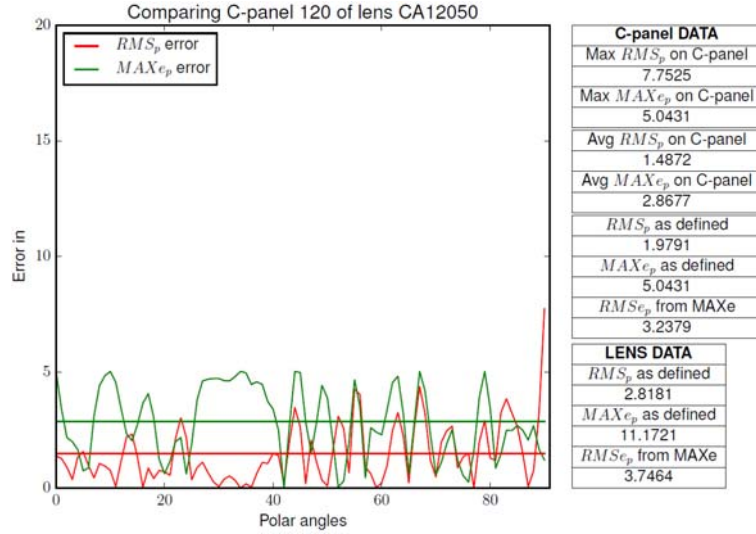


Fig. 4.9. MAXe and RMS error comparison on C-panel 120 of CA12050.

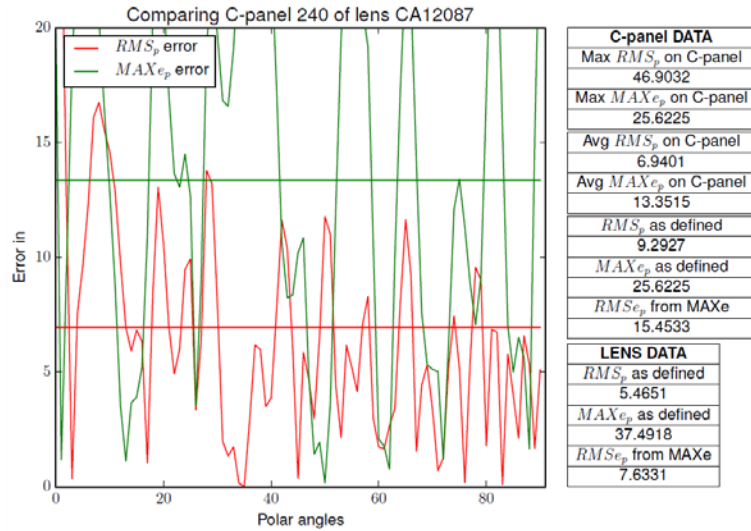


Fig. 4.10. MAXe and RMS error comparison on C-panel 240 of CA12087.

References

- [1]. I. Ashdown, Thinking photometrically - Part II, in *Proceedings of the LIGHTFAIR 2001 Pre-Conference Workshop*, March 2001.
- [2]. G. H. Golub, V. Pereyra, The differentiation of pseudo-inverses and nonlinear least square problems whose variables separate, *SIAM J. Numer. Anal.*, Vol. 16, 1973, pp. 413-432.
- [3]. R. L. Haupt, S. E. Haupt, Practical Genetic Algorithms, 2nd Ed., *John Wiley & Sons*, 2004.
- [4]. D. Kaljun, Optical system design optimization of a LED luminaire, PhD Thesis, *University of Ljubljana*, 2016, (in Slovene).
- [5]. D. Kaljun, J. Žerovnik, Local search optimization of a spatial light distribution model, in *Proceedings of the Bioinspired Optimization Methods and their Application Workshop (BIOMA'14)*, Jožef Stefan Institute, Ljubljana, 13th September 2014, pp. 81-91.
- [6]. D. Kaljun, J. Žerovnik, Function fitting the symmetric radiation pattern of a LED with attached secondary optic, *Optics Express*, Vol. 22, 2014, pp. 29587-29593.
- [7]. D. Kaljun and J. Petrišič and J. Žerovnik, Using Newton's method to model a spatial light distribution of a LED with attached secondary optics, *Strojniški Vestnik - Journal of Mechanical Engineering*, Vol. 62-5, 2016, pp. 307-317.
- [8]. D. Kaljun, J. Žerovnik, Improving approximation by switching between two error functions. *Croatian Operational Research Review CRORR*, Vol. 8, 2017, pp. 107-118.
- [9]. D. Kaljun, T. Novak, J. Žerovnik, Improved approximation of spatial light distribution, *PloS One*, Vol. 12, 2017, pp. 1-16.
- [10]. D. Kaljun, D. Rupnik Poklukar, J. Žerovnik, Heuristics for optimization of LED spatial light distribution model, *Informatica: an International Journal of Computing and Informatics*, Vol. 39, No. 2, 2015, pp. 147-159.
- [11]. D. Kaljun, J. Žerovnik, Developing LED Illumination Optics Design, Chapter 6, in *Advances in Evolutionary Algorithms Research* (G. Papa, Ed.), Vol. 1 (Computer Science, Technology and Applications), *New York: Nova Science*, 2015, pp. 123-138.
- [12]. Ledil Oy, <http://www.ledil.com/products/?y>
- [13]. LightTools, <https://www.synopsys.com/optical-solutions.html>
- [14]. M. Mitchell, An Introduction to Genetic Algorithms, 5th Edition, *The MIT Press*, 1999.
- [15]. N. Mladenović, P. Hansen, Variable neighborhood search, *Comput. Operations Res.*, Vol. 24, 1997, pp. 1097-1100.
- [16]. I. Moreno, C.-C. Sun, Modeling the radiation pattern of leds, *Optics Express*, Vol. 16, 2008, pp. 1808-1819.
- [17]. Speos, <http://www.optis-world.com/product-offering-light-simulation-virtual-reality-software/SPEOS>
- [18]. E.-G. Talbi, Metaheuristics: From Design to Implementation, *John Wiley & Sons*, 2009.
- [19]. TracePro, <https://www.lambdare.com/tracepro/>
- [20]. S. J. Williamson, H. Z. Cummins, Light and Color in Nature and Art, *John Wiley and Sons*, 1983.

Chapter 5

Optical Beams in Linear and Nonlinear Media

Dongmei Deng and Qi Guo

5.1. Introduction

Optical beam propagation is very important and has drawn much attention from many authors for a long time [1-20]. Radially polarized beams (RPBs) have attracted much interest recently due to their unique features [21, 22] in many research fields, for example, acceleration techniques [23, 24], particles guiding or trapping [25, 26], high resolution microscopy [27], and particularly material processing [28, 29]. The elegant Hermite–Laguerre–Gaussian beams (EHLGBs) [30] constitute the exact and continuous transition modes between elegant Hermite–Gaussian and elegant Laguerre–Gaussian beams when an additional parameter continuously changes. Pearcey beams (PBs) [31] possess the self-accelerating, auto-focusing and self-healing properties. The parabolic rotational coordinate beams [32] exhibit the parabolic rotation symmetry during propagation, with their envelopes following along parabolic trajectories. Optical beams which have subwavelength scales yet large bending angles are desirable for many applications such as near-field optical microscopes, plasmonics, and optical manipulation of nanoparticles [33-35]. When the optical beams have large beam divergence or large beam bending, the conventional paraxial theory is no longer appropriate. Therefore, free-space nonparaxial propagation of optical beams continues to be a great deal of attention. There are a few approaches to analyze nonparaxial optical beams including the vector angular spectrum method; based on the high-order corrections to the known paraxial solution [36-38]; the complex point source method [7-9]; the exact analytical solution method from wave equations without the paraxial approximation.

In this chapter, we have given the analytical transverse electric (TE) term and the transverse magnetic (TM) term of the RPBs [39] by means of the vector angular spectrum. We apply the complex point source approach introduced by Deschamps [7] and

systematically extended by Felsen [8] and Shin and Felsen [9] to derive the virtual sources for generation of the EHLGBs [30] and the Pearcey waves [40]. We have demonstrated the nonparaxial beams of parabolic rotational coordinates [32] by solving the exact solution of the Helmholtz equation in parabolic rotational coordinates.

On the other hand, much attention has been paid to the study of the spatial optical solitons which are self-trapped optical beams existing by means of the balance between diffraction and nonlinearity due to their potential applications to photonic switching [41], all-optical switching and logic gating [42], and all-optical signal processing [43]. The propagation of optical beams in the nonlocal nonlinear media satisfies the nonlocal nonlinear Schrödinger equation (NNLSE) [41, 44, 45]. Snyder and Mitchell [41] simplified the NNLSE to the Snyder–Mitchell model (SMM) [46] under the condition of strong nonlocality in 1997. Recent experimental results show that nematic liquid crystals [47, 48] and lead glass [49, 50] are strongly nonlocal nonlinear media (SNNM), which has encouraged further theoretical studies of the intriguing properties of self-trapped optical beams [49–53]. In this chapter, we have derived the Ince-Gaussian beams [54] and elliptic beams [55] that form the third complete family of exact and orthogonal solutions of the SMM by applying an elliptic coordinate transformation; we have obtained the analytical complex-variable-function (CVF)-Gaussian solitons [56, 57] that form the exact solutions of the SMM through a rotating coordinate transformation.

5.2. Nonparaxial Beams in Linear Media

Due to the fast advances of science and technology in micro- and nano-scales, optical beams with extremely small widths, which are comparable or even smaller than the light wavelength, can not avoid a variety of applications including near-field optical microscopes [58] and optical manipulation of nanoparticles [59]. Under such conditions, the divergence angles of the light beams become very large. Consequently, the conventional paraxial theory cannot describe the beam propagation. Therefore, free-space propagation of electromagnetic beams beyond the paraxial approximation continues to keep a topic of central interest in optics.

5.2.1. Nonparaxial Radially Polarized Beams

Here, the analytical TE and TM terms of the RPBs have been obtained by applying the vector angular spectrum. This particular electromagnetic field is entirely TM and it has only an on axis longitudinal (z) electric-field component (i.e., no on axis transverse electric field and no on axis magnetic field at all). The electric field distribution of the RPB at the initial plane ($z = 0$) can be expressed as [10, 15, 39]

$$\vec{E}_{nl}(x, y, 0) = E_{nlx}(x, y, 0)\vec{e}_x + E_{nly}(x, y, 0)\vec{e}_y, \quad (5.1)$$

where

$$E_{n1x}(x, y, 0) = \sqrt{2} E_0 / w_0 L_n^1 \left(\frac{2\rho^2}{w_0^2} \right) \exp \left(-\frac{\rho^2}{w_0^2} \right) x, \quad (5.2)$$

$$E_{n1y}(x, y, 0) = \sqrt{2} E_0 / w_0 L_n^1 \left(\frac{2\rho^2}{w_0^2} \right) \exp \left(-\frac{\rho^2}{w_0^2} \right) y, \quad (5.3)$$

\vec{e}_x, \vec{e}_y are the unit vectors in the x and y directions, respectively; E_0 is an arbitrary constant amplitude, w_0 is the $1/e$ intensity beam radius of the Gaussian term, $\rho = (x^2 + y^2)^{1/2}$ is the radial coordinate, and $L_n^1(\cdot)$ is the Laguerre polynomial of the radial mode number n , angular mode number 1 .

The RPB can be considered as a combining doughnut beam that is generated by coherent addition of the $\tilde{E}_{n1x}(x, y, 0), \tilde{E}_{n1y}(x, y, 0)$ beams. By applying the vector angular spectrum method, the propagating electric field of the RPB can be obtained:

$$\begin{aligned} \tilde{E}_{n1}(\vec{r}, z) = & \int \int_{-\infty}^{\infty} \left\{ \begin{aligned} & \tilde{E}_{n1x}(p_x, p_y) \vec{e}_x + \tilde{E}_{n1y}(p_x, p_y) \vec{e}_y \\ & - \vec{e}_z / \gamma_1 \left[p_x \tilde{E}_{n1x}(p_x, p_y) - p_y \tilde{E}_{n1y}(p_x, p_y) \right] \end{aligned} \right\} \\ & \exp(ikm_1) dp_x dp_y, \end{aligned} \quad (5.4)$$

where $\vec{r} = x\vec{e}_x + y\vec{e}_y$ denotes the transverse displacement vector, p_x, p_y , and $\gamma_1 = \sqrt{1 - p_x^2 - p_y^2}$ are the direction cosines, $m_1 = p_x x + p_y y + \gamma_1 z$; and $k = 2\pi/\lambda$ is the wavenumber; λ is the wavelength in the medium; \vec{e}_z is the unit vector in the z direction, $\tilde{E}_{n1x}(p_x, p_y), \tilde{E}_{n1y}(p_x, p_y)$ denote the x and y components of the vector angular spectrum, respectively, and can be obtained by Fourier transformation of the x and y components of the initial electric field,

$$\begin{aligned} \tilde{E}_{n1x}(p_x, p_y) &= \frac{1}{\lambda^2} \int \int_{-\infty}^{\infty} E_{n1x}(x, y, 0) \exp(-ikv) dx dy \\ &= A p_x L_n^1 \left(\frac{b^2}{2\sigma^2} \right) \exp \left(-\frac{b^2}{4\sigma^2} \right), \end{aligned} \quad (5.5)$$

$$\begin{aligned} \tilde{E}_{n1y}(p_x, p_y) &= \frac{1}{\lambda^2} \int \int_{-\infty}^{\infty} E_{n1y}(x, y, 0) \exp(-ikv) dx dy \\ &= A p_y L_n^1 \left(\frac{b^2}{2\sigma^2} \right) \exp \left(-\frac{b^2}{4\sigma^2} \right), \end{aligned} \quad (5.6)$$

where $\sigma = 1/kw_0 = \alpha_1/2$ denotes a dimensionless perturbation parameter, α_1 is the far-field diffraction angle, $v = p_x x + p_y y$, $A = (-1)^{n+1} 2^{1/2} i E_0 / (8\pi\sigma^3)$, and $b^2 = p_x^2 + p_y^2$. Because the Maxwell's equations can be split into transverse and longitudinal field equations and from the vectorial structure of electromagnetic beams [60–63], the propagating electric field can be written as TE-mode and TM-mode waves:

$$\vec{E}(\vec{r}, z) = \vec{E}_{TE}(\vec{r}, z) + \vec{E}_{TM}(\vec{r}, z), \quad (5.7)$$

where

$$\begin{aligned} \vec{E}_{TE}(\vec{r}, z) = & \int \int_{-\infty}^{\infty} \frac{1}{b^2} \left[p_y \tilde{E}_x(p_x, p_y) - p_x \tilde{E}_y(p_x, p_y) \right] \\ & (p_y \vec{e}_x - p_x \vec{e}_y) \exp(ikm_1) dp_x dp_y, \end{aligned} \quad (5.8)$$

$$\begin{aligned} \vec{E}_{TM}(\vec{r}, z) = & \int \int_{-\infty}^{\infty} \frac{1}{b^2} \left[p_x \tilde{E}_x(p_x, p_y) + p_y \tilde{E}_y(p_x, p_y) \right] \\ & (p_x \vec{e}_x + p_y \vec{e}_y - b^2 / \gamma_1 \vec{e}_z) \exp(ikm_1) dp_x dp_y. \end{aligned} \quad (5.9)$$

Eq. (5.7) demonstrates that the divergence condition of the electric field is appropriate and the polarized direction of every wave component is perpendicular to its own wave vector. From Maxwell equations and Eq. (5.4), the propagating magnetic field can be obtained:

$$\vec{H}(\vec{r}, z) = \vec{H}_{TE}(\vec{r}, z) + \vec{H}_{TM}(\vec{r}, z), \quad (5.10)$$

where

$$\begin{aligned} \vec{H}_{TE}(\vec{r}, z) = & \sqrt{\frac{\varepsilon}{\mu}} \int \int_{-\infty}^{\infty} \frac{1}{b^2} \left[p_y \tilde{E}_x(p_x, p_y) - p_x \tilde{E}_y(p_x, p_y) \right] \\ & (p_x \gamma_1 \vec{e}_x + p_y \gamma_1 \vec{e}_y - b^2 \vec{e}_z) \exp(ikm_1) dp_x dp_y, \end{aligned} \quad (5.11)$$

$$\begin{aligned} \vec{H}_{TM}(\vec{r}, z) = & -\sqrt{\frac{\varepsilon}{\mu}} \int \int_{-\infty}^{\infty} \frac{1}{b^2 \gamma_1} \left[p_x \tilde{E}_x(p_x, p_y) + p_y \tilde{E}_y(p_x, p_y) \right] \\ & (p_y \vec{e}_x - p_x \vec{e}_y) \exp(ikm_1) dp_x dp_y. \end{aligned} \quad (5.12)$$

Eqs. (5.7)-(5.12) are general when the angular spectra denotes those of an arbitrary field. It is easy to find from Eqs. (5.8) and (5.12) that the TE electric field, and similarly the TM magnetic field, has no z component. Therefore, whenever $p_y \tilde{E}_x(p_x, p_y) - p_x \tilde{E}_y(p_x, p_y) = 0$, the TE electric field is zero, with a similar result for the TE magnetic field, i.e., $\vec{E}_{TE}(\vec{r}, z) = 0$, and $\vec{H}_{TE}(\vec{r}, z) = 0$. The field of the RPBs has this specific property, and one can derive

$$\vec{E}_{nl}(\vec{r}, z) = \vec{E}_{TM}(\vec{r}, z), \quad (5.13)$$

$$\vec{H}_{nl}(\vec{r}, z) = \vec{H}_{TM}(\vec{r}, z). \quad (5.14)$$

One can note that since the azimuthally polarized electric field has no longitudinal electric-field component, probably it is entirely a TE field by these definitions. Eqs. (5.13) and (5.14) show that RPBs are comprised of only TM-mode waves. From Eqs. (5.2), (5.3), (5.13), and (5.14), and Ref. [15], the analytical electromagnetic fields of the TM mode are found to be

$$\vec{E}_{TM}(\vec{r}, z) = \frac{(-1)^{n+1} E_0 \sqrt{2} z_R}{r_1^2 (1 - i z_R / r_1)^2} \exp[ikr_1 - c(r_1)] \left\{ \begin{aligned} & \frac{z}{2\sigma r_1} P(r_1) L_n^1[Q(r_1)] (x\vec{e}_x + y\vec{e}_y - \rho^2/z\vec{e}_z) \\ & + i w_0 \vec{e}_z \sum_{m=0}^n \frac{(n+1)!}{(n-m)!m!} \frac{(-1)^{m-n} E_0 \sqrt{2} z_R}{(1 - i z_R / r_1)^m} L_{m+1}^1[c(r_1)] \end{aligned} \right\}, \quad (5.15)$$

where $r_1 = (x^2 + y^2 + z^2)^{1/2}$, $\vec{H}_{TM}(\vec{r}, z)$ can also be obtained, where $z_R = kw_0^2/2$ is the Rayleigh distance, $L_{m+1}^1(\cdot)$ is the $(m+1)^{\text{th}}$ -order Laguerre polynomial, $c(r_1) = s(r_1) / [2(1 - iz_R/r_1)]$, $s(r_1) = \rho^2 / (2r_1^2 \sigma^2)$, $P(r_1) = [(1 + iz_R/r_1) / (1 - iz_R/r_1)]^n$, and $Q(r_1) = s(r_1) / (1 + z_R^2/r_1^2)$. Eq. (5.15) indicates that the polarization of every light ray will keep TM polarized, i.e., a radially polarized beam propagates as a purely radial polarization without azimuthal components through the whole space. Eq. (5.15) demonstrates the analytical expression for the electromagnetic fields of the TM-mode nonparaxial RPB in a homogeneous medium of a refractive index for the case of $z \gg w_0$. The analytical TM term derived is appropriate to the paraxial and the nonparaxial cases. From Eqs. (5.2), (5.3), (5.13), and (5.14), and Ref. [15], the exact on-axis electromagnetic fields of the TM mode can be derived:

$$\vec{E}_{TM}(0, 0, z) = \frac{\sqrt{2} E_0}{w_0} \vec{e}_z \sum_{m=0}^n \frac{\sqrt{\pi} (n+1)! (2u)^m}{(m+1)! (n-m)! m!} \frac{\partial^{m+1}}{\partial u^{m+1}} \left[\sqrt{u} \exp\left(-\frac{k^2}{4u} + z^2 u\right) \text{erfc}\left(z\sqrt{u} - \frac{ik}{2\sqrt{u}}\right) \right], \quad (5.16)$$

$$\vec{H}_{TM}(0, 0, z) = 0, \quad (5.17)$$

where $u = 1/w_0^2$, $\text{erfc}(\cdot)$ denotes the complementary error function [64]. Eqs. (5.16) and (5.17) indicate that the on-axis electric field of the RPBs only has the longitudinal component without the radial component, and the magnetic field of the RPBs has no on-axis component. For the on-axis electric field, as the polarization of every TM-mode

radially polarized light ray only rotates in the plane formed by the wave vector and z axis when it propagates through a homogeneous medium along the z axis, the polarization of every light ray will keep TM polarized, i.e., a RPB propagates as a strictly longitudinal component without azimuthal and radial components.

5.2.2. Nonparaxial Elegant Hermite-Laguerre Gaussian Beams

In free space, the propagation of a general monochromatic scalar light field $U(x,y,z)$ satisfies the homogeneous Helmholtz equation for $z > 0$:

$$\frac{\partial^2 U}{\partial x^2} + \frac{\partial^2 U}{\partial y^2} + \frac{\partial^2 U}{\partial z^2} + k^2 U = 0. \quad (5.18)$$

Let $U(x,y,z) = \Phi(x,y,z)\exp(ikz)$, Eq. (5.18) can be transformed into

$$\frac{\partial^2 \Phi}{\partial x^2} + \frac{\partial^2 \Phi}{\partial y^2} + 2ik \frac{\partial \Phi}{\partial z} = 0. \quad (5.19)$$

Eq. (5.2) has solutions which is a family of paraxial laser beams

$$\begin{aligned} \Phi_{n,m}(x,y,z|\theta) &= T_{nm} \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \theta \right) U_0(x,y,z) = [-C(z)]^{n+m} U_0(x,y,z) \\ &\times \sum_{s=0}^{n+m} i^s \cos^{n-s} \theta \sin^{m-s} \theta P_s^{(n-s,m-s)}(-\cos 2\theta) H_{n+m-s}[C(z)x] H_s[C(z)y], \end{aligned} \quad (5.20)$$

where $T_{nm}(x,y,\theta) = (x \cos \theta + iy \sin \theta)^n (x \sin \theta - iy \cos \theta)^m$, $C(z) = q(z)/w_0$, $q(z) = (1+iz/z_R)^{-1/2}$, $U_0(x,y,z) = q^2(z)\exp(-C^2(z)\rho^2)$ denotes the ordinary Gaussian solution, $P_s(u,v)(\cdot)$ [64] are the Jacobi polynomials, and $H_n(\cdot)$ is the Hermite polynomial.

When $\theta = 0$ and $\pi/4$, Eq. (5.20) can be transformed into elegant Hermite–Gaussian beams (EHGBs) and elegant Laguerre–Gaussian beams (ELGBs). As $m = n = 0$, Eq. (5.20) can be simplified to a Gaussian solution. So, EHLGBs are the more general solutions containing Gaussian solution, EHGBs, ELGBs, and intermediate EHLGBs.

The beam is yielded by a higher-order point source of strength S_{cs} situated at $x = y = 0$ and $z = z_{cs}$. Proper choice of S_{cs} and z_{cs} generates the desired beam.

The wave function satisfies the inhomogeneous Helmholtz equation

$$\frac{\partial^2 U}{\partial x^2} + \frac{\partial^2 U}{\partial y^2} + \frac{\partial^2 U}{\partial z^2} + k^2 U = -S_{cs} T_{nm} \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \theta \right) \delta(x) \delta(y) \delta(z - z_{cs}), \quad (5.21)$$

$$U_{n,m}(x, y, z) = \int \int_{-\infty}^{\infty} \tilde{U}_{nm}(p_x, p_y, z) \exp(-ikv) dp_x dp_y, \quad (5.22)$$

$$\tilde{U}_{nm}(p_x, p_y, z) = \frac{1}{\lambda^2} \int \int_{-\infty}^{\infty} U_{n,m}(x, y, z) \exp(ikv) dx dy. \quad (5.23)$$

From Eq. (5.21), $\tilde{U}_{nm}(p_x, p_y, z)$ is decided and made the substitution of Eq. (5.7), one can obtain

$$U_{n,m}(x, y, z) = \frac{i}{2} S_{cs} (-i2\pi)^{n+m} \int \int_{-\infty}^{\infty} T_{nm}(p_x, p_y, \theta) \exp[i\zeta(z - z_{cs})] / \zeta \exp(-ikv) dp_x dp_y, \quad (5.24)$$

where $\text{Re}(z - z_{cs}) > 0$, $\zeta = (1 - \lambda^2 b^2)^{1/2}$. One can expand ζ into a series and keeping the first and the second terms, we obtain $\zeta = 1 - \frac{1}{2} \lambda^2 b^2$. Under the paraxial approximation, i.e., $\lambda^2 b^2 \ll 1$, we substitute ζ of the exponential part in Eq. (5.24) with the approximation and the other terms with k , Eq. (5.24) reduces to

$$U_{n,m,p}(x, y, z) = \frac{i}{2k} \exp[ik(z - z_{cs})] S_{cs} (-i2\pi)^{n+m} \int \int_{-\infty}^{\infty} \exp(-ikv) T_{nm}(p_x, p_y, \theta) \exp[-i\lambda\pi b^2(z - z_{cs})] dp_x dp_y. \quad (5.25)$$

The additional subscript p in $U_{n,m,p}$ means the paraxial approximation. Calculating the integrals in Eq. (5.25), one can obtain

$$U_{n,m,p}(x, y, z) = \exp(ikz) S_{cs} T_{nm} \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \theta \right) \frac{\exp(-ikz_{cs})}{4\pi(z - z_{cs})} \exp \left[\frac{ik(x^2 + y^2)}{2(z - z_{cs})} \right]. \quad (5.26)$$

To yield the EHLGBs for $z > 0$, we suppose, for the case of the paraxial approximation, the input distribution is provided by Eq. (5.20) for $z = 0$. Parameters S_{cs} and z_{cs} are decided by comparing Eq. (5.26) for $z = 0$ with Eq. (5.20) for $z = 0$; and one can obtain

$$z_{cs} = iz_R, \quad (5.27)$$

$$S_{cs} = -4i\pi z_R w_0^{n+m} \exp(-kz_R). \quad (5.28)$$

Substitution of Eqs. (5.27) and (5.28) into Eq. (5.24), the exact integral expression for U can be obtained

$$U_{n,m}(x, y, z) = 2\pi z_R (-i2\pi w_0)^{n+m} \exp(-kz_R) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} T_{nm}(p_x, p_y, \theta) \exp[i\zeta(z - iz_R)] / \zeta \exp(-ikv) dp_x dp_y. \quad (5.29)$$

From Eq. (5.27), it is easy to find that the source in Eq. (5.21) lies external to $z > 0$, and the solution shown in Eq. (5.29) is the exact solution to the homogeneous Helmholtz Eq. (5.18). The exact solution simplifies to the correct paraxial approximation for the case of excluding all the nonparaxial contributions and the evanescent waves. When $\theta = 0$ and $\pi/4$, Eq. (5.29) simplifies to the special cases obtained by Seshadri [13] and Bandres [6].

Starting from the Green-function approach, we can decide the differential representation of the elegant Hermite–Laguerre–Gaussian (EHLG) waves.

The solution of the differential equation

$$\frac{\partial^2 G}{\partial x^2} + \frac{\partial^2 G}{\partial y^2} + \frac{\partial^2 G}{\partial z^2} + k^2 G = -S_{cs} \delta(x) \delta(y) \delta(z - z_{cs}) \quad (5.30)$$

is expressed in terms of

$$G = S_{cs} \exp(ikR_1) / (4\pi R_1), \quad (5.31)$$

where $R_1 = [x^2 + y^2 + (z - iz_R)^2]^{1/2}$.

Making a comparing between Eq. (5.30) and Eq. (5.21) and applying Eq. (5.28), one can find the differential or multipole representation of the EHLG waves

$$U_{n,m}(x, y, z) = -iz_R w_0^{n+m} \exp(-kz_R) T_{nm} \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \theta \right) \frac{\exp(ikR_1)}{R_1}. \quad (5.32)$$

According to Eq. (5.32), it is easy to find that higher-order EHLG modes can be obtained by making the operator $T_{nm} \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \theta \right)$ to the complex-source-point spherical wave $S_{cs} \exp(ikR_1) / (4\pi R_1)$ of the paraxial Gaussian beam.

By expanding the ς and $\exp[i \varsigma (z-z_{cs})]$ in Eq. (5.24), one can obtain the nonparaxial corrections of EHLG beams. The product of both series terms up to order $(kw_0)^{-2j}$ are remained, and the j^{th} order corrections can be obtained. One can find that the nonparaxial solution approaches the exact solution Eq. (5.24) with the increase of the parameter j . Especially, when j approaches infinity, the nonparaxial solution becomes the exact solution. For simplicity, we obtain the first three nonparaxial corrections of EHLG beams for $j = 3$; Eq. (5.24) can be written in the form

$$U_{n,m}(x, y, z) = \pi w_0^{n+m+2} \exp(ikz) \int_{-\infty}^{\infty} T_{nm}(p_x, p_y, \theta) \exp[-i\lambda\pi b^2(z-z_{cs})] \exp(-ikv) \sum_{j=0}^3 \frac{\tilde{G}^{2j}(\cdot)(4\pi^2 b^2 w_0^2, z)}{(kw_0)^{2j}} dp_x dp_y, \quad (5.33)$$

where $\tilde{G}^{(0)}(t_0, z) = a_{00}$, $\tilde{G}^{(2)}(t_0, z) = a_{21}t_0 - a_{22}t_0^2/q^2(z)$,

$$\tilde{G}^{(4)}(t_0, z) = a_{42}t_0^2 - a_{43}t_0^3/q^2(z) + a_{44}t_0^4/q^4(z),$$

$$\tilde{G}^{(6)}(t_0, z) = a_{63}t_0^3 - a_{64}t_0^4/q^2(z) + a_{65}t_0^5/q^4(z) - a_{66}t_0^6/q^6(z),$$

$a_{00} = 1$, $a_{21} = 1/2$, $a_{22} = 1/16$, $a_{42} = 3/8$, $a_{43} = 1/16$, $a_{44} = 1/512$, $a_{63} = 5/16$, $a_{64} = 5/128$, $a_{65} = 3/1024$, $a_{66} = 1/(4096 \cdot 6)$. From Eq. (5.33), one can obtain

$$U_{n,m}(x, y, z | \theta) = \exp(ikz) [-C(z)]^{n+m} U_0(x, y, z) \sum_{s=0}^{n+m} i^s \cos^{n-s} \theta \sin^{m-s} \theta P_s^{(n-s, m-s)}(-\cos 2\theta) \times \sum_{j=0}^{n+m} (kw_0)^{-2j} (-1)^j q^{2j}(z) f_{n,m}^{(2j)}(x, y, z), \quad (5.34)$$

$$\text{where } f_{n,m}^{(2j)}(x, y, z) = \sum_{t_1=j}^{2j} a_{2j, t_1} \sum_{l=0}^{t_1} \frac{t_1!}{l!(t_1-l)!} H_{n+m-s+2(t_1-l)}[C(z)x] H_{s+2l}[C(z)y].$$

Eq. (5.33) denotes the nonparaxial expression of EHLGBs up to the first three nonparaxial corrections.

5.2.3. Nonparaxial Pearcey Gaussian Beams

The wave function of the Pearcey Gaussian beams is described by the inhomogeneous Helmholtz equation

$$\frac{\partial^2 U}{\partial x^2} + \frac{\partial^2 U}{\partial y^2} + \frac{\partial^2 U}{\partial z^2} + k^2 U = -S_{cs}^{(2)} \exp[-D(x, y)] \delta(x) \delta(y) \delta(z - z_{cs}^{(2)}), \quad (5.35)$$

where $S_{cs}^{(2)} = -8i\pi^2 k w_0^2 \exp(4kc_0 z_R)$ and $D(x, y) = w_0^4 \left(\frac{\partial^4}{\partial x^4} + \frac{\partial^4}{\partial y^4} \right)$, $z_{cs}^{(2)} = -4ic_0 z_R$.

Applying Eq. (5.13), we can derive a beam whose input field distribution in the $z = 0$ plane reads

$$U(x, y, 0) = P e\left(\frac{x}{w_0}, c_0\right) P e\left(\frac{y}{w_0}, c_0\right). \quad (5.36)$$

Making use of the Green function approach, it is easy to see the connection between the (1+ 2)-dimensional Pearcey wave function and the complex source-point spherical wave function. The differential representation of the Pearcey wave can be governed by Eq. (5.30), one solution of Eq. (5.30) can be written

$$G = S_{cs} \exp(ikR_1) / (4\pi R_1), \quad (5.37)$$

where $R_1 = [x^2 + y^2 + (z + 4ic_0 z_R)^2]^{1/2}$.

Comparing Eq. (5.35) and Eq. (5.30), it is straightforward to find the relationship between the two waves as

$$U(x, y, z) = -2i\pi k w_0^2 \exp(4kc_0 z_R) \exp[-D(x, y)] \frac{\exp(ikR_1)}{R_1}. \quad (5.38)$$

Applying the operator $S_{cs}^{(2)} \exp[-D(x, y)]$ to the complex-source-point spherical wave $\exp(ikR_1)/R_1$ relative to the paraxial Gaussian beam, the Pearcey modes can be generated, as is found in Eq. (5.38).

Using the same method as that in the 1.3.2, we can obtain the first three terms in the nonparaxial correction series of the (1+ 2)-dimensional PB for $j = 3$ as

$$\begin{aligned}
U(x, y, z) = & w_0^2 / \sqrt{W_0} \exp(4kc_0 z_R) \\
& \left[1 - \frac{1}{2k^2} \nabla_t^2 + \frac{3}{8k^4} \nabla_t^4 - \frac{(5/k + iz')}{16k^5} \nabla_t^6 - \frac{iz'}{4k^3} \left(1 - \frac{1}{2k^2} \nabla_t^2 \right) \right] \\
& \left[\frac{\partial^4}{\partial x^2 \partial y^2} - \frac{z'^2}{32k^6} \frac{\partial^8}{\partial x^4 \partial y^4} \right] \\
& Pe \left(\frac{x}{\frac{1}{W_0^4}}, \frac{-z'}{2kW_0^2} \right) Pe \left(\frac{y}{\frac{1}{W_0^4}}, \frac{-z'}{2kW_0^2} \right),
\end{aligned} \tag{5.39}$$

where $z' = z + 4ic_0 z_R$, $W_0 = w_0^4 - z'/(8k^3)$, $\nabla_t^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$ is the transverse Laplacian operator.

5.2.4. Nonparaxial Parabolic Rotational Coordinate Beams

In Eq. (5.18), one can establish the parabolic rotational coordinates by introducing $x = \xi \eta \cos \varphi$, $y = \xi \eta \sin \varphi$, $z_0 - z = \frac{\xi^2 - \eta^2}{2}$ [64, 65], z_0 is a positive constant, $\xi \in (0, \infty)$, $\eta \in (0, \infty)$, $\varphi \in (0, 2\pi)$ and supposing $U(\xi, \eta, \varphi) = M(\xi)N(\eta)\Psi(\varphi)$, the Helmholtz Eq. (5.18) can be separated into the following differential equations

$$\frac{d^2 M(\xi)}{d\xi^2} + \frac{1}{\xi} \frac{dM(\xi)}{d\xi} + \left(-\frac{m^2}{\xi^2} + k^2 \xi^2 - \alpha k \right) M(\xi) = 0, \tag{5.40a}$$

$$\frac{d^2 N(\eta)}{d\eta^2} + \frac{1}{\eta} \frac{dN(\eta)}{d\eta} + \left(-\frac{m^2}{\eta^2} + k^2 \eta^2 + \alpha k \right) N(\eta) = 0, \tag{5.40b}$$

$$\frac{d^2 \Psi(\varphi)}{d\varphi^2} + m^2 \Psi(\varphi) = 0, \tag{5.40c}$$

where m and α are the separation constants. Solutions of Eqs. (5.40a) and (5.40 b) are given by hypergeometric functions [64]. Collecting the solutions of Eqs. (5.40 a)-(5.40 c), we obtain the exact solution of Eq. (5.18) as

$$\begin{aligned}
U(\xi, \eta, \varphi) = & \left(\frac{\xi \eta}{w_0} \right)^m {}_1F_1 \left(\frac{i\alpha}{4} + \frac{m+1}{2}, m+1, \mp ik \xi^2 \right) \\
& {}_1F_1 \left(-\frac{i\alpha}{4} + \frac{m+1}{2}, m+1, \mp ik \eta^2 \right) e^{\pm i(\xi^2 + \eta^2)/2 + im\varphi}.
\end{aligned} \tag{5.41}$$

From Eq. (5.41), the nonparaxial parabolic rotational coordinate beams (NPRCBs) as exact solutions of Eq. (5.18) in the laboratory frame can be expressed as

$$U(x, y, z) = (-1)^m \left(\frac{x \pm iy}{w_0} \right)^m {}_1F_1 \left(\frac{i\alpha}{4} + \frac{m+1}{2}, m+1, \mp ik(z_0 - z + r_1) \right) {}_1F_1 \left(-\frac{i\alpha}{4} + \frac{m+1}{2}, m+1, \mp ik(-z_0 + z + r_1) \right) e^{\pm ikr_1}, \quad (5.42)$$

where w_0 is a positive parameter with length dimension, $r_1 = \sqrt{x^2 + y^2 + (z_0 - z)^2}$, and m is related to the azimuthal eigenvalue as referred to the topological charge for optical vortices. At $i\alpha = 0$ and $2(m+1)$, Eq. (5.42) becomes the Bessel wave [65] and the vortex plane wave. When $i\alpha$ is an integer and $i\alpha < -2(m+1)$, Eq. (5.42) describes the Laguerre wave [65]. When $m = 0$, Eq. (5.42) denotes the fundamental solution of a nonparaxial parabolic rotational coordinate beam. We point out that the divergence angle of the beam is decided by the parameters in Eq. (5.42). Under paraxial approximation, Eq. (5.42) reduces to the solution of paraxial beams in parabolic rotational coordinates.

It is easy to deduce from Eq. (5.40) that the NPRCBs are scalable. Assuming the transverse coordinates are normalized by a constant w_0 , the NPRCBs in the normalized coordinate system can be decided by Eq. (5.42) with the solutions recasted by normalizing k with $1/w_0$. Thus, in experiment, one can control the beam focusing and propagation properties simply by squeezing or expanding the transverse coordinates. In our experiment, the NPRCBs are realized by employing the computer generated holography as reported in the previous work [66, 67].

Our theoretical and experimental demonstrations of the zeroth-order NPRCB are summarized in Fig. 5.1. The iso-intensity contour and the side view propagation as shown in Figs. 5.1(a) and 5.1(b) suggest that the evolution of the beam width follows parabolic curves. This is further confirmed by directly calculating the beam width at different propagation distance, as shown in Fig. 5.1(h). The transverse intensity patterns in Figs. 5.1(c)-5.1(g) clearly indicate that the central intensity experiences periodic oscillations during the propagation, which can be also seen in the Figs. 5.1(a), 5.1(b), and 5.1(h). In addition, from Fig. 5.1(h), one can see that the zeroth-order NPRCB diffracts much faster than the Bessel beam does, but much slower than the Gaussian beam does.

5.3. Paraxial Beams in Nonlocal Nonlinear Media

Optical spatial solitons have drawn much attention in the subject of many theoretical [41, 44, 46, 52-57, 68-75] and experimental [42, 43, 45-51, 76-81] demonstration in strongly nonlocal nonlinear medium, mainly because of their novelty for fundamental research and potential applications [41-43]. Optical spatial solitons are self-trapped optical beams when diffraction is balanced by nonlinearity. Many kinds of nonlocal spatial solitons, such as Hermite-Gaussian solitons [52], Laguerre-Gaussian solitons [53], dipole solitons [82], vortex solitons [83], azimuthons [84], ellipticons [85] and rotating solitons

[86, 87] have been demonstrated. In this chapter, we introduce Ince–Gaussian breathers and solitons [54, 55] and the complex variable function (CVF)-Gaussian breathers and solitons [56, 57].

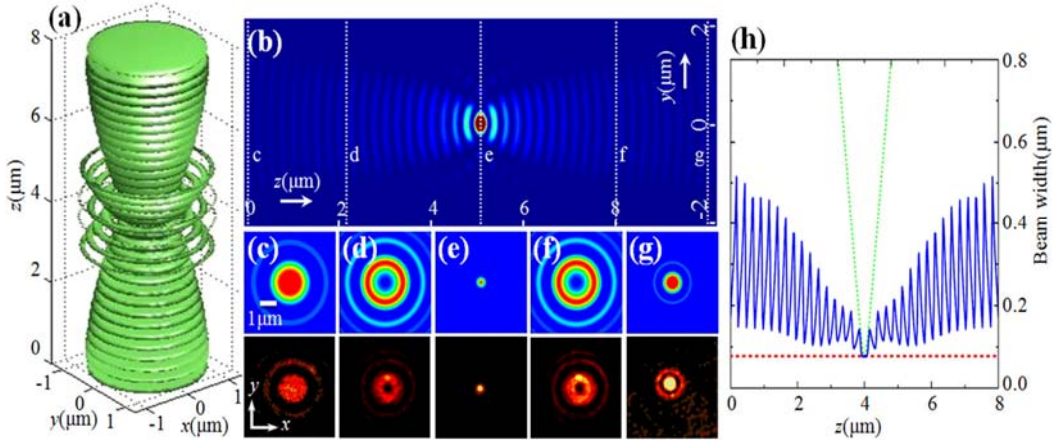


Fig. 5.1. (a) Iso-intensity contour of the exact (00)-mode NPRCBs; (b) Numerical simulated side-view propagation of the exact (00)-mode of NPRCBs along z ; Intensity snapshots from (c)-(g) are corresponding to theoretical (top) and experimental (bottom) results at different longitudinal positions as marked in (a); (h) The beam width of the main lobe of the beam versus the propagation distance z for the (00)-mode NPRCB (blue solid line), the zero-order Bessel beam (red dashed line), and the Gaussian beam (the green dotted line) with the same main lobe size at $z = 0$.

The evolution of the (1+2)-dimensional optical beams in the nonlocal cubic nonlinear media can be described by the NNLSE

$$i \frac{\partial \Phi}{\partial z} + \mu \left(\frac{\partial^2 \Phi}{\partial x^2} + \frac{\partial^2 \Phi}{\partial y^2} \right) + k \frac{\Delta n}{n_0} \Phi = 0, \quad (5.43)$$

where $\Phi = \Phi(x, y, z)$ denotes a paraxial beam, z means the longitudinal coordinate, $\mu = 1/(2k)$, n_0 is the linear refractive index of the media, $\Delta n = n_2 \int R(\vec{r} - \vec{r}') |\Phi(\vec{r}', z)|^2 d^2 \vec{r}'$ is the nonlinear perturbation of the refraction index, n_2 is the nonlinear index coefficient, \vec{r} and \vec{r}' means the two-dimensional transverse coordinate vectors, and R signifies the normalized symmetrical real spatial response function of the media.

For the case of the strong nonlocality, Eq. (5.1) is reduced into the SMM [41],

$$i \frac{\partial \Phi}{\partial z} + \mu \left(\frac{\partial^2 \Phi}{\partial x^2} + \frac{\partial^2 \Phi}{\partial y^2} \right) - \frac{n_2}{n_0} k \mathcal{P}_0 r^2 \Phi = 0, \quad (5.44)$$

where $r = |\vec{r}|$, γ is the material parameter corresponding to R, and P_0 is the input power at $z = 0$.

5.3.1. Ince-Gaussian Breathers and Solitons

Letting that a solution of Eq. (5.44) is a multiplication of two functions $\Psi_F(r, z)$ and $\Psi_G(r, z)$,

$$\Phi = \Psi_F(\vec{r}, z) \Psi_G(\vec{r}, z). \quad (5.45)$$

Substitution of Eq. (5.45) into Eq. (5.44), and applying,

$$i \frac{\partial \Psi_G}{\partial z} + \mu \left(\frac{\partial^2 \Psi_G}{\partial x^2} + \frac{\partial^2 \Psi_G}{\partial y^2} \right) - \frac{n_2}{n_0} k \gamma P_0 r^2 \Psi_G = 0, \quad (5.46)$$

one can obtain

$$i \Psi_G \frac{\partial \Psi_F}{\partial z} + \mu \Psi_G \left(\frac{\partial^2 \Psi_F}{\partial x^2} + \frac{\partial^2 \Psi_F}{\partial y^2} \right) + 2\mu \left(\frac{\partial \Psi_{\Psi_G}}{\partial x} \frac{\partial \Psi_F}{\partial x} + \frac{\partial \Psi_{\Psi_G}}{\partial y} \frac{\partial \Psi_F}{\partial y} \right) = 0. \quad (5.47)$$

Eq. (5.46) has the Gaussian solution

$$\Psi_G = \frac{\sqrt{P_0} \exp[i\theta(z)]}{\sqrt{\pi} w(z)} \exp \left[-\frac{r^2}{w^2(z)} + i c(z) r^2 \right], \quad (5.48)$$

where $w(z)$ represents the beam width of the Gaussian beam, $c(z)$ is the phase front curvature of the beam, $\theta(z)$ denotes the phase of the complex amplitude. They can be expressed in the forms [70, 71], respectively

$$w^2(z) = w_0^2 \left(\cos^2 \beta_0 z + \frac{P_c}{P_0} \sin^2 \beta_0 z \right), \quad (5.49)$$

$$c(z) = \frac{k \beta_0 \left(\frac{P_c}{P_0} - 1 \right) \sin 2\beta_0 z}{4 \left(\cos^2 \beta_0 z + \frac{P_c}{P_0} \sin^2 \beta_0 z \right)}, \quad (5.50)$$

$$\theta(z) = -\arctan \left(\sqrt{\frac{P_c}{P_0}} \tan \beta_0 z \right), \quad (5.51)$$

and

$$P_c = \frac{1}{k^2 m_2 w_0^4}, \quad (5.52)$$

where $\beta_0 = \sqrt{n_2 \gamma P / n_0}$, and P_c denotes the critical power for the soliton propagation. Eq. (5.49) presents that the z -dependent function $w(z)$, which is the beam width of the Gaussian function, oscillates periodically along the propagation z when $P_0 \neq P_c$. Importing the elliptic coordinate transformation in the transverse plane perpendicular to z , where the elliptic coordinate [3, 88–91] is defined as $x = f(z) \cosh \xi \cos \zeta$, $y = f(z) \sinh \xi \sin \zeta$ and $z = z$, $\xi \geq 0$ and $0 \leq \zeta < 2\pi$ are the radial and angular elliptic variables, and semifocal separation $f(z) = f_0 w(z)/w_0$ has the dimension of length, where f_0 denotes the semifocal separation at waist plane $z = 0$. Letting $\Psi_F = F(\xi)N(\zeta) \exp[i\Theta(z)]$, where F , N and Θ signify real functions. Substituting Ψ_F and Eq. (5.48) into Eq. (5.47), one can derive

$$\frac{d^2 F}{d\xi^2} - \varepsilon_1 \sinh(2\xi) \frac{dF}{d\xi} - (a - p\varepsilon_1 \cosh 2\xi) F = 0, \quad (5.53)$$

$$\frac{d^2 N}{d\zeta^2} + \varepsilon_1 \sin(2\zeta) \frac{dN}{d\zeta} + (a - p\varepsilon_1 \cos 2\zeta) N = 0, \quad (5.54)$$

$$-\frac{w^2(z)}{2\mu} \frac{d\Theta}{dz} = p, \quad (5.55)$$

where p and a denote separation constants, and $\varepsilon_1 = f_0^2 / w_0^2$ signifies the ellipticity parameter. The ellipticity ε_1 , the waist spot w_0 , and the semifocal separation f_0 mean the physically important parameters for illustrating the transverse structure of the IG modes. The dimensionless parameter ε_1 adapts to the ellipticity of the mode; the parameters w_0 and f_0 estimate its physical size [90, 91]. The solution of Eq. (5.55) can be given as $\Theta(z) = p\theta(z)$. Eq. (5.54) denotes the Ince equation [3, 89–91], which is a special case of the Hill equation. Let $i\xi = \zeta$, Eq. (5.53) can be changed into Eq. (5.54) and vice versa. The solutions of Eq. (5.54) are honored as the even and odd Ince polynomials of order p and degree m , usually read as C_{mp} and S_{mp} , where $0 \leq m \leq p$ for even function, $1 \leq m \leq p$ for odd function, and indices (p, m) have the same parity, i.e., $(-1)^{p-m} = 1$ [3, 89–91]. To derive three-dimensional solutions, only products of functions of the same parity in ζ and ξ satisfy continuity in the whole space; after adjusting the terms, the general expressions of the even and odd IG solutions read as

$$\Phi_{\text{pm}}^{(\text{e})} = \frac{c_{\text{pm}}^{(\text{e})} C_p^{\text{m}}(i\xi, \varepsilon_1) C_p^{\text{m}}(\zeta, \varepsilon_1) \sqrt{P_0}}{\sqrt{\pi} w(z)} \exp \left[-\frac{r^2}{w^2(z)} + ic(z)r^2 - i(p+1)\theta(z) \right], \quad (5.56)$$

$$\Phi_{\text{pm}}^{(\text{o})} = \frac{s_{\text{pm}}^{(\text{o})} S_p^{\text{m}}(i\xi, \varepsilon_1) S_p^{\text{m}}(\zeta, \varepsilon_1) \sqrt{P_0}}{\sqrt{\pi} w(z)} \exp \left[-\frac{r^2}{w^2(z)} + ic(z)r^2 - i(p+1)\theta(z) \right], \quad (5.57)$$

where $c_{\text{pm}}^{(\text{e})}$ and $s_{\text{pm}}^{(\text{o})}$ are normalization constants which can be determined by

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |\Phi_{\text{pm}}^{(\text{e}, \text{o})}|^2 dS = P_0, \quad dS \text{ denotes the area differential element across the transverse}$$

plane, and the super-indices e and o are even and odd modes, respectively. IG solutions are orthonormal with respect to the indices and the parity at any z-plane, i.e.,

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \Phi_{\text{pm}}^{(\sigma_1)} \Phi_{\text{p'm'}}^{(\sigma_1')} dS = \delta_{\sigma_1 \sigma_1'} \delta_{pp'} \delta_{mm'}; \quad \delta \text{ is the Kronecker delta function and } \sigma = \text{e, o.}$$

To identify the properties of the IG mode discussed above, we have used the split-step Fourier method [94] to decide numerically the three-dimensional field structure of an IG mode within nonlocal nonlinear media. The transverse field is sampled over a grid of 256×256 points. We numerically prove our analytical results by carrying out (2+1)D nonlinear simulation of Eq. (5.43) in the general case of a material with a nonlocal response.

The material response is defined by introducing the Gaussian function [45, 67, 70, 69], i.e., $R(r) = 1/2\pi w_m^2 \exp(-r^2/w_m^2)$, where w_m is the characteristic length of the material response function, and $\alpha = w_0/w_m$ denotes the degree of the material nonlocality. The less α is, the stronger the nonlocality is. The exact analytical IG beams for the SMM with the exact results of numerical simulation of Eq. (5.43) show the same results in the case of SNNM, an input consisting of IG beams at $z = 0$, as is presented in Fig. 5.2. Fig. 5.2 represents that the analytical solutions agree well with the numerical simulations for the case of strong nonlocality. As the degree of nonlocality gets weaker (α becomes bigger), the approximations of the analytical results to the exact ones are a little bit worse, as presented in the Fig. 5.2. When the degree of nonlocality is weak enough, the dynamics of the IG solitons will be unstable.

5.3.2. Complex Variable Function Breathers and Solitons

In Eq. (5.47), importing to a rotating coordinate system (x', y', z) , $x' = w_0(x \cos[\Theta(z)] + y \sin[\Theta(z)]) / w(z)$, $y' = w_0(y \cos[\Theta(z)] - x \sin[\Theta(z)]) / w(z)$, where $\Theta(z)$ depicts the rotation angle of the rotating coordinate, and $d\Theta(z)/dz$ is the angular velocity (AV, the angle rotated per unit propagation distance). In the rotating coordinate system, Eq. (5.47) can be expressed as

$$\frac{d\Theta(z)}{dz} L_z' \Psi_F + \frac{\mu}{w^2(z)} \left(\frac{\partial^2 \Psi_F}{\partial x'^2} + \frac{\partial^2 \Psi_F}{\partial y'^2} \right) - \frac{2\mu}{w^2(z)} \left(x' \frac{\partial \Psi_F}{\partial x'} + y' \frac{\partial \Psi_F}{\partial y'} \right) = 0, \quad (5.58)$$

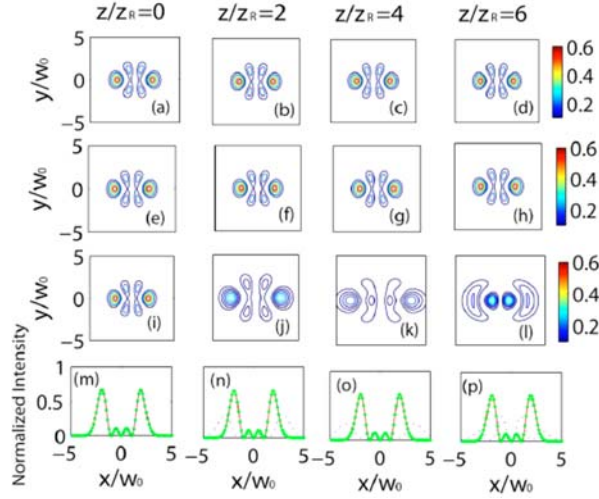


Fig. 5.2. (Color online) Propagation dynamics of the even IG soliton in the Gaussian-shaped response material. (a)-(d) Transverse normalized intensity distributions from the analytical solution, (e)-(h) Distributions from the numerical simulation $\alpha = 0.1$, (i)-(l) distributions from the numerical simulation $\alpha = 0.3$, (m)–(p) normalized intensity distributions at $y/w_0 = 0$. The different columns show the different propagation distances given at the top of the figure. In (m)–(p), the solid, dotted, and dashed curves denote the analytical solutions and the numerical solutions for $a = 0.1$ and $a = 0.3$, respectively; the solid and dotted curves are only a little bit different and cannot be differentiated by one's eyes. The parameters are chosen as $p = 3$, $m = 3$, $P_0/P_c = 1$, $\varepsilon = 2$, and $z_R = kW_0^2$ is the Rayleigh range. This figure comes from Ref. [54].

where $L_z' = -i \left(x' \frac{\partial}{\partial y'} - y' \frac{\partial}{\partial x'} \right)$ is the z component of the angular momentum operator, equivalent to one in quantum mechanics [95]. The solution of Eq. (5.58) can be obtained

$$\Psi_F(x', y', z) = f \left(\frac{x' + iy'}{bw_0} \right), \quad (5.59)$$

$$\Theta(z) = \mp \theta(z), \quad (5.60)$$

where $f(\cdot)$ is an arbitrary analytical function, $b \neq 0$ denotes an arbitrary real parameter whose physical meaning will be discussed later, $\theta(z)$ is given by Eq. (5.51), and \pm presents the rotation direction which accords with the right-hand (left hand) rule corresponding to the propagation direction. From Eqs. (5.45) and (5.59), one can obtain the solution of Eq. (5.46) in the laboratory frame:

$$\Phi(x, y, z) = \frac{c_0 f(\zeta_{\pm})}{w(z)} \exp \left[-\frac{r^2}{w^2(z)} + ic(z)r^2 + i\theta(z) \right], \quad (5.61)$$

where c_0 is the normalization constant which can be obtained by $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |\Phi|^2 ds = P_0$, $\zeta_{\pm} = \frac{x \pm iy}{bw(z)} \exp[\mp i\theta(z)]$.

The solution is named a complex variable function (CVF) Gaussian beam because of the fact that $f(\cdot)$ is an arbitrary analytical complex variable function. The structure of the beam is decided by the product of the function $f(\cdot)$ and a Gaussian function. Solution (5.61) describes an optical beam that rotates with an AV $d\theta(z)/dz$ in the laboratory frame (x, y, z) and is stationary in the rotation frame (x', y', z) . Observed in the laboratory frame, the trajectories of the corresponding points of the beam during propagation rotate generally.

The CVF Gaussian beam can represent different properties for different input power P_0 . When $P_0 < P_c$ [the expression for P_c was given under Eq. (5.52)], the pattern of the CVF Gaussian beam initially broadens and then narrows, experiencing a periodic evolution; whereas the reverse occurs and the pattern initially narrows when $P_0 > P_c$. These two cases are CVF Gaussian breather states. When $P_0 = P_c$, Eq. (5.61) can be simplified to CVF Gaussian solitons.

Fig. 5.3 shows the propagation dynamics of the sine CVF-Gaussian beam. The excellent agreement between the exact analytical sine CVF-Gaussian beam solution of the SMM and the numerical simulation of Eq. (5.43) for the case of strong nonlocality $\alpha = 0.06$ is presented in Fig. 5.3. When the degree of nonlocality becomes weaker $\alpha = 0.2$, the approximations of the analytical results of SMM to the exact ones of Eq. (5.43) are a little bit worse; the sine CVF-Gaussian soliton still exists, but the rotating periodicity turns short. When the degree of nonlocality is weak enough $\alpha = 0.4$, the dynamics of the sine CVF Gaussian beam will be unstable and the rotating periodicity changes shorter.

5.4. Conclusions

In conclusion, the analytical vectorial structure of radially polarized beams has been imported. Our analytical results indicate that the RPBs are composed of only TM-mode waves. The physical pictures of the RPBs are well demonstrated from the vectorial structure. This particular electromagnetic field is entirely transverse magnetic (TM), and on axis it only has a longitudinal (z) electric-field component (i.e., no transverse electric field and no magnetic field at all on axis). The virtual sources that yield the EHLG waves and a Pearcey wave have been introduced. We have obtained the integral and the differential representation for the EHLG waves and the Pearcey wave. From the integral representation of the EHLG waves and the Pearcey wave, we derive the first three orders of nonparaxial corrections for the corresponding paraxial EHLGBs and Pearcey beam. 3D nonparaxial beams of parabolic rotational coordinates have been obtained theoretically by solving the Helmholtz equation and demonstrated experimentally by employing

computer-generated holography. Our experimental results agree well with the theoretical analysis.

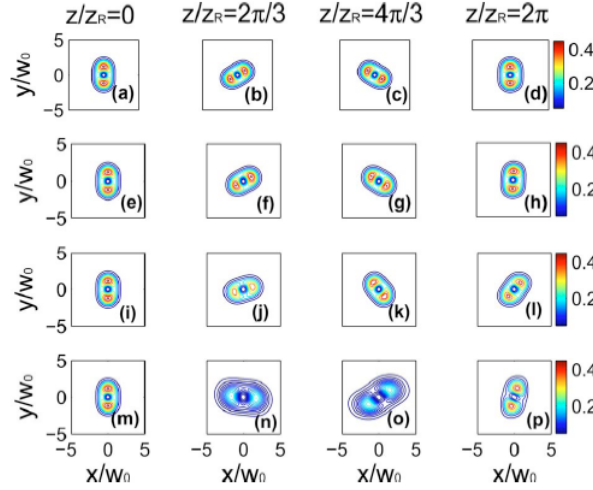


Fig. 5.3. (Color online) Propagation dynamics of the sine CVF-Gaussian beam in the Gaussian-shaped response material. (a)–(d) Transverse normalized intensity distributions from the analytical solution, (e)–(h) distributions from the numerical simulation $\alpha = 0.06$, (i)–(l) distributions from the numerical simulation $\alpha = 0.2$, (m)–(p) distributions from the numerical simulation $\alpha = 0.4$. The different columns represent the different propagation distances given in the top of the figure. The parameters are chosen as $P_0/P_c = 1$ and $b = 1$. This figure comes from Ref. [56].

On the other hand, we have derived spatial IG and CVF-Gaussian breathers and solitons stabilized by strong nonlocal nonlinearity. Their transverse structures of the spatial IG and CVF-Gaussian breathers and solitons are depicted by the Ince polynomials and the CVF. The comparisons of analytical solutions with numerical simulations of the NNLSE show that the analytical IG solutions and CVF-Gaussian solution agree well with the numerical results in the case of strong nonlocality.

Acknowledgements

National Natural Science Foundation of China (NSFC) (11775083, 11374108, 11474109).

References

- [1]. A. E. Siegman, Hermite-Gaussian functions of complex argument as optical-beam eigenfunctions, *J. Opt. Soc. Am. A*, Vol. 63, Issue 9, 1973, pp. 1093-1094.
- [2]. A. E. Siegman, Chapter 16, in *Lasers, University Science*, 1986, pp. 644-652.
- [3]. M. A. Bandres, J. C. Gutiérrez-Vega, Ince-Gaussian beams, *Opt. Lett.*, Vol. 29, Issue 2, 2004, pp. 144-146.

- [4]. M. A. Bandres, Elegant Ince-Gaussian beams, *Opt. Lett.*, Vol. 29, Issue 15, 2004, pp. 1724-1726.
- [5]. T. Takenaka, M. Yokota, O. Fukumitsu, Propagation of light beams beyond the paraxial approximation, *J. Opt. Soc. Am. A*, Vol. 2, Issue 6, 1985, pp. 826-829.
- [6]. M. A. Bandres, J. C. Gutiérrez-Vega, Higher-order complex source for elegant Laguerre-Gaussian waves, *Opt. Lett.*, Vol. 29, Issue 19, 2004, pp. 2213-2215.
- [7]. G. A. Deschamps, Gaussian beam as a bundle of complex rays, *Electron. Lett.*, Vol. 7, Issue 23, 1971, pp. 684-685.
- [8]. L. B. Felsen, Evanescent waves, *J. Opt. Soc. Am.*, Vol. 66, Issue 1, 1976, pp. 751-760.
- [9]. S. Y. Shin, L. B. Felsen, Gaussian beam modes by multipoles with complex source points, *J. Opt. Soc. Am.*, Vol. 67, Issue 5, 1977, pp. 699-700.
- [10]. A. A. Tovar, Production and propagation of cylindrical polarized Laguerre-Gaussian beams, *J. Opt. Soc. Am. A*, Vol. 15, Issue 10, 1998, pp. 2705-2711.
- [11]. S. R. Seshadri, Virtual source for the Bessel-Gauss beam, *Opt. Lett.*, Vol. 27, Issue 12, 2002, pp. 998-1000.
- [12]. S. R. Seshadri, Virtual source for a Laguerre-Gauss beam, *Opt. Lett.*, Vol. 27, Issue 12, 2002, pp. 1872-1874.
- [13]. S. R. Seshadri, Virtual source for a Hermite-Gauss beam, *Opt. Lett.*, Vol. 28, Issue 8, 2003, pp. 595-597.
- [14]. R. Borghi, M. Santarsiero, Nonparaxial propagation of spirally polarized optical beams, *J. Opt. Soc. Am. A*, Vol. 21, Issue 10, 2004, pp. 2029-2037.
- [15]. D. M. Deng, Nonparaxial propagation of radially polarized light beams, *J. Opt. Soc. Am. B*, Vol. 23, Issue 6, 2006, pp. 1228-1234.
- [16]. D. M. Deng, Q. Guo, L. J. Wu, X. B. Yang, Propagation of radially polarized elegant light Beams, *J. Opt. Soc. Am. B*, Vol. 24, Issue 3, 2007, pp. 636-643.
- [17]. D. M. Deng, Q. Guo, S. Lan, X. Yang, Application of the multiscale singular perturbation method to nonparaxial beam propagations in free space, *J. Opt. Soc. Am. A*, Vol. 24, Issue 10, 2007, pp. 3317-3325.
- [18]. S. R. Seshadri, Nonparaxial corrections for the fundamental Gaussian beam, *J. Opt. Soc. Am. A*, Vol. 19, Issue 10, 2002, pp. 2134-2141.
- [19]. R. Borghi, M. Santarsiero, Summing Lax series for nonparaxial beam propagation, *Opt. Lett.*, Vol. 28, Issue 10, 2003, pp. 774-776.
- [20]. Y. C. Zhang, Y. J. Song, Z. R. Chen, J. H. Ji, Z. X. Shi, Virtual sources for a Cosh-Gaussian beam, *Opt. Lett.*, Vol. 32, Issue 3, 2007, pp. 292-294.
- [21]. S. Quabis, R. Dorn, M. Eberler, O. Glöckl, G. Leuchs, The focus of light - theoretical calculation and experimental tomographic reconstruction, *Appl. Phys. B*, Vol. 72, Issue 1, 2001, pp. 109-113.
- [22]. R. Dorn, S. Quabis, G. Leuchs, Sharper focus for a radially polarized light beam, *Phys. Rev. Lett.*, Vol. 91, Issue 23, 2003, pp. 233901 (1-4).
- [23]. C. Varin, M. Piché, Acceleration of ultra-relativistic electrons using high-intensity TM 01 laser beams, *Appl. Phys. B*, Vol. 74, Issue 9, 2002, pp. S83-S88.
- [24]. Y. I. Salamin, Mono-energetic GeV electrons from ionization in a radially polarized laser beam, *Opt. Lett.*, Vol. 32, Issue 1, 2007, pp. 90-92.
- [25]. T. Kuga, Y. Torii, N. Shiokawa, T. Hirano, Y. Shimizu, H. Sasada, Novel optical trap of atoms with a doughnut beam, *Phys. Rev. Lett.*, Vol. 78, Issue 25, 1997, pp. 4713-4716.
- [26]. K. T. Gahagan, G. A. Swartzlander, Jr., Simultaneous trapping of low-index and high-index microparticles observed with an optical-vortex trap, *J. Opt. Soc. Am. B*, Vol. 16, Issue 4, 1999, pp. 533-537.
- [27]. L. Novotny, M. R. Beversluis, K. S. Youngworth, T. G. Brown, *Phys. Rev. Lett.*, Vol. 86, Issue 23, 2001, pp. 5251-5254.

- [28]. A. V. Nesterov, V. G. Niziev, Laser beams with axially symmetric polarization, *J. Phys. D*, Vol. 33, Issue 15, 2000, pp. 1817-1822.
- [29]. V. G. Niziev, A. V. Nesterov, Influence of beam polarization on laser cutting efficiency, *J. Phys. D*, Vol. 32, Issue 32, 1999, pp. 1455-1461.
- [30]. D. M. Deng, Q. Guo, Elegant Hermite-Laguerre-Gaussian beams, *Opt. Lett.*, Vol. 33, Issue 11, 2008, pp. 1225-1227.
- [31]. J. D. Ring, J. Lindberg, A. Mourka, M. Mazilu, K. Dholakia, M. R. Dennis, Auto-focusing and self-healing of Pearcey beams, *Opt. Express*, Vol. 20, Issue 17, 2012, pp. 18955-18966.
- [32]. D. M. Deng, Y. M. Gao, J. Y. Zhao, P. Zhang, Z. G. Chen, Three-dimensional nonparaxial beams in parabolic rotational coordinates, *Opt. Lett.*, Vol. 38, Issue 19, 2013, pp. 3934-3936.
- [33]. W. L. Barnes, A. Dereux, T. W. Ebbesen, Review article: Surface plasmon subwavelength optics, *Nature*, Vol. 424, 2003, pp. 824-830.
- [34]. T. W. Ebbesen, H. J. Lezec, H. F. Ghaemi, T. Thio, P. Wolff, Extraordinary optical transmission through sub-wavelength hole arrays, *Nature*, Vol. 391, Issue 6, 1998, pp. 1114-1117.
- [35]. S. A. Maier, P. G. Kik, H. A. Atwater, S. Meltzer, E. Harel, B. E. Koel, A. A. G. Requicha, Local detection of electromagnetic energy transport below the diffraction limit in metal nanoparticle plasmon waveguides, *Nat. Mater.*, Vol. 2, Issue 4, 2003, pp. 229-232.
- [36]. M. Lax, W. H. Louisell, W. B. McKnight, From Maxwell to paraxial wave optics, *Phys. Rev. A*, Vol. 11, Issue 4, 1975, pp. 1365-1370.
- [37]. A. Ciattoni, B. Crosignani, P. Di Porto, Vectorial free-space optical propagation: a simple approach for generating all-order nonparaxial corrections, *Opt. Commun.*, Vol. 177, Issue 1, 2000, pp. 9-13.
- [38]. Q. Cao, X. Deng, Corrections to the paraxial approximation of an arbitrary free-propagation beam, *J. Opt. Soc. Am. A*, Vol. 15, Issue 5, 1998, pp. 1144-1148.
- [39]. D. M. Deng, Q. Guo, Analytical vectorial structure of radially polarized light beams, *Opt. Lett.*, Vol. 32, Issue 18, 2007, pp. 2711-2713.
- [40]. D. M. Deng, C. D. Chen, X. Zhao, B. Chen, X. Peng, Y. S. Zheng, Virtual source of a Pearcey beam, *Opt. Lett.*, Vol. 39, Issue 9, 2014, pp. 2703-2706.
- [41]. A. W. Snyder, D. J. Mitchell, Accessible solitons, *Science*, Vol. 276, Issue 5318, 1997, pp. 1538-1541.
- [42]. M. Peccianti, C. Conti, G. Assanto, A. D. Luca, C. Umetsu, All-optical switching and logic gating with spatial solitons in liquid crystals, *Appl. Phys. Lett.*, Vol. 81, Issue 18, 2002, pp. 3335-3337.
- [43]. G. Assanto, M. Peccianti, C. Conti, Nematicons: optical spatial solitons in nematic liquid crystals, *Opt. Photonics News*, Vol. 14, Issue 2, 2003, pp. 44-48.
- [44]. A. W. Snyder, Y. Kivshar, Bright spatial solitons in non-Kerr media: stationary beams and dynamical evolution, *J. Opt. Soc. Am. B*, Vol. 11, Issue 14, 1997, pp. 3025-3031.
- [45]. W. Krolikowski, O. Bang, Solitons in nonlocal nonlinear media: exact solutions, *Phys. Rev. E*, Vol. 63, Issue 2, 2000, pp. 016610 (1-6).
- [46]. Y. R. Shen, Optical physics - solitons made simple, *Science*, Vol. 276, Issue 5318, 1997, pp. 1520-1520.
- [47]. C. Conti, M. Peccianti, G. Assanto, Observation of optical spatial solitons in a highly nonlocal medium, *Phys. Rev. Lett.*, Vol. 92, Issue 11, 2004, pp. 113902 (1-4).
- [48]. Q. Guo, W. Hu, D. M. Deng, D. Q. Lu, S. G. Ouyang, Features of strongly nonlocal spatial solitons, Chapter 2, in *Nematicons: Spatial Optical Solitons in Nematic Liquid Crystals* (G. Assanto, Ed.), *John Wiley & Sons*, New Jersey, 2013, pp. 37-69.
- [49]. C. Rotschild, O. Cohen, O. Manela, M. Segev, T. Carmon, Solitons in nonlinear media with an infinite range of nonlocality: first observation of coherent elliptic solitons and of vortexing solitons, *Phys. Rev. Lett.*, Vol. 95, Issue 21, 2005, pp. 213904 (1-4).

- [50]. Q. Guo, D. Q. Lu, D. M. Deng, Nonlocal spatial optical solitons, Chapter 4, in *Advances in Nonlinear Optics* (X. Chen, Qi Guo, W. She, H. Zeng, G. Zhang, Ed.), *De Gruyter*, Berlin, Feb. 2015, pp. 227-305.
- [51]. D. Buccoliero, A. S. Desyatnikov, W. Krolikowski, Y. S. Kivshar, Laguerre and Hermite soliton clusters in nonlocal nonlinear media, *Phys. Rev. Lett.*, Vol. 98, Issue 5, 2007, pp. 053901 (1-4).
- [52]. D. M. Deng, X. Zhao, Q. Guo, S. Lan, Hermite-Gaussian breathers and solitons in strongly nonlocal nonlinear media, *J. Opt. Soc. Am. B*, Vol. 24, Issue 9, 2007, pp. 2537-2544.
- [53]. D. M. Deng, Q. Guo, Propagation of Laguerre-Gaussian beams in nonlocal nonlinear media, *J. Opt. A: Pure Appl. Opt.*, Vol. 10, Issue 3, 2008, pp. 035101 (1-7).
- [54]. D. Deng, Q. Guo, Ince-Gaussian solitons in strongly nonlocal nonlinear media, *Opt. Lett.*, Vol. 32, Issue 21, 2007, pp. 3206-3208.
- [55]. D. Deng, Q. Guo, Propagation of elliptic-Gaussian beams in strongly nonlocal nonlinear media, *Phys. Rev. E*, Vol. 84, 2011, pp. 046604 (1-5).
- [56]. D. Deng, Qi Guo, W. Hu, Complex-variable-function-Gaussian solitons, *Opt. Lett.*, Vol. 34, Issue 1, 2009, pp. 43-45.
- [57]. D. Deng, Qi Guo, W. Hu, Complex-variable-function-Gaussian breathers and solitons, *Phys. Rev. A*, Vol. 79, Issue 2, 2009, pp. 023803 (1-6).
- [58]. A. F. Abouraddy, K. C. Toussaint Jr., Three-dimensional polarization control in microscopy, *Phys. Rev. Lett.*, Vol. 96, Issue 15, 2006, pp. 153901 (1-4).
- [59]. X. L. Wang, J. Chen, Y. N. Li, J. P. Ding, C. S. Guo, H. T. Wang, Optical orbital angular momentum from the curl of polarization, *Phys. Rev. Lett.*, Vol. 105, Issue 25, 2010, pp. 253602 (1-4).
- [60]. R. Martínez-Herrero, P. M. Mejías, S. Bosch, A. Carnicer, Vectorial structure of nonparaxial electromagnetic beams, *J. Opt. Soc. Am. A*, Vol. 18, Issue 7, 2001, pp. 1678-1680.
- [61]. P. M. Mejías, R. Martínez-Herrero, G. Piquero, J. M. Movilla, Parametric characterization of the spatial structure of non-uniformly polarized laser beams, *Prog. Quantum Electron*, Vol. 26, Issue 2, 2002, pp. 65-130.
- [62]. H. M. Guo, J. B. Chen, S. L. Zhuang, Vector plane wave spectrum of an arbitrary polarized electromagnetic wave, *Opt. Express*, Vol. 14, Issue 6, 2006, pp. 2095-2100.
- [63]. G. Q. Zhou, Analytical vectorial structure of Laguerre-Gaussian beam in the far field, *Opt. Lett.*, Vol. 31, Issue 17, 2006, pp. 2616-2618.
- [64]. *Handbook of Mathematical Functions* (M. Abramowitz, I. Stegun, Eds.), *Dover*, 1972.
- [65]. M. Willatzen, L. C. Lew Yan Voon, *Separable Boundary-Value Problems in Physics*, *Wiley*, 2011.
- [66]. P. Zhang, Z. Zhang, J. Prakash, S. Huang, D. Hernandez, M. Salazar, D. N. Christodoulides, Z. Chen, Trapping and transporting aerosols with a single optical bottle beam generated by moiré techniques, *Opt. Lett.*, Vol. 36, Issue 8, 2011, pp. 1491-1493.
- [67]. P. Zhang, J. Prakash, Z. Zhang, M. S. Mills, N. K. Efremidis, D. N. Christodoulides, Z. Chen, Trapping and guiding microparticles with morphing autofocusing Airy beams, *Opt. Lett.*, Vol. 36, Issue 15, 2011, pp. 2883-2885.
- [68]. D. J. Mitchell, A. W. Snyder, Soliton dynamics in a nonlocal medium, *J. Opt. Soc. Am. B*, Vol. 16, Issue 16, 1999, pp. 236-239.
- [69]. W. Krolikowski, O. Bang, N. I. Nikolov, D. Neshev, J. Wyller, J. J. Rasmussen, D. Edmundson, Modulational instability, solitons and beam propagation in spatially nonlocal nonlinear media, *J. Opt. B*, Vol. 6, Issue 5, 2004, pp. S288-S294.
- [70]. Q. Guo, B. Luo, F. Yi, S. Chi, Y. Xie, Large phase shift of nonlocal optical spatial solitons, *Phys. Rev. E*, Vol. 69, Issue 1, 2004, pp. 016602 (1-8).
- [71]. Q. Guo, Nonlocal spatial solitons and their interactions, in *Optical Transmission, Switching, and Subsystems* (C. F. Lam, C. Fan, N. Hanik, K. Oguchi, Eds.), *Proceedings of SPIE*, Vol. 5281, 2004, pp. 581-594.

- [72]. Y. Xie, Q. Guo, Phase modulations due to collisions of beam pairs in nonlocal nonlinear media, *Opt. Quantum Electron.*, Vol. 36, Issue 15, 2004, pp. 1335-1351.
- [73]. D. M. Deng, Q. Guo, Ince-Gaussian beams in strongly nonlocal nonlinear media, *J. Phys. B*, Vol. 41, Issue 14, 2008, pp. 145401 (1-7).
- [74]. S. G. Ouyang, Q. Guo, W. Hu, Perturbative analysis of generally nonlocal spatial optical solitons, *Phys. Rev. E*, Vol. 74, Issue 2, 2006, pp. 036622 (1-13).
- [75]. S. G. Ouyang, W. Hu, Q. Guo, Light steering in a strongly nonlocal nonlinear medium, *Phys. Rev. A*, Vol. 76, Issue 5, 2007, pp. 053832 (1-5).
- [76]. M. Peccianti, A. Dyadyusha, M. Kaczmarek, G. Assanto, Tunable refraction and reflection of self-confined light beams, *Nat. Phys.*, Vol. 2, Issue 11, 2006, pp. 737-742.
- [77]. C. Rotschild, B. Alfassi, O. Cohen, M. Segev, Long-range interactions between optical solitons, *Nat. Phys.*, Vol. 2, Issue 11, 2006, pp. 769-774.
- [78]. C. Rotschild, M. Segev, Z. Y. Xu, Y. V. Kartashov, L. Torner, O. Cohen, Two-dimensional multipole solitons in nonlocal nonlinear media, *Opt. Lett.*, Vol. 31, Issue 22, 2006, pp. 3312-3314.
- [79]. X. Hutsebaut, C. Cambournac, M. Haelterman, A. Adamski, K. Neyts, Single-component higher-order mode solitons in liquid crystals, *Opt. Commun.*, Vol. 233, Issue 1, 2004, pp. 211-217.
- [80]. W. Hu, T. Zhang, Q. Guo, L. Xuan, S. Lan, Single-component higher-order mode solitons in liquid crystals, *Appl. Phys. Lett.*, Vol. 89, Issue 7, 2006, pp. 071111 (1-3).
- [81]. W. Hu, S. G. Ouyang, P. B. Yang, Q. Guo, S. Lan, Short-range interactions between strongly nonlocal spatial solitons, *Phys. Rev. A*, Vol. 77, Issue 3, 2008, pp. 033842 (1-6).
- [82]. S. Lopez-Aguayo, A. S. Desyatnikov, Y. Kivshar, S. Skupin, W. Krolikowski, O. Bang, Stable rotating dipole solitons in nonlocal optical media, *Opt. Lett.*, Vol. 31, Issue 8, 2006, pp. 1100-1102.
- [83]. A. I. Yakimenko, Y. A. Zaliznyak, Y. Kivshar, Stable vortex solitons in nonlocal self-focusing nonlinear media, *Phys. Rev. E*, Vol. 71, Issue 6, 2005, pp. 065603 (1-4).
- [84]. S. Lopez-Aguayo, A. S. Desyatnikov, Y. S. Kivshar, Azimuthons in nonlocal nonlinear media, *Opt. Express*, Vol. 14, Issue 17, 2006, pp. 7903-7908.
- [85]. S. Lopez-Aguayo, J. C. Gutiérrez-Vega, Elliptically modulated self-trapped singular beams in nonlocal nonlinear media: ellipticons, *Opt. Express*, Vol. 15, Issue 26, 2007, pp. 18326-18338.
- [86]. D. Buccoliero, A. S. Desyatnikov, W. Krolikowski, Y. S. Kivshar, Spiraling multivortex solitons in nonlocal nonlinear media, *Opt. Lett.*, Vol. 33, Issue 2, 2008, pp. 198-200.
- [87]. A. Fratalocchi, A. Piccardi, M. Peccianti, G. Assanto, Nonlinearly controlled angular momentum of soliton clusters, *Opt. Lett.*, Vol. 32, Issue 11, 2007, pp. 1447-1449.
- [88]. F. M. Arscott, Periodic Differential Equations, *Pergamon*, Oxford, 1964.
- [89]. F. M. Arscott, The Whittaker-Hill equation and the wave equation in paraboloidal coordinates, *Proceedings of R. Soc. Edinburgh Sect. A*, Vol. 67, Issue 4, 1967, pp. 265-276.
- [90]. M. A. Bandres, J. C. Gutiérrez-Vega, Ince-Gaussian modes of the paraxial wave equation and stable resonators, *J. Opt. Soc. Am. A*, Vol. 21, Issue 5, 2004, pp. 873-880.
- [91]. M. A. Bandres, J. C. Gutiérrez-Vega, Ince-Gaussian series representation of the two-dimensional fractional Fourier transform, *Opt. Lett.*, Vol. 30, Issue 5, 2005, pp. 540-542.
- [92]. U. T. Schwarz, M. A. Bandres, J. C. Gutiérrez-Vega, Observation of Ince-Gaussian modes in stable resonators, *Opt. Lett.*, Vol. 29, Issue 16, 2004, pp. 1870-1872.
- [93]. U. T. Schwarz, M. A. Bandres, J. C. Gutiérrez-Vega, Formation of Ince-Gaussian modes in a stable laser oscillator, *Proceedings Of SPIE*, Bellingham, WA, 2005, Vol. 5708, Issue 5, 2005, pp. 124-131.
- [94]. G. P. Agrawal, Nonlinear Fiber Optics, 2nd Ed., *Academic*, San Diego, CA, 1995.
- [95]. D. V. Skryabin, J. M. McSloy, W. J. Firth, Stability of spiralling solitary waves in Hamiltonian systems, *Phys. Rev. E*, Vol. 66, Issue 5, 2002, pp. 055602(R) (1-4).

Chapter 6

Incoherently Coupled Soliton Families in Photorefractive Media

Aavishkar Katti and R. A. Yadav

6.1. Introduction

The Photorefractive effect is established as a high sensitivity non-linear optical effect, and hence has been quite an attractive medium for holographic storage and other optical processing [1]. With the advancement in synthesis and growth of better and more sensitive photorefractive materials in the past decade, research avenues in such photorefractive crystals are opening up significantly. Photorefractive effect is basically a refractive index change induced through the electro-optic effect due to an electric field produced by a photo generated space charge field. Recent research has progressed broadly under two sub branches. One is the diffusion driven wave-mixing schemes, which make use of the two wave mixing and four wave mixing for various investigations like two-wave mixing gain [2-5], MI gain [6-7], space charge field in presence of an electric field or magnetic field [8-13], holographic photorefractive gratings and their diffraction efficiency [14-19], among others. The other branch in which research has progressed, and more so in the past two decades at a frenetic pace, is the self-trapping in photorefractive materials [20-23]. This phenomenon gives rise to optical spatial solitons.

Optical spatial solitons in photorefractive materials have been studied widely in the recent few decades. On illumination, a space charge field is set up in photorefractive materials, due to which the refractive index changes non-linearly with incident intensity. This is due to the Pockels effect. This process counteracts the effects of diffraction. Hence, a soliton beam can be formed which propagates without changing its shape due to an exact balance between diffraction and self focusing or defocusing non-linearity. Currently, many different types of solitons have been observed, of which some important ones are quasi steady state solitons [24, 25], screening solitons [26-28], photovoltaic solitons [29-32], screening photovoltaic solitons [33-35]. Traditionally, these solitons have been identified in non-centrosymmetric crystals and studied extensively.

Screening solitons in centrosymmetric photorefractive crystals were predicted first by Segev and Agranat [36]. These solitons are governed by the quadratic electro-optic effect i. e. dc Kerr effect and so quite different from the solitons in non-centrosymmetric media which are governed by the Pockels effect. The experimental confirmation came by the studies of Del Re et al. [37, 38] who demonstrated the self trapping of a bright soliton in a KLTN crystal. These screening solitons form due to the non-uniform screening of the external electric field. Recently, the soliton formation due to the combined effect of the drift and diffusion effect in such centrosymmetric photorefractive materials has been investigated [39].

If two mutually incoherent collinear propagating soliton beams have the same frequency and polarization, they create an effective refractive index modulation due to the combination of their intensities. Each of the soliton beams experiences the refractive index modulation effected by both the beams and hence a coupled pair is formed. This can be extended to include incoherently coupled multicomponent solitons in case the incident soliton beams are more than two in number. Incoherently coupled spatial solitons were first investigated by Christodoulides et al. [40] many years ago. Since then, incoherently coupled soliton pairs in various realizations have been studied such as screening-photovoltaic solitons [41-44], screening solitons [45] and photovoltaic solitons [46-48]. Incoherently coupled multicomponent solitons or soliton families have also been investigated in case of screening photovoltaic solitons in both, the symmetric and hybrid realisations [41-44]. Again, all these have been predicted in non-centrosymmetric PR crystals. Incoherently coupled bright-dark soliton pair in centrosymmetric PR medium has been predicted and studied theoretically a few years ago [49]. Recently, it has been found that the two-photon non-linearity can also support these incoherently coupled soliton pairs and multicomponent solitons in centrosymmetric PR media [50]. Also, the incoherent coupling of two Gaussian beams in centrosymmetric PR media have been studied in quite detail [51].

The net internal field inside the crystal is zero at equilibrium in a ferroelectric crystal. This is because the charge distribution on the crystal faces compensates the field due to spontaneous polarisation. In such a crystal, a temperature change causes spontaneous polarisation change and hence, a transient electric field E_{py} . This is called as the pyroelectric field. This field is not compensated immediately and consequently, a drift current can be set up mimicking the effect of an external electric field applied to the crystal. Now, this field is locally screened due to the space charge field formed due to the photorefractive effect and hence, a self-trapped beam results.

Pyroelectric spatial solitons or pyrolitons, originating from the pyroelectric field in photorefractive photovoltaic crystals have been predicted and observed [52, 53]. The theory for spatial solitons in pyroelectric photovoltaic photorefractive media has also been discussed for open circuit crystals [54]. The writing of soliton waveguides due to the pyroelectric effect in lithium niobate has been discussed in [55, 56]. Recently, we investigated the effect of pyroelectricity on solitons in biased photovoltaic photorefractive media [57]. Also, incoherently coupled soliton pairs have been shown to exist

theoretically in open circuit photovoltaic-photorefractive crystals along with the pyroelectric effect [58].

Notable is the fact that [52-58] these researches were in photovoltaic photorefractive crystals. It has been recently predicted that significant pyroelectric effects alone can also support solitons in photorefractive crystals which do not have any bulk photovoltaic effects [59]. In addition, it is emphasized that replacing the external electric field with the temperature change induced pyroelectric field has clear advantages. Firstly, we need not know the direction of the crystal c-axis since the pyroelectric field is always along the c-axis, i. e., in one direction for heating and in the reverse direction for cooling. Secondly, no electrodes are required on the crystal [55, 56]. The coupling of solitons in photorefractive media in which the external electric field is replaced by a temperature induced pyroelectric field has not yet been investigated and is of quite a degree of interest. In this chapter, a theory for incoherently coupled soliton pairs solely due to the pyroelectric effect in photorefractive crystals is formulated. The existence of incoherently coupled soliton pairs in the four realizations, namely, bright-bright, dark-dark, grey-grey, and bright-dark is predicted. The conditions for the formation of each of these soliton pairs are discussed in detail. In addition, the effects of the magnitude of the temperature change and the total beam intensity on the self trapping of the soliton pairs have been studied.

In this chapter, the existence and properties of the incoherently coupled bright, dark and greysoliton pairs in centrosymmetric photorefractive media will be evinced. This formulation is extended to incorporate incoherently coupled multicomponent solitons in centrosymmetric photorefractive media. The stability of these soliton pairs is studied by the modulation instability theory. The local modulation instability gain for all three types of soliton pairs and multicomponent solitons will be derived.

6.2. Theoretical Model

6.2.1. Dynamical Evolution Equation for Soliton families in Centrosymmetric Photorefractive Material

We consider mutually incoherent optical beams propagating along the z -axis. We assume diffraction only in the x direction. The soliton beams are polarized in the x - direction and the external electric field is also applied in the same direction.

Now, in centrosymmetric materials, the change in extraordinary refractive index is [36, 60],

$$\Delta n = n'_e - n_e = -\frac{1}{2} n_e^3 g_{eff} \epsilon_0^2 (\epsilon_r - 1)^2 E_{sc}^2, \quad (6.1a)$$

from which we get, (to the first order of g_{eff}),

$$(n'_e)^2 = n_e^2 - n_e^4 g_{eff} \epsilon_0^2 (\epsilon_r - 1)^2 E_{sc}^2. \quad (6.1b)$$

This relation (6.1a) is analogous to the case of non-centrosymmetric materials where we find that [27, 61]

$$(n'_e)^2 = n_e^2 - n_e^4 g_{eff} E_{sc}. \quad (6.1c)$$

The total electric field $\vec{E} = \vec{E}_1 + \vec{E}_2$ satisfies the Helmholtz equation,

$$\nabla^2 \vec{E} + (k_0 n'_e)^2 \vec{E} = 0. \quad (6.2)$$

As usual, the incident beams are expressed as slowly varying envelopes $\vec{E}_1 = \hat{x} A_1(x, z) \exp(ikz)$ and $\vec{E}_2 = \hat{x} A_2(x, z) \exp(ikz)$ where $k = k_0 n_e$, $k_0 = \frac{2\pi}{\lambda_0}$. Here, n_e is the unperturbed extraordinary refractive index and λ_0 is the free space wavelength.

Substituting the forms of \vec{E} and (6.1b) in (6.2), we get the paraxial equation of diffraction [36, 50-51],

$$\left(i \frac{\partial}{\partial z} + \frac{1}{2k} \frac{\partial^2}{\partial x^2} + \frac{k}{n_e} \Delta n \right) A_j(x, z) = 0, j=1,2 \quad (6.3)$$

with

$$\Delta n = -\frac{1}{2} n_e^3 g_{eff} \epsilon_0^2 (\epsilon_r - 1)^2 E_{sc}^2, \quad (6.4)$$

where E_{sc} is the space charge field in the medium resulting from the external electric field and the optical beams intensity. g_{eff} is the effective quadratic electro-optic coefficient ϵ_0 and ϵ_r are the vacuum permittivity and the dielectric constant.

If the bias field is strong, and the beam is relatively broad, the static space charge field is given as [62],

$$E_{sc} = E_0 \frac{I_\infty + I_d}{I + I_d}, \quad (6.5)$$

where we have neglected the effects of diffusion. I_∞ is the intensity as $x \rightarrow \pm\infty$. I_d is the dark irradiance. E_0 is the value of space charge electric field as $x \rightarrow \pm\infty$. For finding E_0 , consider the circuit shown in Fig. 6.1. We have the following potential condition

$$\mathcal{E} = -\int_{-l/2}^{l/2} E_{sc} dx + RJS, \quad (6.6)$$

where \mathcal{E} is the electro-motive force of the source, R is the resistance connected to the crystal, S is surface area of the electrodes, and l is the width of the crystal between the

electrodes [63]. Since, the space charge field does not depend explicitly on the resistivity of the material, and as KLTN crystal has substantial internal resistance at room temperature, we can also consider R as a combination of the internal resistance of the crystal and the load resistance connected.

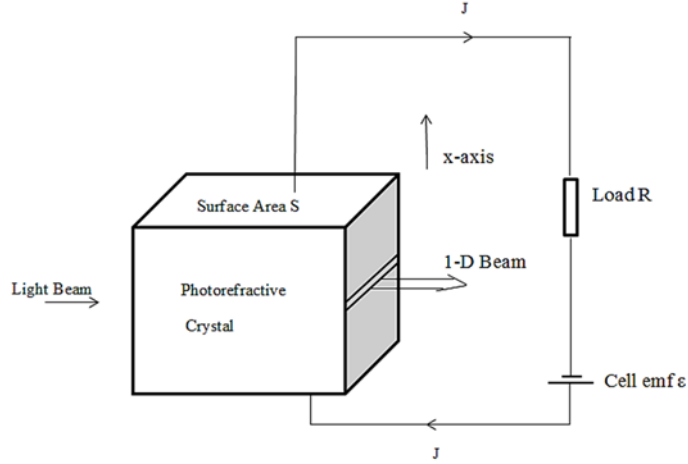


Fig. 6.1. Illustration of the electrical circuit consisting of a centrosymmetric photorefractive crystal biased by a cell of EMF ε and an external resistor R .

Substitution of (6.5) into (6.6) gives,

$$E_{sc} = -(\varepsilon\alpha - RJS\alpha) \frac{I_{\infty} + I_d}{I + I_d}, \quad (6.7)$$

where

$$\alpha = \frac{1}{\int_{-l/2}^{l/2} (I_{\infty} + I_d)(I + I_d)^{-1} dx}. \quad (6.8)$$

The total optical power density for the two mutually incoherent beams can be obtained by adding the two Poynting fluxes,

$$I = \frac{n_e}{2\eta_0} (|A_1|^2 + |A_2|^2), \quad (6.9)$$

with $\eta_0 = (\mu_0/\varepsilon_0)^{1/2}$.

Substituting E_{sc} and Δn in (6.3) and in terms of dimensionless variables, one gets the following equation,

$$iU_\xi + \frac{1}{2}U_{ss} - \frac{\beta(1+\rho)^2 U}{(1+|U|^2 + |V|^2)^2} = 0, \quad (6.10)$$

$$iV_\xi + \frac{1}{2}V_{ss} - \frac{\beta(1+\rho)^2 V}{(1+|U|^2 + |V|^2)^2} = 0, \quad (6.11)$$

where we have written, $A_1 = (2\eta_0 I_d / n_e)^{1/2} U$ and $A_2 = (2\eta_0 I_d / n_e)^{1/2} V$, $\xi = z/kx_0^2$, $s = x/x_0$ and, $U_\xi = \frac{\partial U}{\partial \xi}$, $U_{ss} = \frac{\partial^2 U}{\partial s^2}$, $\rho = I_\infty / I_d$, $\beta = (k_0 x_0)^2 n_e^4 g_{eff} \epsilon_0^2 (\epsilon_r - 1)^2 \frac{(RJS\alpha - \epsilon\alpha)^2}{2}$.

An interesting case to investigate is the case of low amplitude, i. e. $|U|^2, |V|^2 \ll 1$. Following the approach of [45] for investigating the special case of low amplitude solitons, in our case, Eq. (6.10) and (6.11) are modified as,

$$iU_\xi + \frac{1}{2}U_{ss} - \beta(1+\rho)^2 U + 2\beta(1+\rho)^2 (|U|^2 + |V|^2)U = 0, \quad (6.12)$$

$$iV_\xi + \frac{1}{2}V_{ss} - \beta(1+\rho)^2 V + 2\beta(1+\rho)^2 (|U|^2 + |V|^2)V = 0. \quad (6.13)$$

It is to be noted that (6.12) and (6.13) represent the modified form of the Non Linear Schrodinger equation.

Now, we shall work out a formulation for the incoherently coupled multicomponent solitons.

In this case, the $2N$ mutually incoherent beams are made up of $\sum_{j=1}^N \vec{E}_j$ and $\sum_{m=1}^N \vec{E}_m$.

The total electric field component $\vec{E} = \vec{E}_1 + \vec{E}_2 + \vec{E}_3 + \dots + \vec{E}_{2N}$ satisfy the Helmholtz Eq.(6.2), where $k_0 = 2\pi/\lambda_0$ and λ_0 is the free space wavelength.

The electric fields are expressed as slowly varying envelopes $\sum_{j=1}^N \vec{E}_j = \sum_{j=1}^N \hat{x}\varphi_j(x, z) \exp(ikz)$ and $\sum_{m=1}^N \vec{E}_m = \sum_{m=1}^N \hat{x}\psi_m(x, z) \exp(ikz)$.

Substituting the forms of \vec{E} and n' into the Helmholtz equation leads us to the paraxial equation as before,

$$\left(i \frac{\partial}{\partial z} + \frac{1}{2k} \frac{\partial^2}{\partial x^2} + \frac{k}{n_e} \Delta n \right) \varphi_j(x, z) = 0, \quad j = 1, 2, \dots, N, \quad (6.14)$$

$$\left(i \frac{\partial}{\partial z} + \frac{1}{2k} \frac{\partial^2}{\partial x^2} + \frac{k}{n_e} \Delta n\right) \psi_m(x, z) = 0, \quad m=1, 2, \dots, N, \quad (6.15)$$

with

$$\Delta n = -\frac{1}{2} n_e^3 g_{\text{eff}} \epsilon_0^2 (\epsilon_r - 1)^2 E_{sc}^2. \quad (6.16)$$

The total optical power density for the 2N mutually incoherent beams is

$$I = \frac{n_e}{2\eta_0} \left(\sum_{j=1}^N |\varphi_j|^2 + \sum_{m=1}^N |\psi_m|^2 \right), \quad (6.17)$$

with $\eta_0 = (\mu_0/\epsilon_0)^{1/2}$.

As usual, we write $\varphi_j = (2\eta_0 I_d / n_e)^{1/2} U_j$ and $\psi_m = (2\eta_0 I_d / n_e)^{1/2} V_m$, $\xi = z/kx_0^2$, $s = x/x_0$.

Using the value of the space charge electric field derived before and substituting into the paraxial wave equation, we obtain

$$iU_\xi + \frac{1}{2} U_{ss} - \frac{\beta(1+\rho)^2 U}{(1 + \sum_{j=1}^N |U_j|^2 + \sum_{m=1}^N |V_m|^2)^2} = 0, \quad (6.18)$$

$$iV_\xi + \frac{1}{2} V_{ss} - \frac{\beta(1+\rho)^2 V}{(1 + \sum_{j=1}^N |U_j|^2 + \sum_{m=1}^N |V_m|^2)^2} = 0, \quad (6.19)$$

where $U_\xi = \frac{\partial U}{\partial \xi}$, $U_{ss} = \frac{\partial^2 U}{\partial s^2}$, $\rho = I_\infty / I_d$, $\beta = (k_0 x_0)^2 n_e^4 g_{\text{eff}} \epsilon_0^2 (\epsilon_r - 1)^2 \frac{(RJS\alpha - \epsilon\alpha)^2}{2}$.

6.2.2. Modulation Instability

Modulation instability is a fundamental phenomenon associated with wave propagation. Small perturbations originating from the noise on a uniform wave front, which grow under the combined effect of diffraction and non-linearity is known as modulation instability. This breaks up the uniform wave front and leads to formation of multiple filaments of the beam.

Since we neglect the terms related to the diffusion effect in Eq. (6.5) due to a broad beam and strong bias conditions, our investigation of the modulation instability will be local.

A beam travelling in a centrosymmetric photorefractive medium satisfies the paraxial diffraction equation,

$$iU_{\xi} + \frac{1}{2}U_{ss} - \frac{\beta(1+\rho)^2 U}{(1+|U|^2)^2} = 0. \quad (6.20)$$

Now, for a continuous wave, (6.20) takes the form,

$$iU_{\xi} - \frac{\beta(1+\rho)^2 U}{(1+|U|^2)^2} = 0. \quad (6.21)$$

Eq. (6.21) has a steady state solution,

$$U = r'^{1/2} \exp(-i\{\beta(1+\rho)^2 / (1+r')^2\}\xi). \quad (6.22)$$

For studying the modulation instability of the plane wave, we express,

$$U = [r'^{1/2} + \sigma(s, \xi)] \exp(-i\{\beta(1+\rho)^2 / (1+r')^2\}\xi), \quad (6.23)$$

where $\sigma(s, \xi)$ is the weak modulation term added to the steady state solution (6.22). This weak perturbation satisfies

$$|\sigma(s, \xi)|^2 \ll r'^{1/2}, \quad (6.24)$$

and is assumed,

$$\sigma = a(\xi) \exp(ips) + b(\xi) \exp(-ips). \quad (6.25)$$

We want to investigate the possible exponential growth of this perturbation. For this purpose, substituting (6.25) in (6.20) and using the linearizing condition (6.24), we get the following equation for the perturbation $\sigma(s, \xi)$,

$$i \frac{\partial \sigma}{\partial \xi} + \frac{1}{2k} \frac{\partial^2 \sigma}{\partial s^2} + 2\beta(1+\rho)^2 \frac{r'}{(1+r')^3} (\sigma + \sigma^*) = 0. \quad (6.26)$$

Substituting (6.25) in (6.26), we obtain the following coupled equations,

$$i \frac{da}{d\xi} - \frac{1}{2k} p^2 a + 2\beta(1+\rho)^2 \frac{r'}{(1+r')^3} (a + b^*), \quad (6.27)$$

$$i \frac{db}{d\xi} - \frac{1}{2k} p^2 b + 2\beta(1+\rho)^2 \frac{r'}{(1+r')^3} (a^* + b). \quad (6.28)$$

By decoupling equations (6.27) and (6.28), we get simply,

$$\frac{d^2 a}{d\xi^2} = \left[2\beta(1+\rho)^2 \frac{r'}{(1+r')^3} \frac{p^2}{k} - \frac{p^4}{4k^2} \right] a, \quad (6.29)$$

$$\frac{d^2 b}{d\xi^2} = \left[2\beta(1+\rho)^2 \frac{r'}{(1+r')^3} \frac{p^2}{k} - \frac{p^4}{4k^2} \right] b. \quad (6.30)$$

Eqs. (6.29) and (6.30) are of the form of a linear second order differential equation. Hence, their solutions are of the form $\exp(\omega \xi)$ with

$$\omega = \left[2\beta(1+\rho)^2 \frac{r'}{(1+r')^3} \frac{p^2}{k} - \frac{p^4}{4k^2} \right]^{1/2}. \quad (6.31)$$

Hence, the local modulation instability gain is given as,

$$g = \text{Re} \left\{ \left[2\beta(1+\rho)^2 \frac{r'}{(1+r')^3} \frac{p^2}{k} - \frac{p^4}{4k^2} \right]^{1/2} \right\}. \quad (6.32)$$

From (6.32), we can find the maximum modulation instability gain as,

$$g_{\max} = \left[2k_0 E_0^2 (1+\rho)^2 n_e^3 g_{\text{eff}} \epsilon_0^2 (\epsilon_r - 1)^2 \frac{r'}{(1+r')^3} \right]. \quad (6.33)$$

It is notable that in [64], a similar derivation of the MI gain has been done for both local and global space charge field. But they consider only the bright like planar beam as is evinced by the particular expression of the space charge field they use in the dynamical evolution equation, i. e. $I_\infty = 0$. We follow [64] in our derivation but we consider the general expression for the space charge field in the dynamical evolution equation and hence, our derivation is valid for all three cases: bright, dark, and grey like planar beams. We can approximately deduce that the MI gain for each component of the soliton pairs is dependent upon r' , ρ , and also E_0 .

6.2.3. Dynamical Evolution Equation for Soliton Pairs in a Pyroelectric Photorefractive Material

We shall consider two optical beams propagating collinearly in a photorefractive pyroelectric non-photovoltaic crystal. We shall consider the diffraction in x -direction only and that the beams are polarized linearly along the x -direction. The crystal is kept between an insulating plastic cover and a metallic plate whose temperature is controlled accurately. As usual, the incident beams are expressed as slowly varying envelopes

$\vec{E}_1 = \hat{x} A_1(x, z) \exp(ikz)$ and $\vec{E}_2 = \hat{x} A_2(x, z) \exp(ikz)$ where $k = k_0 n_e$, $k_0 = \frac{2\pi}{\lambda_0}$. Here, n_e is

the unperturbed extraordinary refractive index and λ_0 is the free space wavelength. Under these conditions, the evolution of the optical beams is given by [20, 29],

$$\left(i \frac{\partial}{\partial z} + \frac{1}{2k} \frac{\partial^2}{\partial x^2} + \frac{k}{n_e} \Delta n\right) A_j(x, z) = 0, j=1, 2, \quad (6.34)$$

$$\Delta n = -\frac{1}{2} n_e^3 g_{eff} E_{sc}, \quad (6.35)$$

where $E_{sc} = E_{pysc}$ is the space charge field in the medium resulting from the pyroelectric field and the optical beams intensity [44]. We note here that the diffusion effect has been neglected. g_{eff} is the effective linear electro-optic coefficient. For non-photovoltaic photorefractive crystals, E_{pysc} for steady state (for a time $t \gg t_d$ where t_d is the characteristic Maxwell time) has been derived in detail elsewhere as, [44]

$$E_{pysc} = -E_{py} \frac{I}{I + I_d}, \quad (6.36)$$

where E_{py} is the transient pyroelectric field and is given by [37],

$$E_{py} = -\frac{1}{\epsilon_0 \epsilon_r} \frac{\partial P}{\partial T} \Delta T, \quad (6.37)$$

where $\frac{\partial P}{\partial T}$ is the pyroelectric coefficient and ΔT is the magnitude of the temperature change of the crystal and I_d is the dark irradiance, i. e., the intensity at constant illumination region of the crystal, $I(x \rightarrow \infty)$.

From the expression for the space charge field (6.36), we can infer that the value and sign of the space charge field depends upon E_{py} which, in turn can be varied by the change in temperature, i. e., by the requisite heating, or cooling effects.

The total optical power density for the two mutually incoherent beams can be obtained by adding the two Poynting fluxes,

$$I = \frac{n_e}{2\eta_0} \left(|A_1|^2 + |A_2|^2 \right), \quad (6.38)$$

with $\eta_0 = (\mu_0/\epsilon_0)^{1/2}$.

Substituting E_{pysc} and Δn in (6.34) and in terms of dimensionless variables, one gets the following equation,

$$iU_\xi + \frac{1}{2} U_{ss} + \beta \frac{(|U|^2 + |V|^2)}{(1 + |U|^2 + |V|^2)} U = 0, \quad (6.39)$$

$$iV_\xi + \frac{1}{2}V_{ss} + \beta \frac{(|U|^2 + |V|^2)}{(1 + |U|^2 + |V|^2)}V = 0, \quad (6.40)$$

where we have written,

$$A_1 = (2\eta_0 I_d / n_e)^{1/2} U \text{ and } A_2 = (2\eta_0 I_d / n_e)^{1/2} V, \quad \xi = z/kx_0^2, \quad s = x/x_0, \quad U_\xi = \frac{\partial U}{\partial \xi}, \quad U_{ss} = \frac{\partial^2 U}{\partial s^2},$$

$$\rho = I_\infty / I_d, \quad I = I_d |U|^2, \quad \beta = \frac{1}{2}(k_0 x_0)^2 n_e^4 g_{eff} E_{py}$$

As we can see, the intensity scales with the dark irradiance in these dimensionless coordinates.

6.3. Discussion

6.3.1. Coupled Solitons in Centrosymmetric Photorefractive Material

6.3.1.1. Dark-Dark Soliton Pair

In case of a dark-dark incoherently coupled soliton pair, we express the solutions as,

$$U = \rho^{1/2} y(s) \cos(\theta) \exp(i\mu\xi), \quad (6.41)$$

$$V = \rho^{1/2} y(s) \sin(\theta) \exp(i\mu\xi), \quad (6.42)$$

where μ is the nonlinear shift of the propagation constant, θ is the arbitrary projection angle, and $y(s)$ is the normalized bounded function which satisfies $0 \leq y(s) \leq 1$ and $y(\pm\infty) = \pm 1, \dot{y}(\pm\infty) = 0, y(0) = 0, \ddot{y}(\pm\infty) = 0$.

Substituting (6.41) and (6.42) in (6.10) and (6.11), we get a single equation,

$$\frac{1}{2}\ddot{y} - \mu y - \frac{\beta(1+\rho)^2}{(1+\rho y^2)^2} y = 0. \quad (6.43)$$

Integrating once, we get,

$$\frac{1}{2}\dot{y}^2 = \mu y^2 - \frac{\beta(1+\rho)^2}{\rho(1+\rho y^2)} + c, \quad (6.44)$$

where c is the constant of integration. Using the boundary conditions $\dot{y}(\pm\infty) = 0$ and $y(\pm\infty) = \pm 1$ in (6.43), we get,

$$\mu = -\beta. \quad (6.45)$$

Now, using the boundary conditions $\dot{y}(\pm\infty) = 0$, $y(\pm\infty) = \pm 1$ in (6.37), we can find out the value of the constant c as,

$$c = \beta \left(\frac{2\rho + 1}{\rho} \right). \quad (6.46)$$

Putting the value of c in (6.44), we have,

$$\dot{y}^2 = (-2\beta) \left(y^2 + \frac{(1+\rho)^2}{\rho(1+\rho y^2)} - \frac{2\rho+1}{\rho} \right). \quad (6.47)$$

From which we integrate once more to get,

$$(-2\beta)^{1/2} s = \int_y^0 \frac{d\tilde{y}}{\left[\tilde{y}^2 + \frac{(1+\rho)^2}{\rho(1+\rho\tilde{y}^2)} - \frac{2\rho+1}{\rho} \right]^{1/2}}. \quad (6.48)$$

Eq. (6.48) gives us the soliton intensity envelope. Also, as we can see in Eq. (6.47), the second term on the RHS will always be greater than the first term as $y(s)$ is bounded between 0 and 1. Hence, we get the condition that $\beta < 0$ i. e., $g_{eff} < 0$ for such dark incoherently coupled soliton pair. This is consistent with the result obtained by [36] for existence of single dark solitons. We note here that values of parameters ε , α , R , S influence the properties of the coupled solitons. In particular, we shall study the effect of ε , the source EMF; R , the circuit resistance; and the intensity, through the parameter r on the soliton pair.

To illustrate our result, we consider a crystal parameters similar to KLTN crystal but with negative g_{eff} so as to support dark solitons. We have, $n_e = 2.2$, $g_{eff} = -0.12 \text{ m}^4\text{C}^{-2}$, $T = 21 \text{ }^\circ\text{C}$, $\varepsilon_r = 8000$, $\lambda_0 = 0.5 \text{ }\mu\text{m}$, $x_0 = 40 \text{ }\mu\text{m}$. [65]. Now, we assume two cases of $\varepsilon = 2000 \text{ V}$ and 4000 V with $l = 7.5 \text{ mm}$. We also choose parameters in such a way that $RJS\alpha = \frac{1}{2}\varepsilon\alpha$. Apart from the fact that this assumption is physically acceptable, there is nothing special in this choice. One can get other values also depending on the choice of R (and hence J), α and S .

Fig. 6.2 shows the intensity profiles of the incoherently coupled soliton dark pair, for the cases of $\varepsilon = 2000 \text{ V}$ and 4000 V respectively. We find the FWHM of the two components to be $4.8 \text{ }\mu\text{m}$. in the first case when $\varepsilon = 2000 \text{ V}$. This decreases to about $3.2 \text{ }\mu\text{m}$ when $\varepsilon = 4000 \text{ V}$.

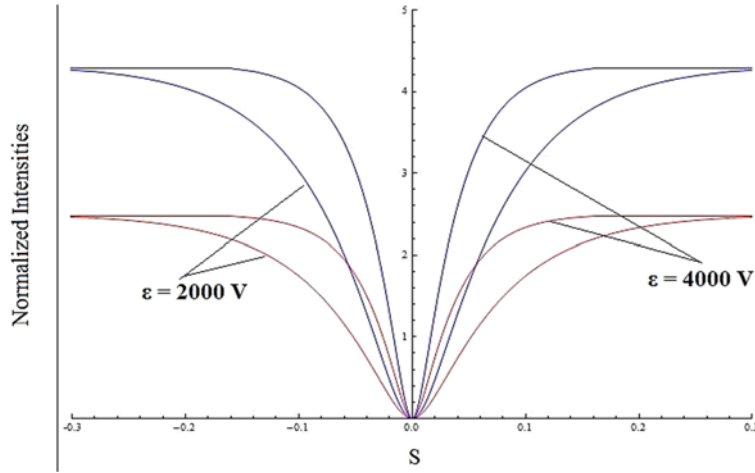


Fig. 6.2. Comparison of Normalized Intensities of the components $|U|^2$ (blue) and $|V|^2$ (red) of the incoherently coupled soliton pair when $\varepsilon = 2000$ V and $\varepsilon = 4000$ V, $\rho = 5$, $\theta = \pi/3$. [20]

We note here, that we have assumed the g_{eff} to be negative for the sake of showing existence of dark solitons, whereas in actuality, KLTN does not support dark solitons as g_{eff} is positive for the crystal. As we know, soliton formation in centrosymmetric photorefractive materials like KLTN is due to the quadratic Kerr non-linearity. There has been recent upsurge in interest in KLTN, namely research on continuous solitons in a lattice nonlinearity and also spatial rogue waves in ferroelectric crystals [66-67], and KLTN remains primarily a self-focusing medium. Now, recently, [68-69] showed the existence of dark solitons in MQW waveguides which are having a quadratic nonlinearity. MQW waveguides have $g_{eff} < 0$ [70]. In addition, we have studied the coherent coupling of dark solitons in these MQW waveguides [71]. To date, [68] remains the only experimental confirmation of dark solitons in a medium with a quadratic non-linearity.

Now, if we take the external resistance as zero and also the internal resistance of the crystal is negligible, then $R \rightarrow 0$ for this case. This affects the value of E_0 in the space charge field. Hence we would expect a change in the characteristics of the coupled pairs for both bias fields now. The normalized intensities are plotted in Fig. 6.3 for the case of $R \rightarrow 0$. The FWHM of the pairs are $2.7 \mu\text{m}$, $1.6 \mu\text{m}$ for $\varepsilon = 2000$, 4000 V respectively. These are much lesser when we compare with the corresponding values of FWHM for the $R \neq 0$ case discussed previously which implies an enhancement in the self trapping as the internal and external resistance decreases.

Now, in the case of low intensity, we plot the intensity profiles of the soliton pair components U and V in Fig. 6.4 for $\varepsilon = 2000$ V and 4000 V when the peak $|U|^2 \ll 1 \approx 0.05$ and peak $|V|^2 \ll 1 \approx 0.08$. We find that the FWHM for these intensities of the coupled pair for $\varepsilon = 2000$ V and 4000 V are $14.4 \mu\text{m}$. and $7.2 \mu\text{m}$. This is much larger than the soliton width in the corresponding case for normal intensities. We can explain this fact by considering the existence curve of the dark solitons. We plot the

existence curve for the incoherently coupled dark soliton pair in Fig. 6.5. We can see that as ρ , the ratio of the total intensity I to the dark irradiance I_d increases, the FWHM of the soliton pair consistently decreases.

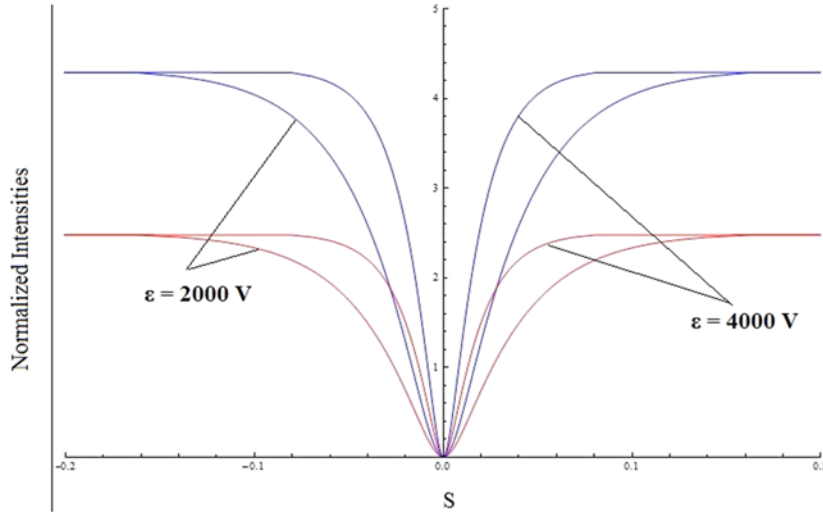


Fig. 6.3. Comparison of Normalized Intensities of the components $|U|^2$ (blue) and $|V|^2$ (red) of the incoherently coupled soliton pair when $\epsilon = 2000$ V and $\epsilon = 4000$ V with $R = 0$, $\rho = 5$, $\theta = \pi/3$. [20]

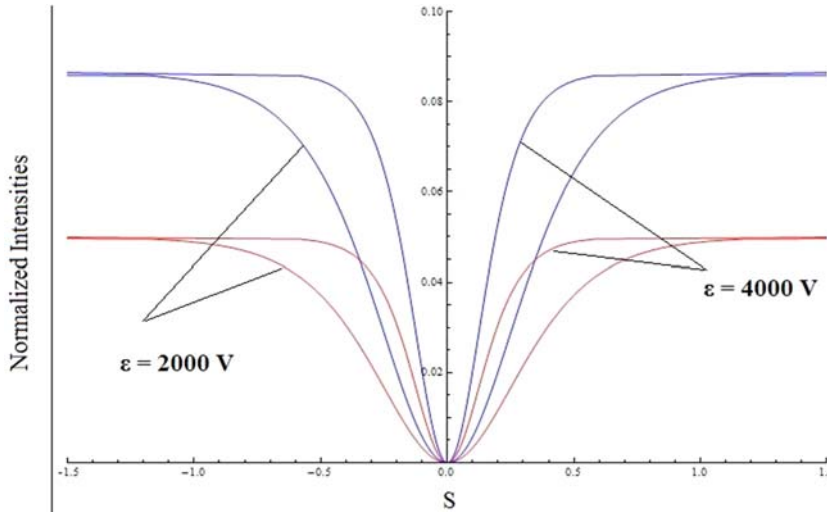


Fig. 6.4. Comparison of Normalized Intensities of the components $|U|^2$ (blue) and $|V|^2$ (red) of the incoherently coupled soliton pair when $\epsilon = 2000$ V and $\epsilon = 4000$ V, $\rho = 0.1$, $\theta = \pi/3$. [20]

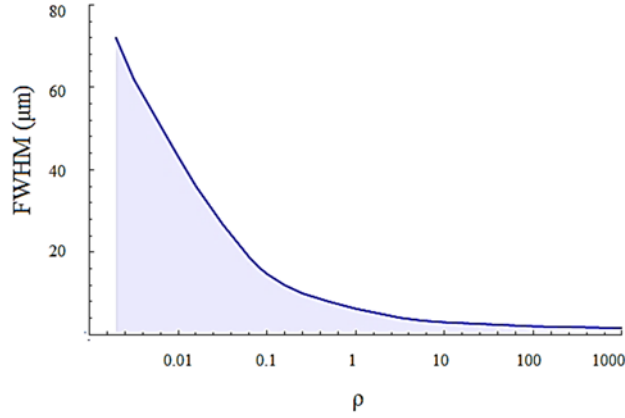


Fig. 6.5. Existence curve for incoherently coupled dark soliton pair when $\varepsilon = 4000$ V. [20]

6.3.1.2. Bright-Bright Soliton Pair

In case of a bright-bright incoherently coupled soliton pair, we express the solutions as,

$$U = r^{1/2} y(s) \cos(\theta) \exp(i\mu\xi), \quad (6.49)$$

$$V = r^{1/2} y(s) \sin(\theta) \exp(i\mu\xi), \quad (6.50)$$

where μ is the nonlinear shift of the propagation constant, θ is the arbitrary projection angle, and $y(s)$ is the normalized bounded function which satisfies $0 \leq y(s) \leq 1$ and $y(\pm\infty) = 0, \dot{y}(0) = 0, \dot{y}(\pm\infty) = 0, y(0) = 1, \ddot{y}(\pm\infty) = 0$,

Substituting (6.49) and (6.50) in (6.10) and (6.11), we get a single equation,

$$\ddot{y} = 2\mu y + \frac{2\beta}{(1 + ry^2)^2} y. \quad (6.51)$$

Integrating (6.51) once, we get,

$$\dot{y}^2 = 2\mu y^2 - \frac{2\beta}{r} \left(\frac{1}{1 + ry^2} \right) + c, \quad (6.52)$$

where c is the constant of integration.

Solving (6.52) by putting $y = y(s \rightarrow \infty)$ and using the boundary conditions $y(\pm\infty) = 0$, $\dot{y}(\pm\infty) = 0$, we get,

$$c = \frac{2\beta}{r}. \quad (6.53)$$

Now, putting $y = y(s = 0)$ and using the boundary conditions $y(0) = 1$, $\dot{y}(0) = 0$, we get,

$$\mu = -\frac{\beta}{1+r}. \quad (6.54)$$

Again, substituting the value of μ and c in (6.52), we get,

$$\dot{y}^2 = \frac{-2\beta(y^2)}{1+r} - \frac{2\beta}{r} \left(\frac{1}{1+ry^2} \right) + \frac{2\beta}{r}. \quad (6.55)$$

We can see, in (6.52) that since y is bounded between 0 and 1, so β has to be greater than zero if the RHS is to remain positive. Hence we get the condition that $g_{eff} > 0$ for a bright soliton pair. This is consistent with the condition for formation of single bright solitons as found in [36].

We can obtain the soliton envelope from (6.55) by integrating once more,

$$\left(\frac{2\beta}{r}\right)^{1/2} s = \pm \int_y^1 \left(1 - \frac{r\tilde{y}^2}{1+r} - \frac{1}{1+r\tilde{y}^2}\right)^{-1/2} d\tilde{y}. \quad (6.56)$$

As before, we assume two cases of $\varepsilon = 2000$ V and 4000 V with $l = 7.5$ mm.

To illustrate our result, we consider crystal parameters of KLTN crystal. We have, $n_e = 2.2$, $g_{eff} = 0.12 \text{ m}^4 \text{ C}^{-2}$, $T = 21^\circ \text{C}$, $\varepsilon_r = 8000$, $\lambda_0 = 0.5 \text{ } \mu\text{m}$, $x_0 = 40 \text{ } \mu\text{m}$ [57].

Fig. 6.6 shows the intensity profiles of the incoherently coupled bright soliton pair, for the cases of $\varepsilon = 2000$ V and 4000 V respectively. We find the FWHM of the two components to be $9.3 \text{ } \mu\text{m}$. in the first case when $\varepsilon = 2000$ V. This decreases to about $4.9 \text{ } \mu\text{m}$ when $\varepsilon = 4000$ V.

Now, if we take the external resistance as zero and also the internal resistance of the crystal is negligible, then $R = 0$ for this case. The normalized intensities are plotted in Fig. 6.7 for the case of $R = 0$. The FWHM of the pairs are $6.7 \text{ } \mu\text{m}$, $3.4 \text{ } \mu\text{m}$. for $\varepsilon = 2000$, 4000 V respectively. These are much lesser when we compare with the corresponding values of FWHM for the $R \neq 0$ case discussed previously which implies an enhancement in the self trapping as the internal and external resistance decreases.

Now, in the case of low intensity, we plot the intensity profiles of the soliton pair components U and V in Fig. 6.8 for $\varepsilon = 2000$ V and 4000 V when the peak $|U|^2 \ll 1 \approx 0.025$ and peak $|V|^2 \ll 1 \approx 0.043$. We find that the FWHM for these intensities of the coupled pair is $26.6 \text{ } \mu\text{m}$, $46.4 \text{ } \mu\text{m}$ for $\varepsilon = 2000$ V and 4000 V. This is larger than the soliton width in the corresponding case for normal intensities. This can be explained by the existence curve for the incoherently coupled bright soliton pair which we plot in Fig. 6.9. As we can see from the Fig. 6.9, the FWHM increases for both very low and very high values of intensity, and reaches a minimum at about $r = 1$.

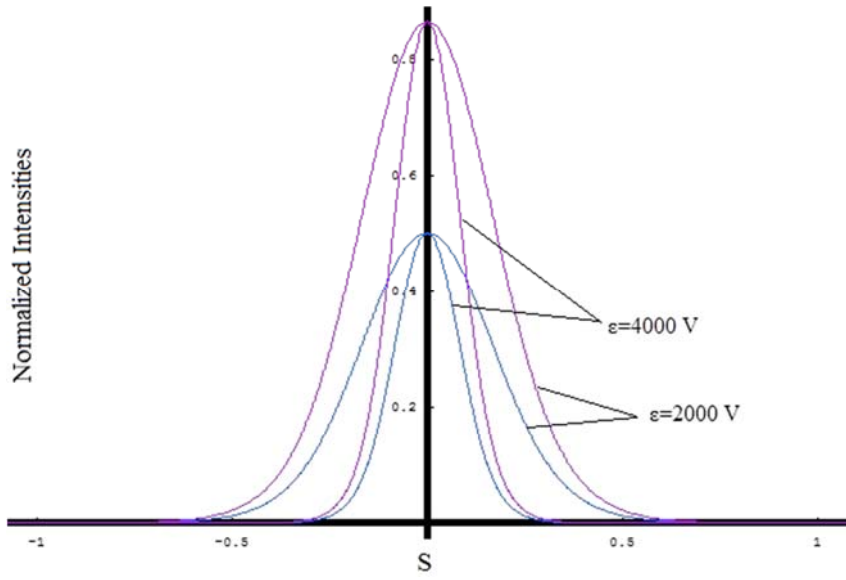


Fig. 6.6. Comparison of Normalized Intensities of the components $|U|^2$ (red) and $|V|^2$ (blue) of the incoherently coupled soliton pair when $\varepsilon = 2000$ V and $\varepsilon = 4000$ V with, $r = 1$, $\theta = \pi/3$. [20]

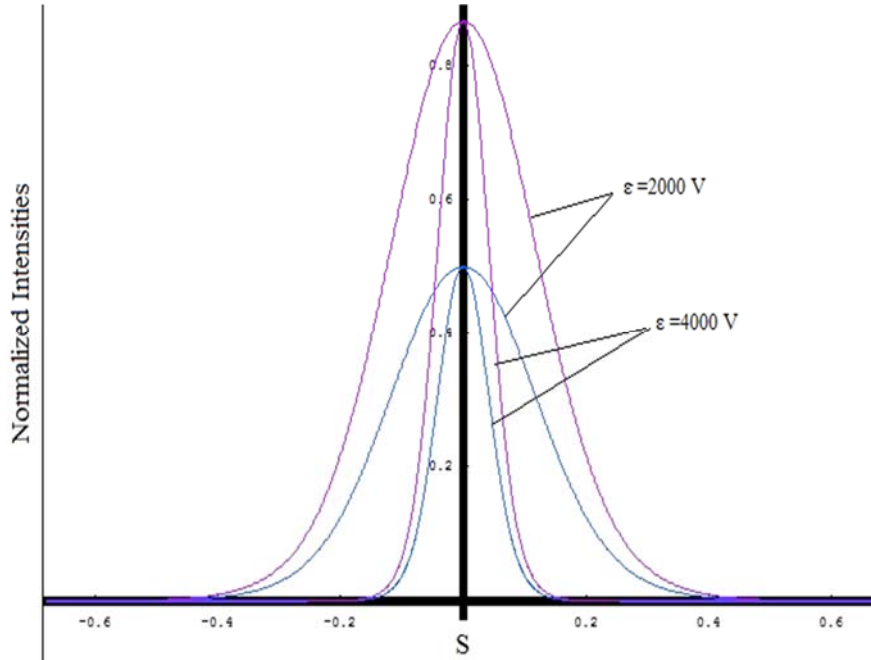


Fig. 6.7. Comparison of Normalized Intensities of the components $|U|^2$ (red) and $|V|^2$ (blue) of the incoherently coupled soliton pair when $\varepsilon = 2000$ V and $\varepsilon = 4000$ V with, $r = 1$, $\theta = \pi/3$, $R = 0$. [20]

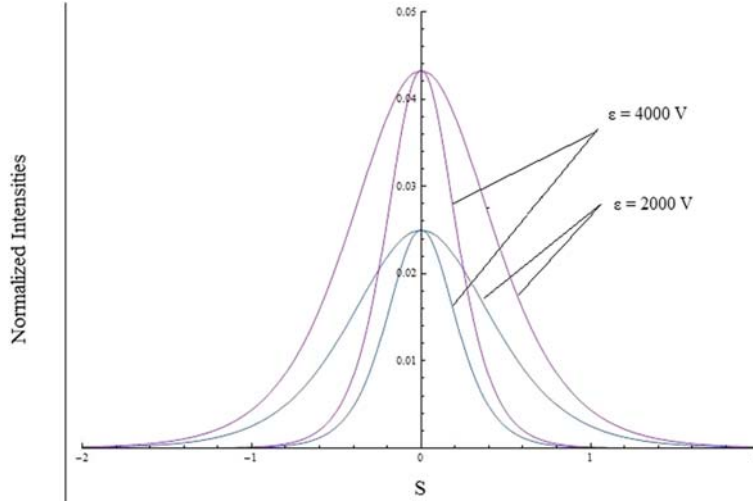


Fig. 6.8. Comparison of Normalized Intensities of the components $|U|^2$ (red) and $|V|^2$ (blue) of the incoherently coupled soliton pair when $\varepsilon = 2000$ V and $\varepsilon = 4000$ V with, $r = 0.05$, $\theta = \pi/3$. [20]

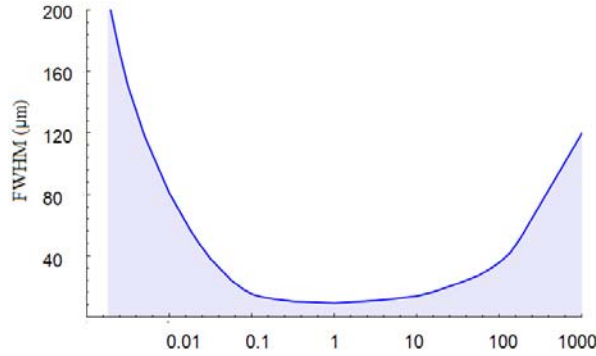


Fig. 6.9. Existence curve for incoherently coupled bright soliton pair when $\varepsilon = 4000$ V. [20]

6.3.1.3. Grey-Grey Soliton Pair

We express the solutions as

$$U = \rho^{1/2} y(s) \cos(\theta) \exp\left[i\left(c\xi + \int_0^s \frac{\Gamma d\tilde{s}}{y^2(\tilde{s})}\right)\right], \quad (6.57)$$

$$V = \rho^{1/2} y(s) \sin(\theta) \exp\left[i\left(c\xi + \int_0^s \frac{\Gamma d\tilde{s}}{y^2(\tilde{s})}\right)\right]. \quad (6.58)$$

Substituting (6.57) and (6.58) in (6.10) and (6.11),

$$\ddot{y} = 2cy + \frac{\Gamma^2}{y^3} + \frac{2\beta(1+\rho)^2}{(1+\rho y^2)^2} y. \quad (6.59)$$

Here, we have the boundary conditions, $y(s \rightarrow \pm\infty) = 1$, $\dot{y}(0) = 0$, $y^2(s=0) = m$, $\ddot{y}(s \rightarrow \pm\infty) = 0$.

Using the boundary conditions $y(s \rightarrow \pm\infty) = 1$, $\ddot{y}(s \rightarrow \pm\infty) = 0$ in (6.59), we can find the value of Γ ,

$$\Gamma^2 = -2c - 2\beta. \quad (6.60)$$

Now, integrating (6.52) once, we get

$$\dot{y}^2 = 2c(y^2) - \Gamma^2 \left(\frac{1}{y^2} \right) - \frac{2\beta(1+\rho)^2}{\rho} \left(\frac{1}{1+\rho y^2} \right) + d, \quad (6.61)$$

where d is the constant of integration.

Let $y = y(0)$, in (6.61), and using $\dot{y}(0) = 0$, $y(0)^2 = m$,

$$2cm - \frac{\Gamma^2}{m} - 2\beta \frac{(1+\rho)^2}{\rho} \left(\frac{1}{1+\rho m} \right) + d = 0, \quad (6.62)$$

where d is the constant of integration.

Let $y = y(s \rightarrow \infty)$, in (6.61), and using $\dot{y}(\infty) = 0$, $y(s \rightarrow \pm\infty)^2 = 1$,

$$2c - \Gamma^2 - 2\beta \frac{(1+\rho)^2}{\rho} + d = 0. \quad (6.63)$$

Solving (6.62) and (6.63), we can find out the values of c and d ,

$$c = \frac{-\Gamma^2}{m} - \frac{2\beta(1+\rho)}{(1+\rho m)}, \quad (6.64)$$

$$d = \frac{\Gamma^2(1+m)}{m} + 2\beta(1+\rho) \left(\frac{(1+\rho) - m(1+\rho m)}{\rho(1+\rho m)(1-m)} \right). \quad (6.65)$$

And, we can finally find the soliton envelope by integrating (6.61) once again and substituting the values of c and d ,

$$s = \pm \int_y^{\sqrt{m}} \left(2c(\tilde{y}^2) + 4\Gamma^2 \left(\frac{1}{\tilde{y}^2} \right) - \frac{2\beta(1+\rho)^2}{\rho} \left(\frac{1}{1+\rho\tilde{y}^2} \right) + d \right)^{-1/2} d\tilde{y}. \quad (6.66)$$

As in the case of dark solitons, we need $\beta < 0$. We take the parameters similar to the case of the dark solitons.

Fig. 6.10 shows the intensity profiles of the incoherently coupled grey soliton pair, for the cases of $\varepsilon = 2000$ V and 4000 V respectively with $l = 7.5$ mm. We find the FWHM of the two components to be $9.77 \mu\text{m}$ the first case when $\varepsilon = 2000$ V. This decreases to about $5.6 \mu\text{m}$ when $\varepsilon = 4000$ V.

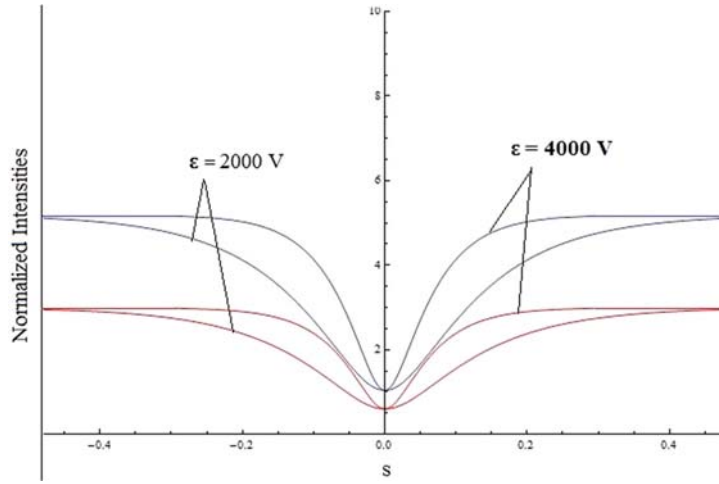


Fig-6.10. Comparison of Normalized Intensities of the components $|U|^2$ (blue) and $|V|^2$ (red) of the incoherently coupled soliton pair when $\varepsilon = 2000$ V and $\varepsilon = 4000$ V with, $\rho = 6$, $m = 0.2$, $\theta = \pi/6$. [20]

Now, we consider the $R \rightarrow 0$ case. The normalized intensities are plotted in Fig. 6.11 for this case. The FWHM of the pairs are $4.70 \mu\text{m}$, $2.87 \mu\text{m}$. for $\varepsilon = 2000$, 4000 V respectively. These are much lesser when we compare with the corresponding values of FWHM for the $R \neq 0$ case discussed previously. Again, this is due to the enhancement of the self trapping as the internal and external resistance decreases.

Now, to consider the low amplitude case, we take, $|U|^2 \ll 1 \approx 0.08$ and $|V|^2 \ll 1 \approx 0.05$.

The intensities of the soliton pair are shown in Fig. 6.12. The FWHM of the soliton pair is $26.6 \mu\text{m}$, $14.5 \mu\text{m}$ for $\varepsilon = 2000$ V and 4000 V. These are larger than the soliton width in the corresponding case for normal intensities. This fact can be explained based on the existence curve for grey-grey soliton pair which is plotted in Fig. 6.13. As we can see from the Fig. 6.13, the FWHM increases for, both low and high values of ρ and reaches a minimum between $0 < \rho < 10$.

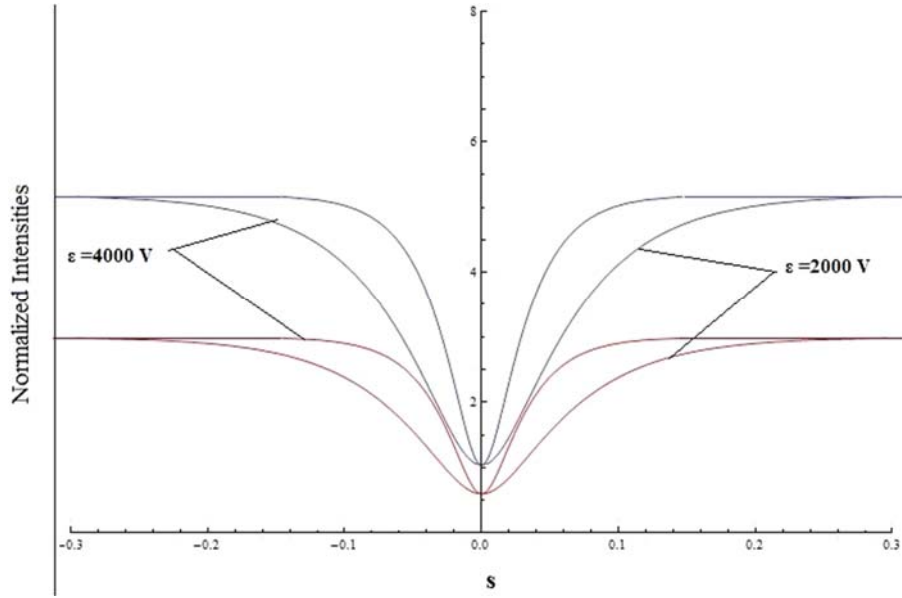


Fig. 6.11. Comparison of Normalized Intensities of the components $|U|^2$ (blue) and $|V|^2$ (red) of the incoherently coupled soliton pair when $\varepsilon = 2000$ V and $\varepsilon = 4000$ V with, $\rho = 6$, $m = 0.2$, $\theta = \pi/6$, $R = 0$. [20]

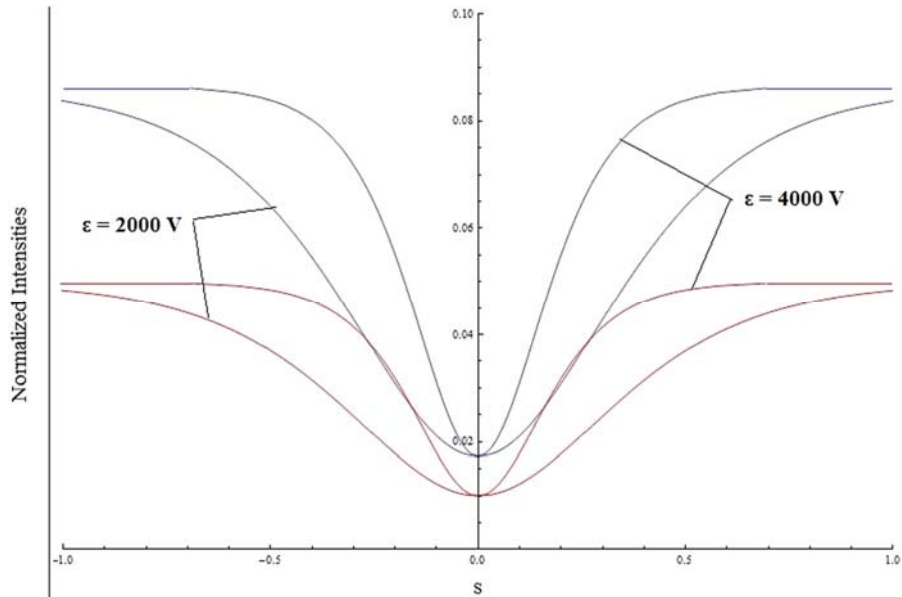


Fig. 6.12. Comparison of Normalized Intensities of the components $|U|^2$ (blue) and $|V|^2$ (red) of the incoherently coupled soliton pair when $\varepsilon = 2000$ V and $\varepsilon = 4000$ V with, $\rho = 0.1$, $m = 0.2$, $\theta = \pi/6$. [20]

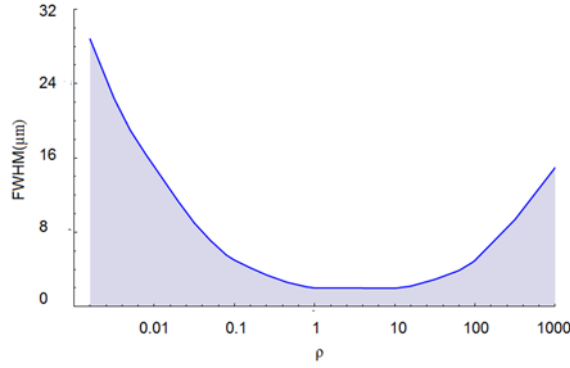


Fig. 6.13. Existence Curve for incoherently coupled grey-grey soliton pair when $\varepsilon = 4000$ V and $m = 0.2$. [20]

6.3.1.4.2N-Component Dark Solitons

Now, to study the incoherently coupled multicomponent solitons, we express the soliton solutions as

$$\sum_{j=1}^N U_j = \sum_{j=1}^N \frac{1}{\sqrt{N}} \rho^{1/2} y(s) \cos(j\theta) \exp(i\mu\xi), \quad (6.67)$$

$$\sum_{m=1}^N V_m = \sum_{m=1}^N \frac{1}{\sqrt{N}} \rho^{1/2} y(s) \sin(m\theta) \exp(i\mu\xi). \quad (6.68)$$

We emphasize that the set (N, θ) has to be selected carefully so that the optical beams

$\sum_{j=1}^N U_j$ and $\sum_{m=1}^N V_m$ include $2N$ components of the different beam intensities.

We substitute these forms of $\sum_{j=1}^N U_j$ and $\sum_{m=1}^N V_m$ into equations (6.18) and (6.19), to get,

$$\frac{1}{2} \ddot{y} - \mu y - \frac{\beta(1+\rho)^2}{(1+\rho y^2)^2} y = 0. \quad (6.69)$$

This equation (6.69) has been discussed before in case of the incoherently coupled dark soliton pair (Eq. (6.43)). We obtain the composite dark soliton components through a θ projection. Since we have obtained the same equation as that of the relevant soliton pair, our previous results of the dark-dark incoherently coupled soliton pair carry forward to the multicomponent solitons also. As an illustration, these are shown in Fig. 6.14 for $N = 2$, $\theta = \pi/7$. As mentioned above, with a judicious selection of N and θ , one can find out $2N$ components for any N .

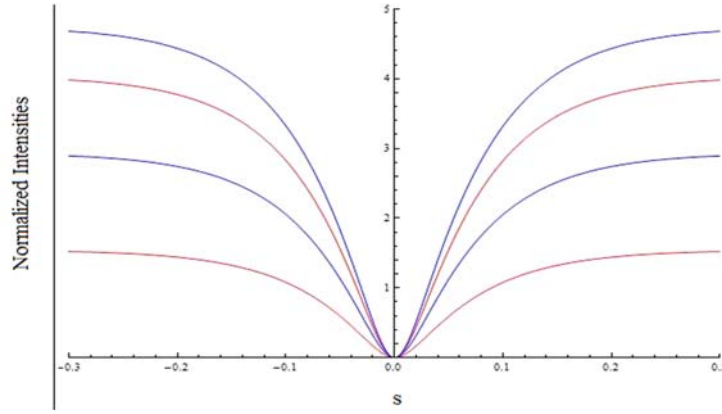


Fig. 6.14. Normalized Intensities of the components $\sum |U_j|^2$ (blue) and $\sum |V_m|^2$ (red) of the incoherently coupled multi-component solitons when $\varepsilon = 2500$ V with, $\rho = 5$, $\theta = \pi/5$. [20]

6.3.1.5.2N-Component Bright Solitons

Now, to study the incoherently coupled multicomponent solitons, we express the soliton solutions as

$$\sum_{j=1}^N U_j = \sum_{j=1}^N \frac{1}{\sqrt{N}} r^{1/2} y(s) \cos(j\theta) \exp(i\mu\xi), \quad (6.70)$$

$$\sum_{m=1}^N V_m = \sum_{m=1}^N \frac{1}{\sqrt{N}} r^{1/2} y(s) \sin(m\theta) \exp(i\mu\xi). \quad (6.71)$$

We emphasize that the set (N, θ) has to be selected carefully so that the optical beams $\sum_{j=1}^N U_j$ and $\sum_{m=1}^N V_m$ include $2N$ components of the different beam intensities.

We substitute these forms of $\sum_{j=1}^N U_j$ and $\sum_{m=1}^N V_m$ into Eqs. (6.18) and (6.19), we get

$$\ddot{y} = 2\mu y + \frac{2\beta}{(1 + ry^2)^2} y. \quad (6.72)$$

This Eq. (6.72) has been discussed before in case of the incoherently coupled bright soliton pair (Eq. (6.51)). We obtain the composite bright soliton components through a θ projection. Since we have obtained the same equation as that of the relevant soliton pair, our previous results of the bright-bright incoherently coupled soliton pair carry forward to the multicomponent solitons also.

As an illustration, we show the normalized intensities of the multicomponent bright solitons for the case of $N = 2$, i. e. 4 components and for $\theta = \pi/7$ in Fig. 6.15. As mentioned above, with a judicious selection of N and θ , one can find out $2N$ components for any N .

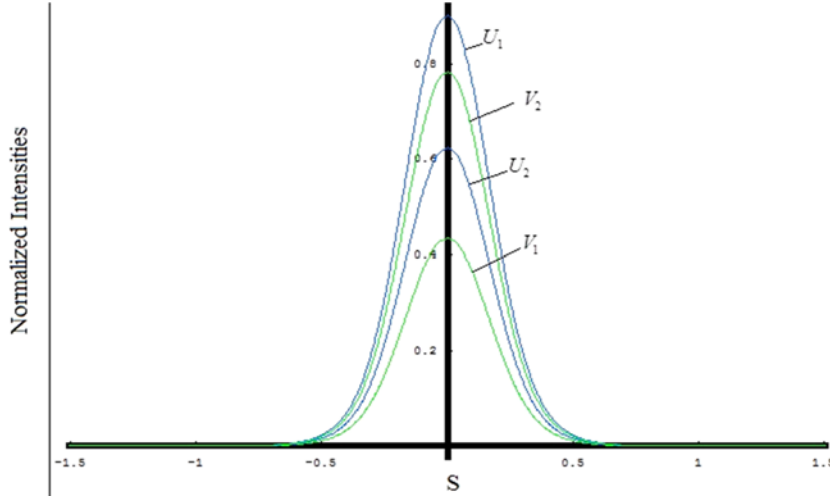


Fig. 6.15. Normalized Intensities of the components $\sum |U_j|^2$ (blue) and $\sum |V_m|^2$ (green) of the incoherently coupled multi-component solitons when $\varepsilon = 2000$ V with, $r = 1$, $\theta = \pi/7$. [20]

6.3.1.6. N Component Grey Solitons

Now, to study the incoherently coupled multicomponent solitons, we express the soliton solutions as

$$U_j \left\{ \begin{aligned} &= \frac{1}{\sqrt{N/2}} \rho^{1/2} y(s) \cos(j\theta) \exp[i(\delta\xi + \int_0^s \frac{\Gamma d\tilde{s}}{y^2(\tilde{s})})], j \leq N/2 \\ &= \frac{1}{\sqrt{N/2}} \rho^{1/2} y(s) \sin((N+1) - j\theta) \exp[i(\delta\xi + \int_0^s \frac{\Gamma d\tilde{s}}{y^2(\tilde{s})})], j > N/2 \end{aligned} \right\}, \quad (6.73)$$

($j = 1, 2, \dots, N$; $N = \text{even numbers}$)

$$U_j \left\{ \begin{aligned} &= \frac{1}{\sqrt{(N+1)/2}} \rho^{1/2} y(s) \cos((j-1)\theta) \exp[i(\delta\xi + \int_0^s \frac{\Gamma d\tilde{s}}{y^2(\tilde{s})})], j \leq (N+1)/2 \\ &= \frac{1}{\sqrt{(N+1)/2}} \rho^{1/2} y(s) \sin((N+1) - j\theta) \exp[i(\delta\xi + \int_0^s \frac{\Gamma d\tilde{s}}{y^2(\tilde{s})})], j > (N+1)/2 \end{aligned} \right\}, \quad (6.74)$$

($j = 1, 2, \dots, N$; $N = \text{odd numbers}$), where we modify Eq. (6.10) as,

$$i \frac{\partial U_j}{\partial \xi} + \frac{1}{2} \frac{\partial^2 U_j}{\partial s^2} - \frac{\beta(1+\rho)^2 U_j}{(1 + \sum_{j=1}^N |U_j|^2)^2} = 0. \quad (6.75)$$

This is nothing but a reorganization of Eq. (6.10) assimilating the components of the two beams U and V into a single beam U 's components.

We emphasize that the set (N, θ) has to be selected carefully so that the optical beams $\sum_{n=1}^N U_j$ include N components of the different beam intensities.

Now, substitute these forms (6.73) and (6.74) of $\sum_{n=1}^N U_j$ into Eq. (6.75),

$$\ddot{y} = 2cy + \frac{\Gamma^2}{y^3} + \frac{2\beta(1+\rho)^2}{(1+\rho y^2)^2} y. \quad (6.76)$$

Since we have obtained the same equation as that of the relevant soliton pair (Eq. (6.59)), our previous results of the grey-grey incoherently coupled soliton pair carry forward to the multicomponent solitons also. We obtain the composite grey soliton components through a θ projection.

As an illustration, we show the normalized intensities of the multicomponent grey solitons for the case of $N = 3$, i. e. 3 components and for $\theta = \pi/3$ in Fig. 6.16. As mentioned above, with a judicious selection of N and θ , one can find out N components for any N .

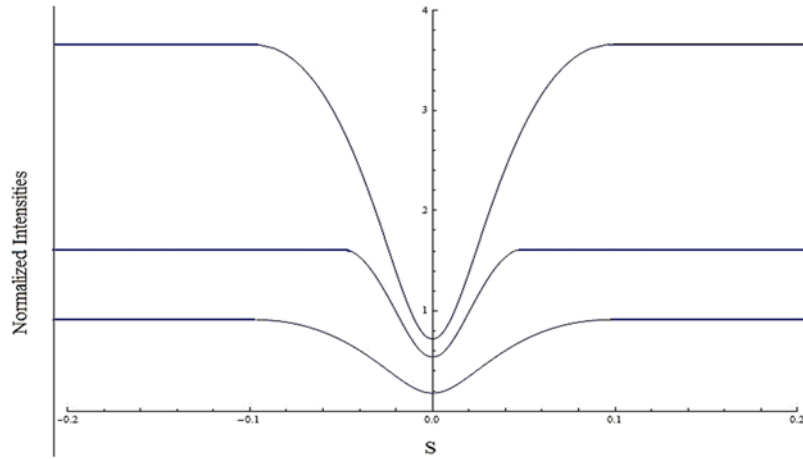


Fig. 6.16. Normalized Intensities of the components $\sum |U_j|^2$ (blue) of the incoherently coupled multi-component solitons when $\varepsilon = 2000$ V with, $\rho = 3.6$, $m = 0.2$, $\theta = \pi/3$ [20].

6.3.1.7. Modulation Instability

From (6.32)-(6.33), we can see that the maximum MI gain is dependent on the value of $E_0 = E(x \rightarrow \pm\infty)$, r' and ρ .

In case of a bright-bright soliton pair, $\rho = 0$ and so the MI gain is dependent on $r' (= r)$ and E_0 . As the MI gain varies as $\sim E_0^2$, the modulation instability gain will be more for the case of $\varepsilon = 4000$ V than for $\varepsilon = 2000$ V considered in our investigation. Also, the MI gain varies as order of $\sim \frac{1}{r^2}$, if we consider the dependence on r . Hence, in the low intensity limit, in which we consider $|U|^2, |V|^2 \ll 1$, we conclude that the MI gain will be increased significantly.

For studying the MI in case of grey-grey soliton pairs, $r' < \rho$ and hence, as the uniform background intensity increases, the modulation instability becomes larger. Also, since the MI is dependent on the value of E_0 , hence it will be lesser for $\varepsilon = 2000$ V than for $\varepsilon = 4000$ V.

In the case of dark soliton pair, we can study the modulation instability of the uniform background on which the dark notch is embedded. Hence, $r' = \rho$, and the MI gain g varies as $\sim \frac{\rho}{1 + \rho}$. Hence, in the low intensity limit, the modulation instability is lesser than that at normal intensities.

6.3.2. Coupled Solitons in Pyroelectric Photorefractive Materials

6.3.2.1. Bright Soliton Pair

In case of a bright-bright incoherently coupled soliton pair, we express the solutions as,

$$U = r^{1/2} y(s) \cos(\theta) \exp(iv\xi), \quad (6.77)$$

$$V = r^{1/2} y(s) \sin(\theta) \exp(iv\xi), \quad (6.78)$$

where $r = \frac{I_0}{I_d}$ which stands for the ratio of the maximum intensity to the dark irradiance,

v is the nonlinear shift of the propagation constant, $y(s)$ is the normalized bounded function which satisfies $0 \leq y(s) \leq 1$ and $y(\pm\infty) = 0$, $\dot{y}(0) = 0$, $\dot{y}(\pm\infty) = 0$, $y(0) = 1$, $\dot{y}(\pm\infty) = 0$ and θ is the arbitrary projection angle.

Substituting (6.77) and (6.78) in (6.39) and (6.40), we get a single equation,

$$\frac{d^2 y}{ds^2} - 2\nu y + 2\beta \frac{ry^3}{1+ry^2} = 0. \quad (6.79)$$

Integrating (6.79) once,

$$\frac{1}{2} \left(\frac{dy}{ds} \right)^2 = \nu y^2 - \beta y^2 + \frac{\beta}{r} \log(1+ry^2) = 0. \quad (6.80)$$

Using the boundary conditions of bright solitons, we have,

$$\nu = \beta - \frac{\beta}{r} \log(1+r). \quad (6.81)$$

And we get, substituting the value in (6.80),

$$\frac{1}{2} \left(\frac{dy}{ds} \right)^2 = -\frac{\beta}{r} y^2 \log(1+r) + \frac{\beta}{r} \log(1+ry^2). \quad (6.82)$$

Integrating once more, we get,

$$s = \pm \int_y^1 \left[\frac{2\beta}{r} \{ \log(1+r\tilde{y}^2) - \tilde{y}^2 \ln(1+r) \} \right]^{-1/2} d\tilde{y}. \quad (6.83)$$

The envelope $y(s)$ can be found out from (6.83) by straightforward numerical integration. Further, the components of the soliton pair can be found by substituting $y(s)$ in (6.77) and (6.78) by means of a θ -projection. Since $y(s)$ is bounded between 0 and 1, we can see from (6.80) that $\beta > 0$ for keeping the RHS positive. Hence, the condition for bright soliton pair formation is that $\beta > 0$. This means that if the pyroelectric coefficient is negative, then change in temperature $\Delta T > 0$ and if the pyroelectric coefficient is positive, then change in temperature $\Delta T < 0$. We shall consider the SBN crystal for illustration. This is a non-photovoltaic photorefractive crystal with a negative pyroelectric coefficient. Hence, we take the following parameters [45, 46], $n_e = 2.35$, $\lambda_0 = 532 \text{ nm}$, $x_0 = 20 \text{ }\mu\text{m}$, $g_{\text{eff}} = 237 \times 10^{-12} \text{ m/V}$, $\varepsilon_0 = 8.85 \times 10^{-12} \text{ F/m}$, $\varepsilon_r = 3400$, $\frac{\partial P}{\partial T} = -3 \times 10^{-4}$, $r = 10$. For the $\Delta T = 10^\circ\text{C}$, 20°C ; we can arrive at the values for $\beta = 20.1$, 40.2 respectively.

Fig. 6.17 shows the normalized intensities of the bright soliton pair.

For studying the global properties of these bright soliton pairs, we plot the existence curve for the incoherently coupled bright solitons in Fig. 6.18.

As we see, in Fig. 6.18, we have plotted the graph on a log-linear scale. The soliton FWHM (full width at half maximum) is a marker of the self trapping. As the value of r

increases, the FWHM decreases till about $r = 1$ to $r = 10$. For $r > 10$, the FWHM again increases non-linearly with an increase in r .

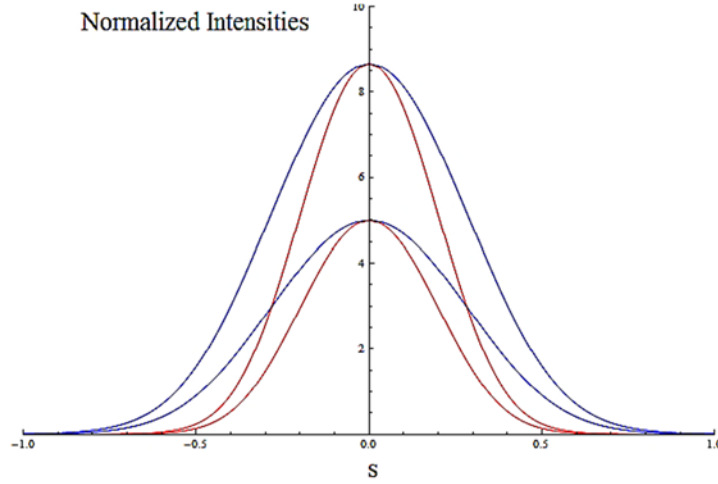


Fig. 6.17. Normalized Intensities of the bright incoherently coupled soliton pair under different temperature changes $\Delta T = 10^0$ C (Blue), $\Delta T = 20^0$ C (Red) for $r = 10$, $\theta = \pi/3$. [22]

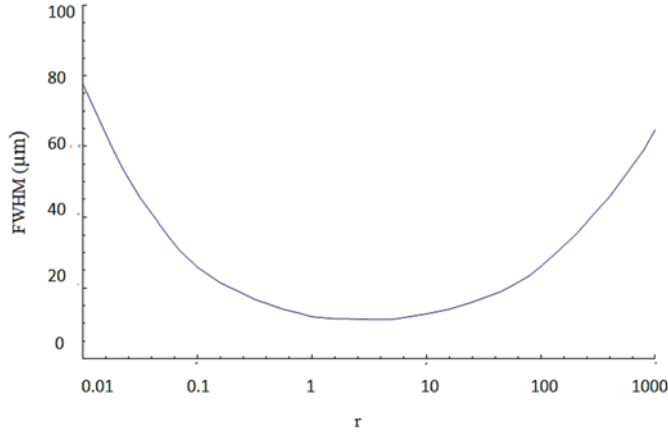


Fig. 6.18. Existence curve for the incoherently coupled bright pyroelectric soliton pair when $\Delta T = 10^0$ C. [22]

6.3.2.2. Dark Soliton Pair

In case of a dark-dark incoherently coupled soliton pair, we express the solutions as,

$$U = \rho^{1/2} y(s) \cos(\theta) \exp(i\mu\xi), \quad (6.84)$$

$$V = \rho^{1/2} y(s) \sin(\theta) \exp(i\mu\xi), \quad (6.85)$$

where μ is the nonlinear shift of the propagation constant, θ is the arbitrary projection angle, and $y(s)$ is the normalized bounded function which satisfies $0 \leq y(s) \leq 1$ and $y(\pm\infty) = \pm 1, \dot{y}(\pm\infty) = 0, y(0) = 0, \ddot{y}(\pm\infty) = 0$.

Substituting (6.84) and (6.85) in (6.39) and (6.40), we get a single equation,

$$\frac{d^2 y}{ds^2} - 2\mu y + 2\beta \frac{\rho y^3}{1 + \rho y^2} = 0. \quad (6.86)$$

Using the boundary conditions of dark solitons, we have,

$$\mu = \frac{\beta \rho}{1 + \rho}. \quad (6.87)$$

Integrating (6.86) once,

$$\frac{1}{2} \left(\frac{dy}{ds} \right)^2 = \left(\frac{\beta \rho}{1 + \rho} \right) y^2 - \beta y^2 + \frac{\beta}{\rho} \log(1 + \rho y^2) + c, \quad (6.88)$$

where c is the constant of integration. Evaluating the constant of integration c by using the boundary conditions of the dark soliton,

$$c = \frac{\beta}{1 + \rho} - \frac{\beta}{\rho} \log(1 + \rho), \quad (6.89)$$

and we get, substituting the value of c in (6.88),

$$\left\langle \frac{1}{2} \left(\frac{dy}{ds} \right)^2 = \left(\frac{\beta \rho}{1 + \rho} \right) y^2 - \beta y^2 + \frac{\beta}{\rho} \log(1 + \rho y^2) + \frac{\beta}{1 + \rho} - \frac{\beta}{\rho} \log(1 + \rho) \right\rangle. \quad (6.90)$$

Integrating (6.90) once more,

$$s = \pm \int_y^0 \left[-2\beta \left\{ \frac{\tilde{y}^2 - 1}{1 + \rho} - \frac{1}{\rho} \ln \left(\frac{1 + \rho \tilde{y}^2}{1 + \rho} \right) \right\} \right]^{-1/2} d\tilde{y}. \quad (6.91)$$

The envelope $y(s)$ can be found out from (6.91) by straightforward numerical integration. Further, the components of the soliton pair can be found by substituting $y(s)$ in (6.84) and (6.85) by means of a θ -projection. Since $y(s)$ is bounded between 0 and 1, we can see from (6.90) that $\beta < 0$ for keeping the RHS positive. Hence, the condition for dark soliton pair formation is that $\beta < 0$. This means that if the pyroelectric coefficient is negative, then change in temperature $\Delta T < 0$ and if the pyroelectric coefficient is positive, then change in temperature $\Delta T > 0$. In the present illustration as mentioned in Section 6.3.1, since we consider the SBN crystal, for $\Delta T = -10^\circ \text{C}$, -20°C ; we can arrive at the values for

$\beta = -20.1, -40.2$ respectively. Since ΔT is negative, hence the dark soliton pair results from cooling the crystal between the specified temperature differences.

Fig. 6.19 shows the normalized intensities of the dark soliton pair.

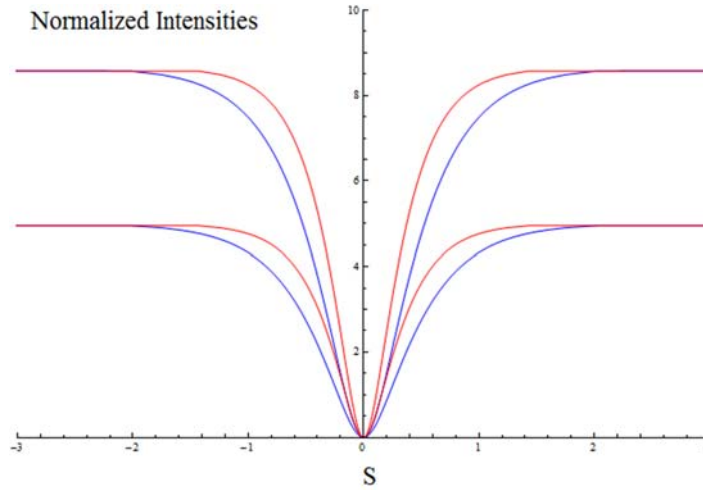


Fig. 6.19. Normalized Intensities of the dark incoherently coupled soliton pair under different temperature changes $\Delta T = -10^\circ \text{C}$ (Blue), $\Delta T = -20^\circ \text{C}$ (Red) for $\rho = 10$, $\theta = \pi/3$. [22]

For studying the global properties of the incoherently coupled dark soliton pair, we plot the existence curve in Fig. 6.20.

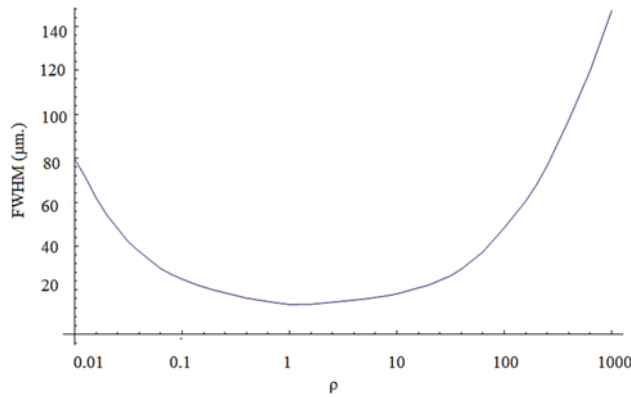


Fig. 6.20. Existence curve for the dark incoherently coupled pyroelectric soliton pair when $\Delta T = -10^\circ \text{C}$. [22]

From Fig. 6.20, we can see that the FWHM of the soliton pair decreases with an increase in ρ for low values of ρ till about $\rho \cong 1$ after which for greater values of ρ , the FWHM rapidly increases.

6.3.2.3. Grey Soliton Pair

We express the solutions as

$$U = \rho^{1/2} y(s) \cos(\theta) \exp[i(\mu \xi + \int_0^s \frac{\Gamma d\tilde{s}}{y^2(\tilde{s})})], \quad (6.92)$$

$$V = \rho^{1/2} y(s) \sin(\theta) \exp[i(\mu \xi + \int_0^s \frac{\Gamma d\tilde{s}}{y^2(\tilde{s})})]. \quad (6.93)$$

Substituting (6.92) and (6.93) in (6.39) and (6.40),

$$\ddot{y} = 2\mu y + \frac{\Gamma^2}{y^3} + 2\beta \frac{\rho y^3}{1 + \rho y^2} = 0. \quad (6.94)$$

Here, we have the boundary conditions $y(s \rightarrow \pm\infty) = 1$, $\dot{y}(0) = 0$, $y^2(s=0) = m$, $\ddot{y}(s \rightarrow \pm\infty) = 0$.

Using the boundary conditions, we get,

$$\Gamma^2 = 2\beta \frac{\rho}{1 + \rho} - 2\mu, \quad (6.95)$$

$$\mu = \frac{1}{2(m-1)^2} \left\{ (1-m) \left(\frac{2\beta\rho}{1+\rho} \right) + \frac{2m\beta}{\rho} \left[\rho(m-1) - \log\left(\frac{1+\rho m}{1+\rho} \right) \right] \right\}. \quad (6.96)$$

Now, integrating (6.94) once and substituting the value of Γ and μ , we get

$$\left(\frac{dy}{ds} \right)^2 = 2\mu(y^2 - 1) - \Gamma^2 \left(\frac{1}{y^2} - 1 \right) - \frac{2\beta}{\rho} \left(\rho(y^2 - 1) - \log\left(\frac{1+\rho y^2}{1+\rho} \right) \right). \quad (6.97)$$

And, we can finally find the envelope $y(s)$ by integrating (6.97) once again and substituting the values of μ and Γ ,

$$\left\langle \left\langle s = \pm \int_y^{\sqrt{m}} \left\{ 2\mu(\tilde{y}^2 - 1) - \Gamma^2 \left(\frac{1}{\tilde{y}^2} - 1 \right) - \frac{2\beta}{\rho} \left(\rho(\tilde{y}^2 - 1) - \log\left(\frac{1+\rho\tilde{y}^2}{1+\rho} \right) \right) \right\}^{-1/2} d\tilde{y} \right\rangle \right\rangle. \quad (6.98)$$

The envelope $y(s)$ can be found out from (6.98) by numerical integration. Further, the components of the soliton pair can be found by substituting $y(s)$ in (6.92) and (6.93) by means of a θ -projection. In case of an SBN crystal, as in the case of dark solitons, we need $\beta < 0$. For $\Delta T = -10^\circ\text{C}$, -20°C ; we can arrive at the values for $\beta = -20.1$, -40.2 respectively. Since ΔT is negative, hence the grey soliton pair results from cooling the crystal between the specified temperature differences.

Fig. 6.21 shows the normalized intensities of the grey coupled soliton pair.

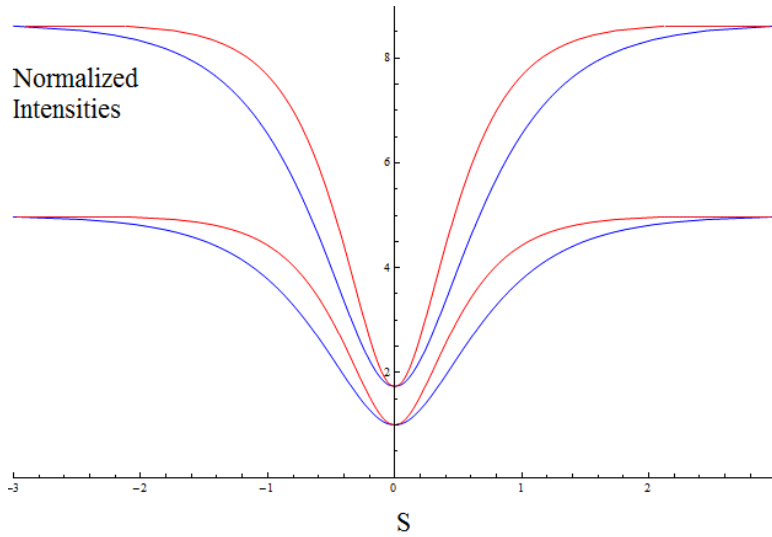


Fig. 6.21. Normalized Intensities of the dark incoherently coupled soliton pair under different temperature changes $\Delta T = -10^\circ\text{C}$ (Blue), $\Delta T = -20^\circ\text{C}$ (Red) for $\rho = 10$, $m = 0.2$, $\theta = \pi/3$. [22]

For studying the global properties of the incoherently coupled grey soliton pair, we plot the existence curve in Fig. 6.22.

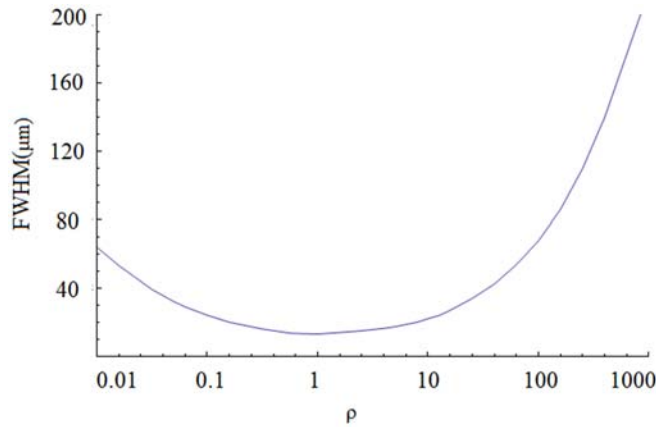


Fig. 6.22. Existence curve for the incoherently coupled grey pyroelectric soliton pair. $m = 0.2$, $\Delta T = -10^\circ\text{C}$. [22]

From Fig. 6.22, we see that the FWHM of the soliton pair decreases as ρ increases till $\rho \sim 1$ after which the FWHM of the soliton pair rapidly increases with increasing ρ .

6.3.3. Bright-Dark Soliton Pair

We express the two wave envelopes as,

$$U = r^{1/2} f(s) \exp(i\mu\xi), \quad (6.99)$$

$$V = \rho^{1/2} g(s) \exp(i\nu\xi), \quad (6.100)$$

where μ and ν are the shift in the propagation constants of the bright and dark soliton beams. $f(s)$ is the bright beam and satisfies the boundary conditions $f(0)=1, \dot{f}(0)=0, f(s \rightarrow \pm\infty)=0$. $g(s)$ is the dark beam and satisfies the boundary conditions $g(0)=0, \dot{g}(0)=0, g(s \rightarrow \pm\infty)=\pm 1$. Substituting the expressions for U and V in (6.39) and (6.40), we have,

$$\frac{d^2 f}{ds^2} = 2 \left\{ \mu - \beta + \frac{\beta}{1 + rf^2 + \rho g^2} \right\} f, \quad (6.101)$$

$$\frac{d^2 g}{ds^2} = 2 \left\{ \nu - \beta + \frac{\beta}{1 + rf^2 + \rho g^2} \right\} g. \quad (6.102)$$

We put the condition that $f^2 + g^2 = 1$. Thus, we get,

$$\frac{d^2 f}{ds^2} = 2 \left\{ \mu - \beta + \frac{\beta}{1 + \rho} \left(\frac{1}{1 + \sigma f^2} \right) \right\} f, \quad (6.103)$$

$$\frac{d^2 g}{ds^2} = 2 \left\{ \nu - \beta + \frac{\beta}{1 + \rho} \left(\frac{1}{1 + \sigma(1 - g^2)} \right) \right\} g, \quad (6.104)$$

where we have taken,

$$\sigma = (r - \rho) / (1 + \rho). \quad (6.105)$$

Integrating (6.103) once,

$$\left(\frac{df}{ds} \right)^2 = 2(\mu - \beta)(f^2 - 1) + \frac{2\beta}{\sigma(1 + \rho)} \log \left(\frac{1 + \sigma f^2}{1 + \sigma} \right). \quad (6.106)$$

Using the boundary conditions for $f(s)$ and $g(s)$ in (6.106) and (6.104) respectively,

$$\mu = \beta - \frac{\beta}{\sigma(1 + \rho)} \log(1 + \sigma), \quad (6.107)$$

and

$$\nu = \frac{\beta\rho}{(1+\rho)}. \quad (6.108)$$

If $|\sigma| \ll 1$, then the peak intensities of the two beams are nearly equal. In this limit of $|\sigma| \ll 1$, μ can be expanded by Taylor series and neglecting higher order terms,

$$\mu \approx \beta - \frac{\beta}{1+\rho} \left(1 - \frac{\sigma}{2}\right). \quad (6.109)$$

Due to $|\sigma| \ll 1$, and substituting (6.108)-(6.109) in (6.101)-(6.102), we have,

$$\frac{d^2 f}{ds^2} = \frac{\beta\sigma}{1+\rho} (f - 2f^3), \quad (6.110)$$

$$\frac{d^2 g}{ds^2} = -\frac{2\beta\sigma}{1+\rho} (g - 2g^3) \quad (6.111)$$

Eqs. (6.110)-(6.111) have a closed form solution,

$$f(s) = \text{sech} \left[\left\{ \frac{\beta\sigma}{1+\rho} \right\}^{1/2} s \right], \quad (6.112)$$

$$g(s) = \tanh \left[\left\{ \frac{\beta\sigma}{1+\rho} \right\}^{1/2} s \right]. \quad (6.113)$$

We can obtain the components of the soliton pair by substituting (6.112)-(6.113) in (6.99)-(6.100),

$$U(s, \xi) = r^{1/2} \text{sech} \left[\left\{ \frac{\beta\sigma}{1+\rho} \right\}^{1/2} s \right] \exp \left\{ i\beta \left(1 - \frac{1}{1+\rho} \left(1 - \frac{\sigma}{2} \right) \right) \xi \right\}, \quad (6.114)$$

$$V(s, \xi) = \rho^{1/2} \tanh \left[\left\{ \frac{\beta\sigma}{1+\rho} \right\}^{1/2} s \right] \exp \left\{ i\beta \left(\frac{\rho}{1+\rho} \right) \xi \right\}. \quad (6.115)$$

The FWHM as a function of s can be obtained from (6.114)-(6.115) as,

$$FWHM = \Delta s = 2 \log(1 + \sqrt{2}) \left\{ \frac{1+\rho}{\beta\sigma} \right\}^{1/2}. \quad (6.116)$$

From (6.114)-(6.115), we can see that for formation of bright-dark soliton pairs, we must have, $\beta(r - \rho) > 0$.

If we take $r > \rho$, that is, the maximum intensity of the respective bright component is slightly larger than the maximum intensity of the respective dark component, then $\beta > 0$ has to be satisfied for the bright-dark soliton pair's existence. In the case of a SBN crystal, this implies that $\Delta T > 0$, which in turn implies a heating of the crystal. On the other hand, if we take $\rho > r$, that is, the maximum intensity of the respective dark component is slightly larger than that of the respective bright component, then we need $\beta < 0$, and hence we have, $\Delta T < 0$, which in turn implies a cooling of the crystal. Thus, we can infer that the bright-dark soliton pair can exist for both, a controlled heating and a controlled cooling of the pyroelectric crystal. Fig. 6.23 shows the normalized intensities of the bright –dark incoherently coupled soliton pair for the case of heating of the crystal, i. e. $\Delta T = 10^\circ\text{C}, 20^\circ\text{C}$.

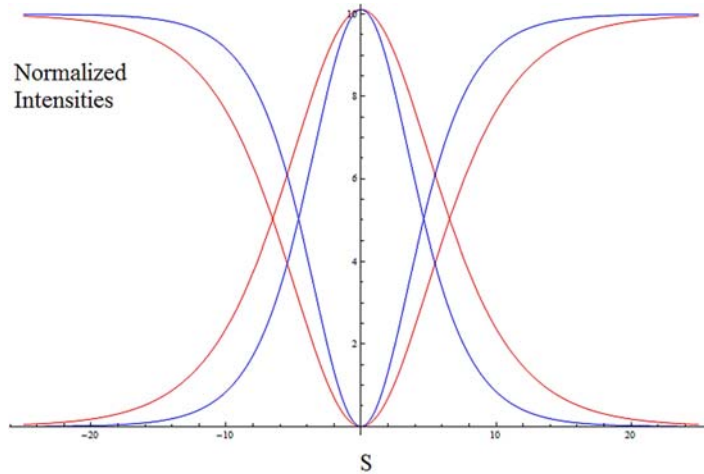


Fig. 6.23. Normalized intensities for the bright-dark incoherently coupled pyroelectric soliton pair for $\Delta T = 10^\circ\text{C}$ (Red), $\Delta T = 20^\circ\text{C}$ (Blue), $r > \rho$, $\rho = 10$. [22]

Fig. 6.24 shows the normalized intensities of the bright-dark incoherently coupled pyroelectric soliton pair for the case of cooling of the crystal, i. e. $\Delta T = -10^\circ\text{C}, -20^\circ\text{C}$.

We plot (6.116) to obtain the existence curve for the bright-dark incoherently coupled pyroelectric soliton pair for a value of $\sigma = 0.01$. The existence curve is plotted in Fig. 6.25.

From Fig. 6.25, we see that the FWHM of the bright-dark incoherently coupled pyroelectric soliton pair remains relatively constant for low values of ρ , which is the maximum intensity of the dark soliton component, and increases with increase of ρ for $\rho > 1$. Interestingly, the FWHM of the bright-dark soliton pair depends upon both r and ρ , the maximum intensity of bright soliton component and dark soliton component respectively.

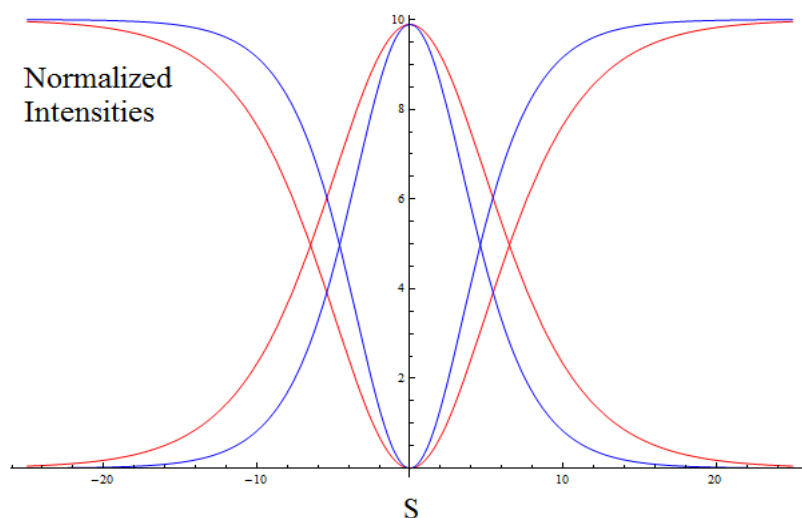


Fig. 6.24. Normalized intensities for the bright-dark incoherently coupled pyroelectric soliton pair for $\Delta T = -10$ °C (Red), $\Delta T = -20$ °C (Blue), $r < \rho$, $\rho = 10$. [22]

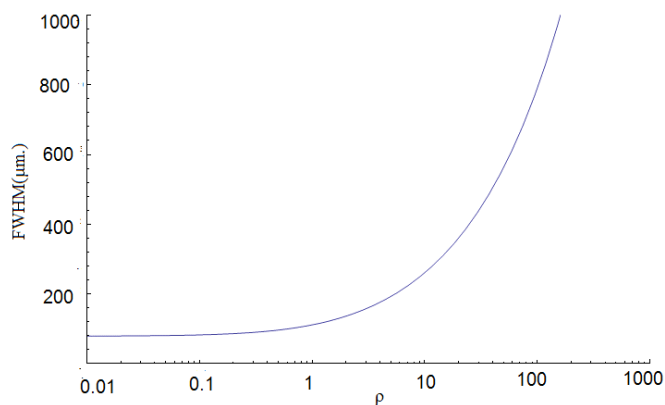


Fig. 6.25. Existence curve for the bright-dark incoherently coupled bright-dark pyroelectric soliton pair, $\Delta T = -10$ °C. [22]

6.3.4. Observation of Separate Components of Incoherently Coupled Solitons

The incoherently coupled bright or dark soliton pairs and families which are observed in pyroelectric materials are expected to be observed in the experimental set-up shown in Fig. 6.26. The incident laser beam is split into $M + N$ mutually incoherent optical beams. A glass slide is inserted in one half of the beam that passes through variable attenuator VA2; by tilting the glass slide the phase across the beam can be adjusted [72], and thus the dark notch with the necessary π phase jump in its center can then be obtained in every one of the N beams to the left-hand side of the glass slide.

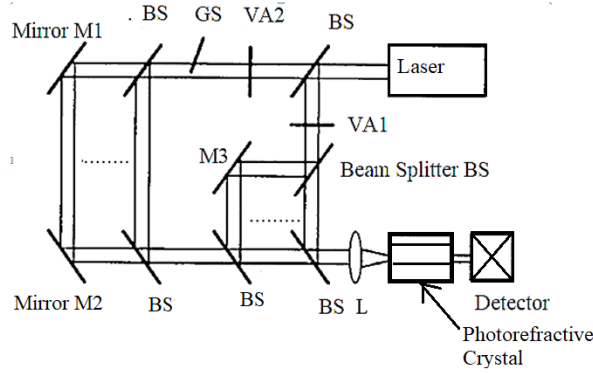


Fig. 6.26. Schematic diagram of the experimental setup for the observation of incoherently coupled soliton families in photorefractive crystals. M1-M3: 100 % reflecting mirrors, BS: Beam Splitters, VA1-VA2: Variable Attenuators, L: Ideal aberration free lens, PRC: Photorefractive Crystal with external bias/temperature change, D: Detector.

The total intensities of the M bright/dark/grey beams and of the N bright/dark/grey ones can be adjusted by variable attenuators VA1 and VA2, respectively. Since the family of collinear beams have the same wavelength and polarization, the researcher can distinguish an arbitrary one among them in the output by blocking the other beams with mechanical shutters, and sampling the specially designated beam within a time interval much shorter than the dielectric relaxation time (response time) of the crystal. This permits viewing every soliton beam separately, because the refractive-index modulation created by all of the soliton beams remains unaffected by the rapid change in the intensity within such a short time interval. The components of the bright-bright, dark-dark, or grey-grey soliton families in either centrosymmetric photorefractives or pyroelectric photorefractives can hence be observed separately. Also, the bright-dark hybrid soliton family solutions of (6.114)-(6.115) are obtained when the peak value of the total intensity of the M bright beams is approximately equal to that of the N dark ones. Under this condition, the intensities of the M bright beams and the N dark ones can be freely adjusted in a certain range [73].

6.4. Conclusions

We have shown that dark-dark, bright-bright and grey-grey incoherently coupled soliton pairs can be supported by a (one-photon) centrosymmetric photorefractive crystal. These can be established provided the incident beams are of same polarization, wavelength, and are mutually incoherent. These incoherently coupled soliton pairs are governed by the quadratic electro-optic effect. We have studied the effect of bias field, beam intensity, and the internal resistance of the crystal along with any external load on the characteristics of the soliton pairs. We find that these parameters have a profound effect on the soliton pairs. We have also proved that our results carry forward to incoherently coupled multicomponent composite solitons which are also supported in such crystals. Finally, we investigate the modulation instability and MI gain for the incoherently coupled soliton

pairs. The MI depends upon the external bias field and the intensity of the solitons. Since, there have been investigations into these soliton pairs before our work [49-51], we would like to enumerate, among other things, certain major differences in our work and the previous research. Ref [50] has investigated the incoherently coupled solitons in two photon Photorefractive media using Castro-Camus model. Ref [49] proves the existence of bright-dark incoherently coupled soliton pair in centrosymmetric Photorefractive media. Ref [51] investigates the bright incoherently coupled Gaussian solitons in centrosymmetric photorefractive media. Our work is different from all of the above mentioned, in that, firstly, our results pertain to photorefractive media exhibiting the one-photon photorefractive effect. Secondly, we prove the existence for incoherently coupled pairs in all the three symmetric realizations, viz. bright, dark and grey in centrosymmetric PR media. It is notable to mention here that, though bright and dark solitons have been predicted and investigated theoretically in centrosymmetric photorefractive media long ago [36], the dark-dark, and particularly the grey-grey soliton pairs along with dark and grey multicomponent solitons have not been investigated theoretically till now. We investigate various characteristics of these soliton pairs in quite a detailed manner. Thirdly, we follow Ref. [62] and [26] in obtaining the soliton solutions directly by numerical methods and do not assume a Gaussian or hyperbolic secant approximation for the input beam. Fourthly, we also present a formulation for multicomponent incoherently coupled solitons in all the three realizations, bright, dark and grey. Lastly, we present a detailed discussion on the modulation instability and the local MI gain in case of all the three soliton pair realizations.

In the second part of this chapter, we have predicted the existence of four types of incoherently coupled soliton pairs, namely, bright-bright, dark-dark, grey-grey and bright-dark, due to only the pyroelectric effects in photorefractive pyroelectric non-photovoltaic crystals. The existence of incoherently coupled bright-bright soliton pair requires a heating of the crystal while the existence of incoherently dark-dark and grey-grey soliton pairs require a cooling of the crystal. The bright-dark soliton pair can exist for both heating and cooling of the crystal. The FWHM of the bright, dark, grey and bright-dark soliton pairs depend upon the intensity of each soliton component incident upon the crystal.

Acknowledgement

This book chapter is based on our original research work published in *Optik-International Journal of Light and Electron Optics* and *Journal of Nonlinear Optical Physics and Materials*.

The content and images in our article [20], is copyright of Elsevier and as the chapter is open access, it amounts to a non-commercial use for scholarly purpose.

Republished with permission of World Scientific, from ‘Incoherently coupled photorefractive spatial solitons supported by pyroelectric effects’, Aavishkar Katti, R. A. Yadav, *Journal of Nonlinear Optical Physics & Materials* 26.01 (2017): 1750002, permission conveyed through Copyright Clearance Center, Inc.

References

- [1]. D. D. Nolte, Photorefractive Effects, Materials, *Springer Science+Business Media*, New York, 1995.
- [2]. M. K. Maurya, R. A. Yadav, Effect of photoconductivity and dielectric constant of the photorefractive materials on two-beam coupling gain and phase-shifts for a single unidirectional photorefractive ring resonator, *Opt. Laser Tech.*, Vol. 42, Issue 6, 2010, pp. 883-893.
- [3]. M. K. Maurya, T. K. Yadav, R. A. Yadav, Oscillation dependence of two wave mixing gain for unidirectional ring resonator in photorefractive materials, *Opt. Laser Tech.*, Vol. 42, Issue 3, 2010, pp. 465-476.
- [4]. T. K. Yadav, M. K. Maurya, R. A. Yadav, Effect of photoconductivity and oscillation frequency shift on the signal beam intensity in two beam coupling in photorefractive materials, *Optik*, Vol. 122, 2011, pp. 1607-1614.
- [5]. T. K. Yadav, M. K. Maurya, R. A. Yadav, Dynamic study of the gain in two beam coupling in photorefractive materials, *Optik*, Vol. 123, 2012, pp. 1329-1335.
- [6]. M. K. Maurya, T. K. Yadav, Ruchi Singh, R. A. Yadav, D. P. Singh, Study of modulation instability in a temporally incoherent photorefractive ring resonator below the threshold with incoherent light, *Opt. Comm.*, Vol. 283, Issue 11, 2010, pp. 2416-2424.
- [7]. M. K. Maurya, T. K. Yadav, R. A. Yadav, Dependence of gain and phase-shift on crystal parameters and pump intensity in unidirectional photorefractive ring resonators. *Pramana - Journal of Physics*, 72, 4, 2009, pp. 709-726.
- [8]. Ruchi Singh, M. K. Maurya, T. K. Yadav, R. A. Yadav, D. P. Singh, Dependence of space charge field and gain coefficient on the applied electric field in photorefractive materials, *Opt. Laser Tech.*, Vol. 43, Issue 1, 2011, pp. 95-101.
- [9]. M. Wichtowski, E. Weinert-Raczka, Analytical solution for temporal photorefractive response in the diffusion regime, *Phot. Lett. Pol.*, 1, 4, 2009, pp. 169-171.
- [10]. P. M. Johansen, A. S. Jensen, Space-charge field in photorefractive media with a constant applied magnetic field, *J. Opt. Soc. Am. B*, Vol. 8, Issue 11, 1991, pp. 2342-2354.
- [11]. D. Sharma, D. Mohan, U. Gupta, Photorefractive space charge electric field in GaAs:Cr with external electric and magnetic fields, *Appl. Phys. Lett.*, Vol. 98, 2011, 211119.
- [12]. I. Aubrecht, L. Solymar, A. Grunnet-Jepsen, Space charge field enhancement in photorefractive materials by applied sinusoidal fields: An approximate analytical solution, *Opt. Comm.*, Vol. 139, 1997, pp. 73-76.
- [13]. E. Serrano, M. Carrascosa, F. Agullo-Lopez, Analytical and numerical study of photorefractive kinetics at high modulation depths, *J. Opt. Soc. Am. B*, Vol. 13, Issue 11, 1996, pp. 2587-2594.
- [14]. M. K. Maurya, T. K. Yadav, R. A. Yadav, Minimization of the fluctuation in the signal beam intensity of a nonlinear optical medium with a transmission grating, *Opt. Laser Tech.*, Vol. 42, Issue 5, 2010, pp. 775-782.
- [15]. M. K. Maurya, R. A. Yadav, Feedback method of the noise suppression in wave mixing amplifiers based on nonlinear materials with photorefractive response in a reflection grating configuration, *Opt. Comm.*, Vol. 283, Issue 12, 2010, pp. 2615-2621.
- [16]. C.-H. Kwak, S. J. Lee, Approximate analytic solution of photochromic and photorefractive gratings in photorefractive materials, *Opt. Comm.*, Vol. 183, Issue 5-6, 2000, pp. 547-554.
- [17]. K. Meerholz, B. L. Volodin, B. Kippelen, N. Peyghambarian, A photorefractive polymer with high optical gain and diffraction efficiency near 100%, *Nature*, Vol. 371, 1994, pp. 497-500.
- [18]. A. Sobolewska, A. Miniewicz, Analysis of the kinetics of diffraction efficiency during the holographic grating recording in azobenzene functionalized polymers, *J. Phys. Chem. B*, Vol. 111, Issue 7, 2007, pp. 1536-1544.

- [19]. T. K. Yadav, M. K. Maurya, R. A. Yadav, Enhancement of the diffraction efficiency of the hologram recorded in photorefractive media under the action of external electric field, *Optik*, Vol. 123, 2012, pp. 1120-1127.
- [20]. A. Katti, R. A. Yadav, and D. P. Singh, Theoretical investigation of incoherently coupled solitons in centrosymmetric photorefractive crystals, *Optik-International Journal for Light and Electron Optics*, 136, 2017, pp. 89-106.
- [21]. Z. Chen, M. Segev, D. N. Christodoulides, Optical spatial solitons: historical overview and recent advances, *Rep. Prog. Phys.*, Vol. 75, Issue 8, 2012, 086401.
- [22]. A. Katti, and R. A. Yadav, Incoherently coupled photorefractive spatial solitons supported by pyroelectric effects, *Journal of Nonlinear Optical Physics & Materials*, 26.01, 2017, 1750002. (Order Detail ID: 71028235, Journal of Nonlinear Optical Physics & Materials by WSPC. Reproduced with permission of WSPC in the format Republish in a book via Copyright Clearance Center).
- [23]. W. Krolikowski, B. Luther-Davies, C. Denz, Photorefractive solitons, *IEEE J. Quantum Electron.*, Vol. 39, 2003, pp. 3-12.
- [24]. M. Segev, B. Crosignani, A. Yariv, B. Fischer, Spatial solitons in photorefractive media, *Phys. Rev. Lett.*, Vol. 68, 1992, pp. 923-926.
- [25]. G. C. Duree, J. L. Shultz, G. J. Salamo, M. Segev, A. Yariv, B. Crosignani, P. Di Porto, E. J. Sharp, R. Neurgaonkar, Observation of self-trapping of an optical beam due to the photorefractive effect, *Phys. Rev. Lett.*, Vol. 71, 1993, pp. 533-536.
- [26]. M. Segev, G. C. Valley, B. Crosignani, P. Di Porto, A. Yariv, Steady-state spatial screening solitons in photorefractive materials with external applied field, *Phys. Rev. Lett.*, Vol. 73, 1994, pp. 3211-3214.
- [27]. D. N. Christodoulides, M. I. Carvalho, Bright, dark, and gray spatial soliton states in photorefractive media, *J. Opt. Soc. Am. B*, Vol. 12, 1995, pp. 1628-1633.
- [28]. A. G. Grandpierre, D. N. Christodoulides, T. N. Coskun, M. Segev, Y. S. Kivshar, Gray spatial solitons in biased photorefractive media, *J. Opt. Soc. Am. B*, Vol. 18, 2001, pp. 55-63.
- [29]. G. C. Valley, M. Segev, B. Crosignani, A. Yariv, M. M. Fejer, M. C. Bashaw, Dark and bright photovoltaic spatial solitons, *Phys. Rev. A*, Vol. 50, 1994, pp. R4457-R4460.
- [30]. M. Taya, M. C. Bashaw, M. M. Fejer, M. Segev, G. C. Valley, Observation of dark photovoltaic spatial solitons, *Phys. Rev. A*, Vol. 52, 1995, pp. 3095-3100.
- [31]. M. Segev, G. C. Valley, M. C. Bashaw, M. Taya, M. M. Fejer, Photovoltaic spatial solitons, *J. Opt. Soc. Am. B*, Vol. 14, 1997, pp. 1772-1781.
- [32]. W. L. She, K. K. Lee, W. K. Lee, Observation of two-dimensional bright photovoltaic spatial solitons, *Phys. Rev. Lett.*, Vol. 83, 1999, pp. 3182-3185.
- [33]. J. Liu, K. Lu, Screening-photovoltaic spatial solitons in biased photovoltaic-photorefractive crystals and their self-deflection, *J. Opt. Soc. Am. B*, Vol. 16, 1999, pp. 550-555.
- [34]. J. Liu, Z. Hao, Evolution of separate screening soliton pairs in a biased series photorefractive crystal circuit, *Phys. Rev. E*, Vol. 65, 2002, 066601.
- [35]. K. Lu, T. Tang, Y. Zhang, One-dimensional steady-state spatial solitons in photovoltaic photorefractive materials with an external applied field, *Phys. Rev. A*, Vol. 61, 2000, 053822.
- [36]. M. Segev, A. J. Agranat, Spatial solitons in centrosymmetric photorefractive media, *Opt. Lett.*, Vol. 22, 1997, pp. 1299-1301.
- [37]. E. DelRe, B. Crosignani, M. Tamburrini, M. Segev, M. Mitchell, E. Refaeli, A. J. Agranat, One-dimensional steady-state photorefractive spatial solitons in centrosymmetric paraelectric potassium lithium tantalate niobate, *Opt. Lett.*, Vol. 23, 1998, pp. 421-423.
- [38]. E. Delre, M. Tamburrini, M. Segev, E. Refaeli, A. J. Agranat, Two-dimensional photorefractive spatial solitons in centrosymmetric paraelectric potassium-lithium-tantalate-niobate, *Appl. Phys. Lett.*, Vol. 73, 1998, pp. 16-18.

- [39]. S. Shwetanshumala, N. Asif, S. Konar, & A. Biswas, Bright spatial solitons in biased centrosymmetric photorefractive medium under drift as well as diffusion effects, *Optik-International Journal for Light and Electron Optics*, Vol. 124, Issue 3, 2013, pp. 229-233.
- [40]. D. N. Christodoulides et al., Incoherently coupled soliton pairs in biased photorefractive crystals, *Applied Physics Letters*, Vol. 68, 1996, p. 1763.
- [41]. S. Konar, S. Jana, S. Shwetanshumala, Incoherently coupled screening photovoltaic spatial solitons in biased photovoltaic photorefractive crystals, *Opt. Comm.*, Vol. 273, 2007, pp. 324-333.
- [42]. H. Chun-Feng, L. Bin, S. Xiu-Dong, J. Yong-Yuan, X. Ke-Bin, Incoherently coupled screening-photovoltaic soliton families in biased photovoltaic photorefractive crystals, *Chin. Phys.*, Vol. 10, 2001, pp. 310-313.
- [43]. L. Keqing, Z. Yanpeng, T. Tiantong, L. Bo, Incoherently coupled steady-state soliton pairs in biased photorefractive-photovoltaic materials, *Phys. Rev. E*, Vol. 64, 2001, 056603.
- [44]. C. Hou, Z. Zhou, X. Sun, B. Yuan, Incoherently coupled grey-grey screening-photovoltaic soliton pairs in biased photovoltaic-photorefractive crystals, *Optik*, Vol. 112, 2001, pp. 17-20.
- [45]. D. N. Christodoulides, M. I. Carvalho, M. Segev, S. R. Singh, Incoherently coupled soliton pairs in biased photorefractive crystals, *Appl. Phys. Lett.*, Vol. 68, 1996, p. 1763.
- [46]. H. Chun-Feng, P. Yan-Bo, Z. Zhong-Xiang, S. Xiu-Dong, Bright dark incoherently coupled photovoltaic soliton pair, *Chin. Phys.*, Vol. 14, 2005, pp. 349-352.
- [47]. H. Wang, X. Peng, Incoherently coupled soliton pairs in photovoltaic crystals, *Chin. J. Phys.*, 50, 2012, p. 589.
- [48]. Keqing Lu, Wei Zhao, Yanlong Yang, Guangde Chen, Jingjun Xu, Yanpeng Zhan, Xun Hou, Incoherently coupled gray-gray soliton pairs in biased photorefractive-photovoltaic crystals, *Opt. Mat.*, 27, 2005, p. 1845.
- [49]. C. F. Hou, C. G. Du, Abdurusul, S. Q. Li, Incoherently coupled bright-dark soliton pairs in biased centrosymmetric photorefractive media, *Chin. Phys. Lett.*, Vol. 18, 2001, pp. 1607-1609.
- [50]. K. Zhan, C. Hou, H. Tian, S. Pu, Y. Du, Spatial solitons in centrosymmetric photorefractive crystals due to the two-photon photorefractive effect, *Journal of Optics*, Vol. 12, Issue 1, 2009, 015203.
- [51]. K. Zhan, C. Hou, Y. Zhang, Incoherently coupled spatial soliton families in centrosymmetric two-photon photorefractive crystals, *Journal of Optics*, Vol. 12, Issue 3, 2010, 035208.
- [52]. J. Safioui, F. Devaux, M. Chauvet, Pyroliton: pyroelectric spatial soliton, *Opt. Exp.*, Vol. 17, Issue 24, 2009, pp. 22209-22216.
- [53]. J. Safioui, E. Fazio, F. Devaux, M. Chauvet, Surface-wave pyroelectric photorefractive solitons, *Opt. Lett.*, Vol. 35, Issue 8, 2010, pp. 1254-1256.
- [54]. Q. Jiang, Y. Su, X. Ji, Pyroelectric photovoltaic spatial solitons in unbiased photorefractive crystals, *Phys. Lett. A*, Vol. 376, Issue 45, 2012, pp. 3085-3087.
- [55]. S. T. Popescu, A. Petris, V. I. Vlad, Fast writing of soliton waveguides in lithium niobate using low-intensity blue light, *Applied Physics B*, Vol. 108, Issue 4, 2012, pp. 799-805.
- [56]. S. T. Popescu, A. Petris, V. I. Vlad, Recording of self-induced waveguides in lithium niobate at 405 nm wavelength by photorefractive-pyroelectric effect, *Journal of Applied Physics*, Vol. 113, Issue 21, 2013, 213110.
- [57]. A. Katti, R. A. Yadav, Spatial solitons in biased photovoltaic photorefractive materials with the pyroelectric effect, *Phys. Lett. A*, Vol. 381, Issue 3, 2017, pp. 166-170.
- [58]. Y. Su, Q. Jiang, X. Ji, Incoherently coupled photorefractive spatial soliton pairs based on the combination of pyroelectric and photovoltaic effect, *Act. Photon. Sin.*, Vol.3, 2014, 014.
- [59]. A. Agranat, V. Leyva, A. Yariv, Voltage-controlled photorefractive effect in paraelectric $\text{KTa}_{1-x}\text{Nb}_x\text{O}_3$:Cu, V, *Opt. Lett.*, Vol. 14, 1989, pp.1017-1019.

- [60]. K. Zhan, et al., Incoherently coupled Gaussian soliton pairs in biased centrosymmetric photorefractive crystals, *Optics & Laser Technology*, Vol. 42, Issue 7, 2010, pp. 1176-1179.
- [61]. Photorefractive Materials and Their Applications I and II, Topics in Applied Physics (P. Gunter, J. P. Huignard, Eds.), Vol. 61-62, *Springer-Verlag*, 1998.
- [62]. D. N. Christodoulides, M. I. Carvalho, Bright, dark, and gray spatial soliton states in photorefractive media, *J. Opt. Soc. Am. B*, Vol. 12, 1995, pp. 1628-1633.
- [63]. M. Segev, G. C. Valley, M. C. Bashaw, M. Taya, M. M. Fejer, Photovoltaic spatial solitons, *J. Opt. Soc. Am. B*, Vol. 14, 1997, pp. 1772-1781.
- [64]. K. Zhan, C. Hou, One-dimensional modulational instability of broad optical beams in biased centrosymmetric photorefractive crystals, *Phys. Lett. A*, Vol. 374, 2009, pp. 169-172
- [65]. E. DelRe, B. Crosignani, M. Tamburrini, M. Segev, M. Mitchell, E. Refaeli, A. J. Agranat, One-dimensional steady-state photorefractive spatial solitons in centrosymmetric paraelectric potassium lithium tantalate niobate, *Opt. Lett.*, Vol. 23, 1998, pp. 421-423.
- [66]. D. Pierangeli, M. Flammini, F. Di Mei, J. Parravicini, C. E. M. de Oliveira, A. J. Agranat, E. DelRe, Continuous solitons in a lattice nonlinearity, *Phys. Rev. Lett.*, Vol. 114, 2015, 203901.
- [67]. D. Pierangeli, F. Di Mei, C. Conti, A. J. Agranat, E. DelRe, Spatial rogue waves in photorefractive ferroelectrics, *Phys. Rev. Lett.*, Vol. 115, 2015, 093901.
- [68]. A. Ziolkowski, E. Weinert-Raczka, Dark screening solitons in multiple quantum well planar waveguide, *J. Opt. A: Pure Appl. Opt.*, Vol. 9, Issue 7, 2007, pp. 688-698.
- [69]. A. Ziolkowski, Temporal analysis of solitons in photorefractive semiconductors, *J. Opt.*, Vol. 14, Issue 3, 2012, 035202.
- [70]. J. E. Zucker, T. L. Hendrickson, C. A. Burrus, Electro-optic phase modulation in GaAs/AlGaAs quantum well waveguides, *Appl. Phys. Lett.*, Vol. 52, 1988, pp. 945-947.
- [71]. A. Katti, R. A. Yadav, Coupled spatial solitons in photorefractive multiple quantum well planar waveguides, in *Proceedings of the 13th International Conference on Fibre Optics and Photonics*, *OSA Technical Digest*, Optical Society of America, 2016, p. Tu3d.3.
- [72]. G. C. Duree, M. Morin, G. Salamo, M. Segev, B. Crosignani, P. Di Porto, E. Sharp, A. Yariv, Dark photorefractive spatial solitons and photorefractive vortex solitons, *Phys. Rev. Lett.*, Vol. 74, 1995, pp. 1978-1981.
- [73]. C. Hou, Z. Zhou, B. Yuan, X. Sun, Incoherently coupled bright–dark hybrid soliton families in biased photovoltaic–photorefractive crystals, *Appl. Phys. B*, Vol. 72, 2001, pp. 191–194.

Chapter 7

Vectorial Complex Ray Model for Light Scattering of Nonspherical Particles

Kuan Fang Ren and Claude Rozé

7.1. Introduction

Geometrical optics is a very simple and intuitive method for treating the interaction of an object with light or electromagnetic waves when the dimension of the object is much larger than the wavelength [1, 2]. One of its main advantages over the other methods is that it can be applied to objects of complex shape, which are hard or even impossible to be dealt with by rigorous theories or most numerical techniques. The variable separation methods based on the solution of Maxwell equations (or its equivalents) are limited to objects that can be described in a coordinate system of the same geometry, such as sphere, spheroid, ellipsoid, or circular or elliptical cylinder. Even in these “simple” cases, the numerical calculation remains another obstacle. Except for the sphere and the infinite circular cylinder, the calculable size of the scatterer can hardly exceed a few tens of wavelengths. Numerical methods such as T matrix, discrete multipole approximation, etc., can be applied to non - spherical particles, but the size parameter of the scatter is also severely limited [3].

Many researchers have contributed to the improvement of geometrical optics. Some take into account the forward diffraction or other particular wave effects (Airy theory for the rainbow [4] and Marston’s model for the critical scattering [5]). Others combine directly geometrical optics with the electromagnetic wave method [6]. However, in these studies interference effects of all order rays are rarely taken into account. We have shown that, by taking into account the interferences between all scattered rays, as well as forward diffraction, we can predict correctly the scattering diagram of a sphere in all directions [7, 8], although the scattering diagram near the critical and rainbow angles is still to be improved. But, as soon as the geometrical optics is extended to a three - dimensional (3D) object of irregular shape, three difficulties are encountered: (1) calculation of local

divergence factors for smooth dielectric surfaces; (2) phase shift due to focal lines; and (3) determination of reflection and refraction angles. If the last one concerns just a technical realization, the two others are inherent problem of ray models. To overcome these obstacles, we have been developing a so - called Vectorial Complex Ray Model (VCRM) in which the wave front curvature is introduced as an intrinsic property of a ray [9]. The calculation of the divergence factor in VCRM is just the ratio of the Gauss curvatures of the wave front surfaces and the phase shift due to the focal line is a count of the sign changes of the wave front curvature radii [10]. The direction of a ray and the Fresnel coefficients are determined simply by the tangent and normal components of the wave vector. The total scattered field is the superposition of the contributions of all complex rays. This model makes it possible to calculate the scattering of any irregularly shaped 3D objects illuminated by a plane wave or a shaped beam. In this chapter, we present the fundamentals of VCRM for an irregularly shaped 3D object, its applications to the scattering of non - spherical particles and characterization of liquid droplets in fluid mechanics.

The structure of the chapter is as follows. In Section 7.2, the fundamentals of geometrical optics and its applications to the scattering of a sphere or an infinite circular cylinder are presented in a way to ease the understanding of VCRM. Section 7.3 is devoted to the description of the Vectorial Complex Ray model, including its fundamental laws and its applications in simple cases of image formation to illustrate its power. The application of VCRM in the light scattering by an elliptical cylinder and a spheroidal particle in a symmetric plane is given in Section 7.4. The last section is the conclusions.

7.2. Fundamentals of Geometrical Optics

In this section we will present the essence of the geometrical optics (GO), or ray model in general sense, then apply it in the light scattering by particles and show that this model can predict very precisely the scattering diagram of a sphere or an infinite cylinder of circular section. These simple cases are very helpful to understand the Vectorial Complex Ray Model.

In ray models, a wave is considered as bundles of rays and each ray is characterized by four parameters.

The direction is usually described by an angle relative to the normal of the diopter. This is sufficient since the incident, reflected and refracted rays remain all in the same plane, called incident plane which is defined by the incident ray and the normal of the diopter.

The amplitude of the wave represented by the ray changes each time it is reflected or refracted by a diopter. It decreases when the light propagates in an absorbing medium. The amplitude evolves also along its path when the wave is convergent or divergent.

The polarization state of the light representing the vibration direction of the electric field is essential in the determination of the amplitudes of reflected and refracted waves by Fresnel coefficients.

The phase of the ray plays an important role in the interference of different orders if the coherent length of the incident wave is larger than the dimension of the particle. The interference can be constructive or destructive according to the phases of the rays. The three main factors which contribute to the phase of a ray are the optical path, the focal line/point and the Fresnel reflection coefficients.

We will see that careful count of these parameters permits to describe well the interaction of light with particle. A detailed description of these parameters will be given later. A special attention is payed to the scattering of light by particle. The sphere and the infinite circular cylinder are taken as examples to show that GO can be applied to deal with the scattering of light with good precision if all the properties are correctly counted. Unfortunately, they are also the only cases that can be treated “rigorously” in the scope of classical ray model. The barrier relies on the fact that in the classical ray model there is no parameter to take into account the divergence/convergence of the wave the rays represent. This will be possible with VCRM described in the next section.

7.2.1. Snell Laws and Fresnel Formulas

In the regime of ray model, the wavelength is much greater than the dimension of the object, the diopter surface can be considered as a plane tangent to the surface (here the convergence of the wave is not concerned). The directions of the reflected and refracted rays are related to that of the incident ray and the relative refractive index between the two media according to the Snell law.

$$\theta_i = \theta_l, \quad (7.1)$$

$$\sin \theta_i = m \sin \theta_r, \quad (7.2)$$

where θ_i , θ_l and θ_r are the incident angle, the reflection angle and the refraction angle respectively (Fig. 7.1).

The amplitude of the reflected wave and the refracted wave relative to the incident amplitude are given by the Fresnel formulas according to the state of polarization:

$$r_{\perp} = \frac{E_{\perp}^r}{E_{\perp}^i} = \frac{\cos \theta_i - m \cos \theta_r}{\cos \theta_i + m \cos \theta_r} = -\frac{\sin(\theta_i - \theta_r)}{\sin(\theta_i + \theta_r)}, \quad (7.3)$$

$$t_{\perp} = \frac{E_{\perp}^r}{E_{\perp}^i} = \frac{2 \cos \theta_i}{\cos \theta_i + m \cos \theta_r} = \frac{2 \sin \theta_r \cos \theta_i}{\sin(\theta_i + \theta_r)}, \quad (7.4)$$

$$r_{\parallel} = \frac{E_{\parallel}^r}{E_{\parallel}^i} = \frac{m \cos \theta_i - \cos \theta_r}{m \cos \theta_i + \cos \theta_r} = -\frac{\tan(\theta_i - \theta_r)}{\tan(\theta_i + \theta_r)}, \quad (7.5)$$

$$t_{\parallel} = \frac{E_{\parallel}^r}{E_{\parallel}^i} = \frac{2 \cos \theta_i}{m \cos \theta_i + \cos \theta_r} = \frac{2 \sin \theta_r \cos \theta_i}{\sin(\theta_i + \theta_r) \cos(\theta_i - \theta_r)}. \quad (7.6)$$

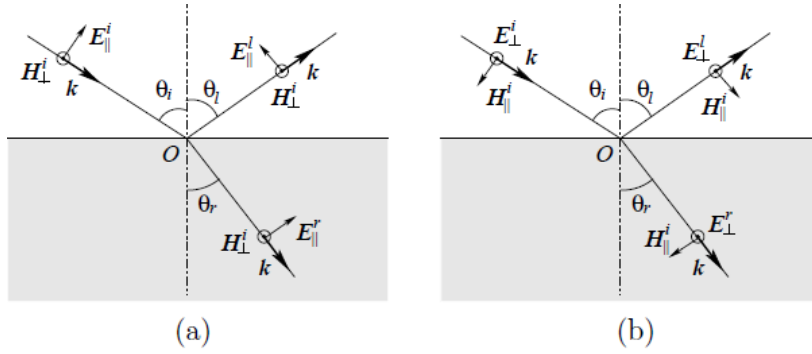


Fig. 7.1. Schema for derivation of Fresnel law.

The ratios of reflected and transmitted energy flux are described respectively by the reflectivity R_X and the transmissivity T_X :

$$R_X = |r_X|^2, \quad (7.7)$$

$$T_X = m |t_X|^2 \frac{\cos \theta_r}{\sin \theta_i}, \quad (7.8)$$

where the index X stands for the state of polarization: the electric field parallel \parallel or perpendicular \perp to the incident plane. An attention should be paid to the fact that the transmissivity is not equal to the square of the amplitude ratio of the refracted (transmitted) wave to the incident wave.

7.2.2. Light Scattering by a Sphere and a Circular Cylinder

The geometrical optics is well known as a very simple and instructive method in dealing with the reflection and refraction of light from an object and widely used in the image formation. It has also been applied in some techniques for optical particle metrology, such as in the prediction of the positions of rainbow or the phase shift in the phase Doppler Anemometry. However, since longtime, the geometrical optics has been considered applicable only in narrow region in forward direction [11, 12]. In the early 2000's we have shown that the geometrical optics is capable to predict the scattering diagram of a sphere in all direction if the interference and diffraction effects are appropriately counted. This method can be easily extended to the scattering of a shaped beam [7, 8], or a multilayered sphere [12] or a cylinder [13]. However, it cannot be applied or extended to the particles without circular symmetry [14]. This will be the subject of the next section.

We deal with the scattering of plane wave by a homogeneous sphere and an infinite circular cylinder at normal incidence. These are the simplest scatterers, one in two dimensions and the other in three dimensions. Thanks to the symmetric property of the problem, the calculations of the deviation of the rays, the phases due to the optical path

and the amplitude ratios of reflection and refractions are the same at each interaction. These three aspects common to the two kinds of particle will be dealt with in the first subsection. The particular problems related to the shape of the particle, such as the phase due to the focal line, the divergence factor and the calculation of the total field will be discussed in the two subsections which follow.

7.2.2.1. Deviation of Rays on Particle Surface

Consider a particle having a circular section of radius a and refractive index m illuminated by a plane wave as shown in Fig. 7.2. The particle can be a sphere or an infinite circular cylinder of axis perpendicular to the plane of paper. We note the order of the emergent rays by p which indicates the emergent rays after $p + 1$ interactions with the particle surface. So the reflected ray corresponds to the order $p = 0$, the order of the first refracted rays is 1, and etc. Due to the symmetry of the problem, the angle of any emergent rays with the normal of the particle surface is constant and equal to the incident angle. The angle between any rays in the particle with the normal of the surface is also constant and equal to the refraction angle. Here we adopt the notations of van de Hulst [1] and note the angle between the incident ray and the tangent plane by τ and the angle between refracted ray and the tangent plane by τ' . They are related to the incident angle θ_i and the refraction angle θ_r by $\tau = \pi/2 - \theta_i$ and $\tau' = \pi/2 - \theta_r$.

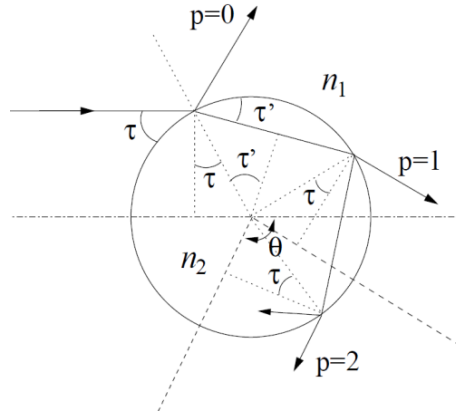


Fig. 7.2. Ray in a spherical/circular cylinder.

When a ray arrives at the surface of the particle, the reflected ray turns in the counterclockwise direction of an angle 2τ . The refracted ray of order p turns in clockwise direction an angle equal to p times $2\tau'$. The deviation angle is therefore given by

$$\theta = 2(\tau - p\tau'). \quad (7.9)$$

However, in practice, the scattered light is observed in 0 to 360° . More particularly, the scattering angle is usually scaled between 0 to 180° due to the symmetry of the sphere and the circular cylinder, and given by following relation

$$\theta_p = 2q_p(\tau - p\tau') + 2k_p\pi, \quad (7.10)$$

where k_p is the integer representing the times the emergent ray crosses the x axis and q_p takes +1 or -1.

7.2.2.2. Amplitudes of Reflected and Refracted Rays

When a ray interacts with the surface of the particle, it is usually divided into two parts: reflected and refracted rays. The amplitudes of the reflected and refracted wave depend on the incident angle and the polarization of the incident wave. The ratios of the reflected and refracted wave amplitudes to the incident one are calculated by the Fresnel formula (7.3) - (7.6).

In the case under study, the Fresnel coefficients are constant for all orders of rays since the incident and refraction angles of all orders are the same. The reflection ratios on the outside of the particle surface are r_{\perp} and r_{\parallel} respectively for the two polarizations. The reflection ratios on the internal surface are the same as the first reflection, but the sign of both ratios are reversed, i.e. the two reflection ratios are respectively $-r_{\perp}$ and $-r_{\parallel}$. All orders of rays enter and exit one time the particle surface, so the refraction ratios are $1 - r_{\perp}^2$ and $1 - r_{\parallel}^2$. In summary, the ratio of the amplitude of the emergent wave of order p to the amplitude of the incident wave is given by

$$\mathcal{E}_{X,p} = \begin{cases} r_X & p=0 \\ (1 - r_X^2)(-r_X)^{p-1} & p \geq 1 \end{cases}, \quad (7.11)$$

where r_X stands for the reflection ratio of perpendicular polarization r_{\perp} or of parallel polarization r_{\parallel} .

7.2.2.3. Phases of Rays

To take into account the interference between different orders or rays, the phase of the rays must be counted. The phase shifts of a ray can be classified into three kinds [1].

1. *Phase due to reflection and refraction Φ_R* : It is well known that when a wave is reflected on the surface from optically thinner medium to optically denser medium, there is a half - wave loss. In fact, this is true only when the incident angle is small. In more general case, this kind of phase shift is accounted in the Fresnel coefficients. When the total reflection occurs, the Fresnel coefficients are complex and the phase shift is to be deduced from the complex values of the coefficients.

2. *Phase due to optical path Φ_P* : This type of phase shift is due to the difference of optical path of rays and evaluated relatively to a reference. The usual reference taken in the literature and chosen here is the virtual ray which (i) propagates in surrounding medium (without any particle), (ii) arrives at the center in the same direction as the incident ray,

and (iii) exits in the same direction as the emergent ray. Therefore, the optical path of the reflected ray ($p = 0$) has a shorter path than the reference ray. This difference is equal to $2as\sin\tau$. Thus it has the positive phase shift equal to $2kas\sin\tau$. On the other hand, the rays of order $p \geq 1$ undergo $p-1$ reflections in the particle. The distance between two successive interactions of a ray with the particle surface is $2as\sin\tau'$, so the supplementary optical path relative to the reflection ray $p = 0$ is equal to $2apms\sin\tau'$. Therefore the phase shift due to the optical path is

$$\Phi_p = 2ka(\sin\tau - pm\sin\tau'). \quad (7.12)$$

3. *Phase due to focal lines* Φ_F : When a ray passes through a focal line, its phase advances by $\pi/2$ [1]. The total phase shift due to focal lines depends on the order of the ray, the position of the ray and the shape of the particle. To illustrate the divergence and convergence of the bundle of rays, the trajectories of two adjacent rays impinging on a particle at different position are shown in Fig. 7.3. The focal lines of the rays (times the rays cross) in the plane of paper in Fig. 7.3(a) for $p = 0$ to 3 are respectively 0, 1, 1, 2 while the corresponding focal lines in Fig. 7.3(b) are 0, 1, 2, 3. In the cases of plane wave scattering by an infinite circular cylinder or a sphere, this phase shift can be calculated analytically according to the deviation of the rays. The details of the calculation will be given later in this section.

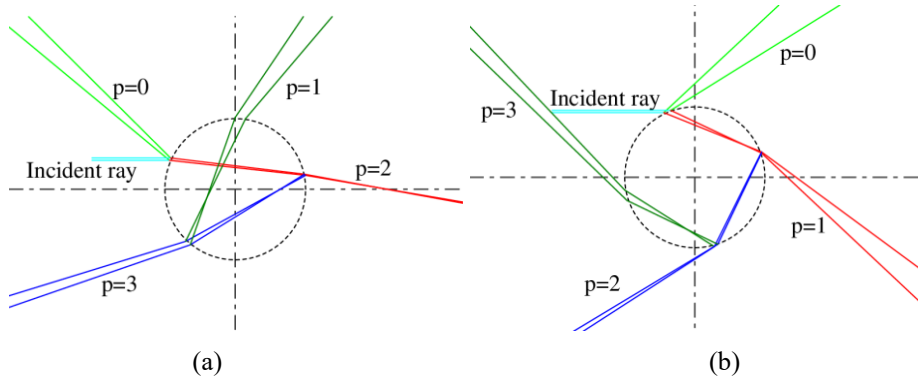


Fig. 7.3. Rays in a circular section.

The total phase of an emergent ray of order p is the summation of the three types:

$$\Phi_p = \Phi_R + \Phi_P + \Phi_F. \quad (7.13)$$

The two first phase shifts are the same for an infinite circular cylinder and a sphere while the last one depends on the shape of the particle. When the incident wave is not a plane wave, the phase of the incident wave must be counted also.

7.2.2.4. Scattering of an Infinite Circular Cylinder

When an infinite circular cylinder is illuminated by a plane wave, the scattering can be dealt with by tracing of rays in the circular section perpendicular to its symmetric axis. The waves being converged or diverged only in one direction, the phase shift due to the focal lines is given by [1]

$$\Phi_F = \frac{\pi}{2} \left[p - \frac{1}{2}(1-s) \right], \quad (7.14)$$

where s is the sign of the angle derivative $d\theta_p/d\tau$ that will be discussed later in this section.

When a wave arrives on a curved surface it will be converged or diverged. The amplitude of the reflected and the refracted wave will change consequently. This variation is described by the divergence factor according to the balance of the energy.

To determine the divergence factor, we consider an incident beam of section dA_i illuminating the particle surface of an area $dA = a d\tau dz$. All the flux in the beam spreads, after interactions with the particle, into an area dA_s in far distance r such that $dA_s = r d\theta dz$. Suppose that the intensity of the ray in far field is $I_p(\theta)$, then according to the energy balance we have $I_0 \varepsilon_{X,p}^2 a \sin \tau d\tau dz = I_p(\theta) r d\theta dz$. It follows that the scattered intensity of order p is given by

$$I_p(\theta) = \frac{\varepsilon_{X,p}^2 I_0 a \sin \tau d\tau}{r d\theta} = \frac{a}{r} I_0 \varepsilon_{X,p}^2 D, \quad (7.15)$$

where $\varepsilon_{X,p}^2$ is the coefficient related to the reflection on and transmission through the particle surface determined by the Fresnel formulas according to Eq. (7.11). The divergence factor D is defined by

$$D = \frac{\sin \tau}{|d\theta / d\tau|}. \quad (7.16)$$

The derivative $d\theta/d\tau$ is deduced from Eq. (7.9) and the Snell law (7.2), and is given by

$$\frac{d\theta}{d\tau} = 2 \left(p \frac{\tan \tau}{\tan \tau'} - 1 \right). \quad (7.17)$$

The divergence factor is therefore

$$D = \frac{m \cos \theta_i \cos \theta_r}{2(p \cos \theta_i - m \cos \theta_r)}. \quad (7.18)$$

In the special case of external reflection ($p = 0$), the divergence factor is simplified to

$$D = \frac{\cos \theta_i}{2}. \quad (7.19)$$

It is independent of the refractive index as it should be.

7.2.2.5. Scattering of a Sphere

When a sphere is illuminated by a plane wave, the trajectories of rays are symmetric around the diameter along the incident direction and can be treated by ray tracing in the circular section. The calculation of the scattering angle and the phase due to the optical path are the same as for a circular cylinder discussed above. But the phase due to the focal points/lines and the divergence factor are different.

The focal lines in the scattering of a plane wave by a sphere can be classified into two sets according to van de Hulst [1]. The first set, noted by set a, is due to the intersection of two adjacent rays in a meridional cross section. The full focal curve is a circle around the axis in a plane perpendicular to the axis. The phase shift due to these focal lines of a ray Φ_{Fa} is the same as for a circular cylinder and given by Eq. (7.14). For a spherical particle, the point where a ray interacts the axis is a focal line, noted as set b, because corresponding rays in other meridional sections have the same point of intersection. The full focal line is the full axis, both before and beyond the sphere. The phase due to these focal lines are

$$\Phi_{Fb} = \frac{\pi}{2} \left[-2k_p + \frac{1}{2}(1 - q_p) \right], \quad (7.20)$$

where k_p and q_p are determined from Eq. (7.10). The total phase shift due to the focal lines of the ray of order p is the summation of the Φ_{Fa} and Φ_{Fb} .

To determine the divergence factor of a sphere, we consider an incident beam of section dA_i illuminating an area $dA = a^2 \cos \tau d\tau d\phi$ on the sphere. All the flux in the beam, after interaction with the particle, spreads into a solid angle $d\Omega$ which corresponds to a surface $dA_s = r^2 d\Omega = r^2 \sin \theta d\theta d\phi$ in far distance r from the sphere. According to the balance of energy, the scattered intensity of order p is given by

$$I_p(\theta) = \frac{\varepsilon_{X,p}^2 I_0 a^2 \sin \tau \cos \tau d\tau d\phi}{r^2 \sin \theta d\theta d\phi} = \frac{a^2}{r^2} I_0 \varepsilon_{X,p}^2 D, \quad (7.21)$$

where the divergence factor D is defined by

$$D = \frac{\sin \tau \cos \tau}{\sin \theta |d\theta / d\tau|}. \quad (7.22)$$

The derivative $d\theta / d\tau$ is given in Eq. (7.17). We find finally the divergence factor of order p in terms of incident and refraction angles as follows

$$D = \frac{\sin(2\theta_i)}{4 \sin \theta \left(p \frac{\tan \theta_r}{\tan \theta_i} - 1 \right)} = \frac{\sin(2\theta_i)}{4 \sin \theta \left(p \frac{\cos \theta_i}{m \cos \theta_r} - 1 \right)}. \quad (7.23)$$

In the special case of the external reflection ($p = 0$), the deviation angle $\theta = \pi - 2\theta_i$, the divergence factor is a constant equal to $1/4$. The deviation angle of the refraction ray is $\theta = 2(\theta_r - \theta_i)$, the divergence factor for the refracted ray $p = 1$ is written as

$$D = \frac{\sin(2\theta_i)}{4 \sin[2(\theta_r - \theta_i)]} \frac{m \cos \theta_r}{\cos \theta_i - m \cos \theta_r}. \quad (7.24)$$

These results will be applied in the next section to check the formulas of VCRM.

7.2.3. Comparison of Scattering Diagrams with Lorenz - Mie Theory

In this section we will present some scattering diagrams calculated by geometrical optics and compare them with the Lorenz - Mie theory to show its applicability.

The scattering diagrams of an infinite cylinder calculated by the geometrical optics described in the above section and by Lorenz - Mie theory for two different sizes are shown in Fig. 7.4. The incident wavelength is $0.6328 \mu\text{m}$ and the refractive index of the particle is 1.33 . It is clear that when the size of the particle is much greater than the wavelength ($a = 50 \mu\text{m}$ in Fig. 7.4(a)) the agreement between GO and the LMT is very satisfactory in almost all directions except in vicinity of rainbow angles. From Fig. 7.4(b) ($a = 10 \mu\text{m}$) we can see that the GO can predict correctly the scattering for particle of size about 10 times the wavelength.

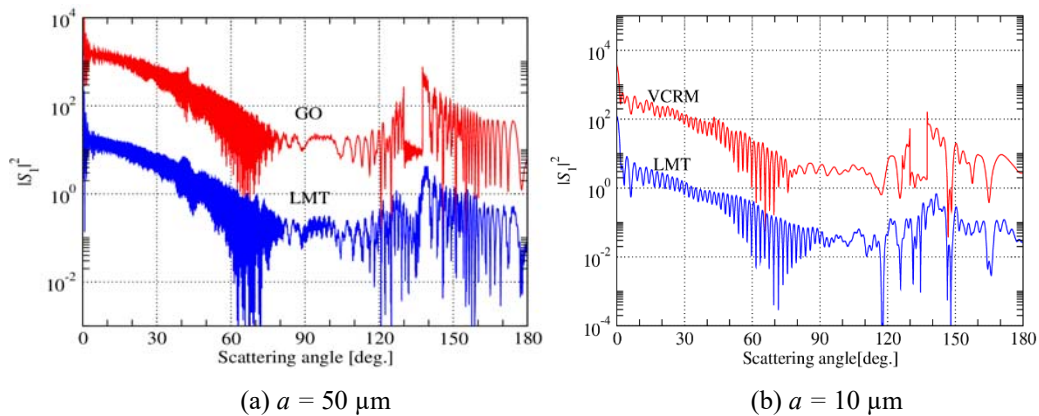


Fig. 7.4. Comparison of the scattering diagrams computed by GO and LMT for an infinite circular cylinder of water ($m = 1.33$) of radius a illuminated by a plan wave of wavelength $\lambda = 0.6328 \mu\text{m}$. The results of LMT and GO are shifted by 10^{-2} and 10^2 respectively for clarity [13].

The geometrical optics can easily be extended to the scattering of shaped beam, called therefore Extended Geometrical Optics Approximation (EGOA) [7, 8]. When the divergence of the incident beam is small, we suppose that the rays in the particle propagate rectilinearly. In that case, we need only to take into account the local amplitude and phase of each ray at incident point. The incident angle is the angle between the normal of the particle surface and the normal of the incident wave front surface which is determined by the gradient of the phase function of the incident wave.

We compare in Fig. 7.5 the scattering diagrams calculated by EGOA and Generalized Lorenz - Mie theory (GLMT) for a spherical water droplet. The agreement is excellent in all directions and still better than the case of plane wave. This is because the waist radius of the incident beam is smaller than the radius of the particle and the intensity of the rays which contribute to scattering around the rainbow angles is much smaller than that on the axis.

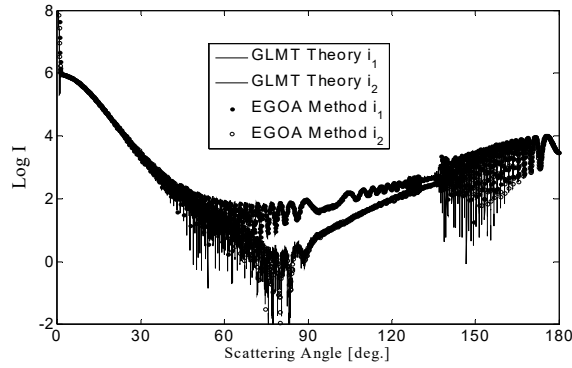


Fig. 7.5. Comparison of the scattering intensities calculated by GLMT and EGOA for a water droplet ($m = 1.333$) of radius $a = 25 \mu\text{m}$ illuminated by a Gaussian beam of waist radius $w_0 = 10 \mu\text{m}$ and wavelength $\lambda = 0.6328 \mu\text{m}$. The particle is located at the center of the beam.

7.3. Vectorial Complex Ray Model

We have seen in the previous section that, if the amplitude and the phase shifts of all the rays are correctly counted, the classical geometrical optics can predict the scattering of a wave by an infinite circular cylinder and a sphere with good precision when the particle size is large compared to the wavelength. This is based on the fact that the divergence factor and the phase shift due to the focal lines can be calculated analytically.

In the case where the particle has no such circular or spherical symmetry (referred in the following as non - spherical particle), the incident angle changes at each interaction of a ray with the particle surface. The divergence factor depends on the local curvature of the particle surface, the wave front curvature of the incident wave and the incident angle. No analytical expression of divergence factor is possible for arbitrarily shaped particle. We have shown also that these two properties cannot be achieved by pure numerical technique [14].

In principle, the phase shift due to focal lines is just a count of the passage of focal lines. Accordingly it is dependent on the convergence of the waves in and out of the particle. Therefore, to count correctly the variation of the amplitude of a ray and its phase shift due to the focal lines, we must be able to predict the curvature of the wave front of the wave that a ray represents.

In the Vectorial Complex Ray Model (VCRM) that will be presented in this section, the wave front curvature is considered as a new intrinsic property of a ray besides the four properties described in Section 7.2. Furthermore, in VCRM all properties of a ray: propagation direction, polarization, amplitude and phase will be described in vector, their components and complex numbers.

7.3.1. Snell Law and Fresnel Formulas in Vector Form

Knowing the fact that $k_v \sin \theta_v$ ($v = i, l$ or r represent respectively incident, reflected or refracted wave) is the tangent component of the wave vector \mathbf{k}_v in the tangent plane to the diopter and $k_r = mk_i$, the Snell laws of reflection and refraction (7.1) and (7.2) can be written simply as

$$k_{i\tau} = k_{l\tau} = k_{r\tau}, \quad (7.25)$$

where the index τ stands for the tangent component. The normal components of the reflected and refracted waves are respectively $k_{ln} = -k_{in}$ and $k_{rn} = \sqrt{k_r^2 - k_{i\tau}^2}$.

Similarly, since $k_v \cos \theta_v$ is the component of the wave vector in the normal direction of the diopter surface, the Fresnel formulas (7.3)-(7.6) can therefore be written as function of the normal components of wave vectors

$$r_{\perp} = \frac{k_{in} - k_{rn}}{k_{in} + k_{rn}}, \quad (7.26)$$

$$t_{\perp} = \frac{2k_{in}}{k_{in} + k_{rn}}, \quad (7.27)$$

$$r_{\parallel} = \frac{m^2 k_{in} - k_{rn}}{m^2 k_{in} + k_{rn}}, \quad (7.28)$$

$$t_{\parallel} = \frac{2mk_{in}}{m^2 k_{in} + k_{rn}}. \quad (7.29)$$

When the total reflection occurs, the tangent component of the incident wave vector is greater than the wave number of the refracted wave $k_{\tau} > k_r$. The normal component of the refracted wave becomes a pure imaginary number if the two media are both transparent.

By taking into account the time convention, it can be written as $k_{rn} = -i\sqrt{k_{i\tau}^2 - k_r^2}$. The

Fresnel coefficients become complex. Therefore, the phase shifts due to the reflection and refraction vary, in general, as function of the incident angle and must be counted correctly in the phase calculation of the rays.

It is worth to note that though Eqs. (7.25)-(7.29) are equivalent to Eqs. (7.1)-(7.6) they are much more convenient to the numerical calculation, especially for the scattering of 3D irregular particles since only four basic operations are necessary.

7.3.2. Wave Front Equation

When a wave arrives on a curved surface, the reflected wave and the refracted wave will be converged or diverged according to the curvature of the surface, i.e. the wave front curvature will change and the amplitude of the emergent wave in far field will be more or less important. To describe this property, we introduce the wave front curvature as a new property of a ray. By matching the phase between the incident wave and the reflected or refracted wave, we can establish the wave front equation. The derivation of this equation is tedious and will be omitted here. We will focus our attention to its physical interpretation and applications.

Consider an arbitrary wave whose wave front curvature at the incident point is described by the curvature matrix Q in the base $(\mathbf{t}_1, \mathbf{t}_2)$ (Fig. 7.6)¹. The curvature of the dioptric surface is given by the curvature matrix C in its base $(\mathbf{s}_1, \mathbf{s}_2)$. The curvature matrix of the wave after refraction or reflection Q' is given by the wave front equation [9]

$$(\mathbf{k}_r - \mathbf{k}) \cdot \mathbf{n} C = k_r \Theta'^T Q' \Theta' - k_i \Theta^T Q \Theta, \quad (7.30)$$

where the letters with prime represent the quantities after refraction or reflection, the superscript T the transpose of the matrix, Θ the projection matrix between the unitary vectors of the coordinates systems on the planes tangent to the wave front $(\mathbf{t}_1, \mathbf{t}_2)$ and the dioptric surface $(\mathbf{s}_1, \mathbf{s}_2)$

$$\Theta = \begin{pmatrix} \mathbf{t}_1 \cdot \mathbf{s}_1 & \mathbf{t}_1 \cdot \mathbf{s}_2 \\ \mathbf{t}_2 \cdot \mathbf{s}_1 & \mathbf{t}_2 \cdot \mathbf{s}_2 \end{pmatrix}. \quad (7.31)$$

In the case of the plane wave scattering by an ellipsoid or when the axis of an axis-symmetric beam passes by the symmetric axis of the ellipsoid, the rays in the plane defined

¹ When $(\mathbf{t}_1, \mathbf{t}_2)$ are the principal directions of the wave front, the matrix Q is diagonal and may be written as $Q = \begin{pmatrix} 1/R_1 & 0 \\ 0 & 1/R_2 \end{pmatrix}$, where R_1 and R_2 are the principal curvature radii of the wave front. Similar notation is applied to the dioptric surface. For example, the curvature matrix C of a sphere of radius a is $C = \begin{pmatrix} 1/a & 0 \\ 0 & 1/a \end{pmatrix}$, and that of a infinite cylinder is $C = \begin{pmatrix} 1/a & 0 \\ 0 & 0 \end{pmatrix}$.

by the beam axis and the ellipsoid axis remain always in this plane for any orders. This is a special case but very interesting because it simplifies considerably the problem. It permits a good understanding of the essential concepts of VCRM and reveals certain physical phenomena.

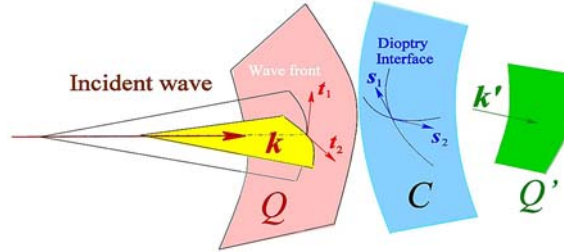


Fig. 7.6. Schema of the wave fronts and the dioptric surface.

Without loss of generality, we suppose that s_2 and t_2 are in the incident plane, that means $s_1 \cdot t_1 = 1$ and $s_2 \cdot t_2 = \cos \theta$. Similar relations can be found for the refracted wave. The wave front equation (7.30) is simplified to two scalar equations. The relation between the curvature radii in the plane perpendicular to the incident plane is given by [9, 10]:

$$\frac{k_r}{R_{r1}} = \frac{k_i}{R_{i1}} + \frac{k_{rn} - k_{in}}{\rho_1}, \quad (7.32)$$

and the relation of the curvature radii in the incident plane reads as:

$$\frac{k_{rn}^2}{k_r R_{r2}} = \frac{k_{in}^2}{R_{i2}} + \frac{k_{rn} - k_{in}}{\rho_2}, \quad (7.33)$$

where R stands for the curvature radius of the wave front and ρ that of the diopter. The index 1 and 2 indicate respectively the values in the plane perpendicular or parallel to the symmetric plane. It is important to note that the wave front curvature in the direction perpendicular to the scattering plane evolves also at each interaction with particle surface. This is different from the pure two dimension problem as an infinite cylinder of circular, elliptical or any other section shape.

7.3.3. Amplitude and Phase of a Ray

The amplitude of a ray may change during the propagation due to the convergence and divergence. In the case of light scattering of a plane wave by an infinite circular cylinder or by a sphere, the divergence factor has been introduced to describe the variation of the intensity (or amplitude) of the ray. A phase shift of focal line has also been calculated according to the suggestion of van de Hulst [1]. In the aforementioned cases, both the divergence factor and the total phase shift due to the focal lines are given in analytical

expressions. In the framework of VCRM for non spherical particle, these two properties are to be evaluated step by step according to the wave front curvature.

7.3.3.1. Amplitude

When a wave propagates from one point to another, its amplitude, therefore the intensity evolves according to the divergence of the wave and the distance between the two points. The relation of the intensities between points A and B along a ray can be deduced from the energy balance as shown in Fig. 7.7.

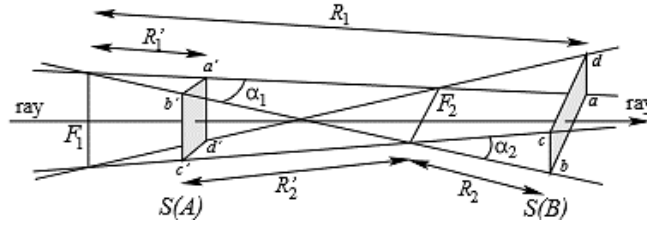


Fig. 7.7. Variation of pencil cross section and phase shift due to focal line.

The two focal lines are F_1 and F_2 . We note the two principal radii of the curvature at A by R_1' and R_2' , and those at B are R_1 and R_2 , then the surface $a'b'c'd'$ at A is $S(A) = R_1'\alpha_1 R_2'\alpha_2$ and the surface $abcd$ at B is $S(B) = R_1\alpha_1 R_2\alpha_2$. The convergent or divergent characteristic of the wave can be noted by the sign of the curvature radii. In this chapter we adopt the convention that the curvature is positive if the focal line is after the considered point. Therefore, R_2' in Fig. 7.7 is positive while R_1' , R_1 and R_2 are all negative. The sign of the curvature radii permits to count the phase of focal line that will be discussed later. For what concerns the intensity, only their absolute values matter. According to the conservation of energy, the energy flux passing through the surface $S(A)$ is equal to that through the surface $S(B)$, i.e. $I(A)[R_1'\alpha_1 R_2'\alpha_2] = I(B)[R_1\alpha_1 R_2\alpha_2]$. We deduce therefore

$$I(B) = I(A) \left| \frac{R_1' R_2'}{R_1 R_2} \right|. \quad (7.34)$$

In the light scattering by a particle, the intensity of the reflected wave at the first reflection is the product of the intensity of the incident wave on the particle surface I_0 and the intensity reflection ratio $|r_{x,0}|^2$. If we note the two radii of the reflected wave front by R_{11}^e and R_{21}^e , the scattered intensity at distance r is then

$$I(r) = I_0 |r_{x,0}|^2 \left| \frac{R_{11}^e R_{21}^e}{(r - R_{11}^e)(r - R_{21}^e)} \right|, \quad (7.35)$$

since the curvature radii of the reflected wave at distance r are $(r - R_{11}^e)$ and $(r - R_{21}^e)$.

Similarly, the intensity of the scattered light of order $p = 1$ is

$$I(r) = I_0 \left| t_{X,0} t_{X,1} \right|^2 \left| \frac{R_{11}' R_{21}'}{R_{11} R_{21}} \cdot \frac{R_{12}^e R_{22}^e}{(r - R_{12}^e)(r - R_{22}^e)} \right| \quad (7.36)$$

where the factors $t_{X,0}$ and $t_{X,1}$ are respectively the Fresnel transmission coefficient for $p = 0$ and 1. The term $\frac{R_{11}' R_{21}'}{R_{11} R_{21}}$ is the ratio of the intensities at two successive points.

In general, if we note the two curvature radii of the incident wave front at j^{th} interaction point by R_{1j} and R_{2j} ($j = 1, 2, \dots, q$), those of the reflected or refracted wave by R_{1j}' and R_{2j}' , and the two curvature radii of the emergent wave by R_{1j}^e and R_{2j}^e , then the intensity of the emergent ray after q interactions with the diopter is given by

$$I(r) = I_0 \left| \varepsilon_{X,p} \right|^2 D', \quad (7.37)$$

where $p = q - 1$ and the divergence factor D is defined by

$$D' = \left| \frac{R_{11}' R_{21}'}{R_{12} R_{22}} \cdot \frac{R_{12}' R_{22}'}{R_{13} R_{23}} \dots \frac{R_{1q}^e R_{2q}^e}{(r - R_{1q}^e)(r - R_{2q}^e)} \right|. \quad (7.38)$$

The factor $\varepsilon_{X,p}$ is due to the reflection and refraction and given by

$$\varepsilon_{X,p} = \begin{cases} r_{X,0} & p = 0 \\ t_{X,0} t_{X,p} \prod_{i=1}^{p-1} r_{X,i} & p \geq 1 \end{cases}. \quad (7.39)$$

It is worth to note that $r_{X,0}$ is the Fresnel reflection coefficient on the surface outside of the particle while $r_{X,i}$ with $i \geq 1$ is the coefficient of the internal reflection. In the special case of light scattering of plane wave by an infinite circular cylinder or a spherical particle, the reflection coefficient is constant for all orders of rays. Eq. (7.39) is therefore reduced to Eq. (7.11).

The definition of the divergence factor (7.38) includes directly the size of the particle and is consistent to the divergence factor of the infinite circular cylinder and the sphere defined in Section 7.2. It includes already the prefactor a/r for the circular cylinder (see Eq. (7.18)) and a^2/r^2 for the sphere (see Eq. (7.23)). The detailed derivation of the divergence factor

of classical geometrical optics for the scattering of an infinite circular cylinder and a sphere from Eq. (7.38) will be given in the next subsection.

It is worth to point out that in light scattering theories, we talk often about the scattering diagram described by a function $F(\theta, \phi)$, such that (see Section 2.1 in [1])

$$I(\theta, \phi) = I_0 \frac{F(\theta, \phi)}{k^2 r^2}, \quad (7.40)$$

which is independent of r in far field. In Lorenz - Mie theory for sphere, $F(\theta, \phi)$ is equal to $|S_1(\theta, \phi)|^2$ for perpendicular polarization and $|S_2(\theta, \phi)|^2$ for parallel polarization.

In VCRM, if we are interested only the scattering in far field, the term $(r - R_{1q}^e)(r - R_{2q}^e)$ is eliminated in the same way. The amplitude of scattered wave of order p is then calculated by

$$A_p = A_0 k |\varepsilon_{X,p}| \sqrt{D''}. \quad (7.41)$$

Where the new divergence factor is defined by

$$D'' = \left| \frac{R'_{11} R'_{21}}{R_{12} R_{22}} \cdot \frac{R'_{12} R'_{22}}{R_{13} R_{23}} \cdots R_{1q}^e R_{2q}^e \right|. \quad (7.42)$$

In the case of 2D, i.e. scattering by an infinite cylinder (circular, or elliptical or of any section shape), Eq. (7.41) is reduced to

$$A_p = A_0 |\varepsilon_{X,p}| \sqrt{\frac{\pi k}{2} \left| \frac{R'_1}{R_2} \cdot \frac{R'_2}{R_3} \cdots R'_{1q} \right|}, \quad (7.43)$$

since the convergence occurs only in one direction. The factor $\pi/2$ under the square root is necessary for the results to be consistent with LMT (see Eq. (8.40) in [15]).

7.3.3.2. Phase

The phase of rays plays a critical rule in counting the wave effect to predict the fine structure of the light scattering diagrams, such as the supernumerary bows, the fringes near the critical angle or the interference structure near different kinds of caustics. The phase of a ray in VCRM is counted in four parts:

The first is the phase due to optical path Φ_P which is computed directly according to the optical path, usually relative to a reference ray which arrives at the particle center in the same direction as the incident ray and goes out in the same direction as the emergent ray.

Let \mathbf{r}_i and $\hat{\mathbf{k}}_i$ be respectively the position vector of i^{th} interaction point of the ray with particle surface and the normalized wave vector of the emergent ray from that point, the phase due to the optical path of a ray after q interactions with the particle is given by

$$\Phi_p = -k_0(\hat{\mathbf{k}}_0 \cdot \mathbf{r}_1 - \hat{\mathbf{k}}_q \cdot \mathbf{r}_q) - k_p \sum_{i=1}^{q-1} \hat{\mathbf{k}}_i \cdot (\mathbf{r}_{i+1} - \mathbf{r}_i), \quad (7.44)$$

where $\hat{\mathbf{k}}_0$ is the normalized wave vector of the incident ray. In the case of scattering of the plane wave by an infinite circular cylinder or a sphere, $\hat{\mathbf{k}}_0 \cdot \mathbf{r}_1 = -\hat{\mathbf{k}}_q \cdot \mathbf{r}_q = -a \cos \theta_i$. The equation (7.23) of the phase due to the optical path of geometrical optics is recovered.

The second concerns the phase due to the reflection Φ_R which is calculated directly according to the Fresnel coefficient as described in the previous section. If no total reflection occurs, only a constant phase π is added to the perpendicular component for any incident angle and to the parallel component if the incident angle is larger than the Brewster angle. When the total reflection occurs, the phase shifts depend on the incident angle and the phase shifts are to be calculated and added to both polarizations. This is responsible of the effects of the tunneling (Goos Hänchen's shift) [16].

As for the phase due to the focal line, it is easy to be calculated in VCRM, because we need only to count the number n_f of sign changes of the wave front curvature radii between successive interactions of ray with the particle surface. And the total phase shift is given by

$$\Phi_F = n_f \frac{\pi}{2}. \quad (7.45)$$

Finally, if the incident wave is a shaped beam, the phase of the incident wave Φ_i is also to be counted.

In conclusion, the total phase of a ray in VCRM is summation of four kinds of phase shifts:

$$\Phi = \Phi_p + \Phi_R + \Phi_F + \Phi_i. \quad (7.46)$$

The phase due to optical path is independent of polarization and divergence of the wave, while the computation of the phases of reflection and focal lines is delicate, especially for problems without symmetry (or simply called 3D scattering) since both depend on the polarization of the wave.

7.3.4. Simple Applications of the Wave Front Equation

We will apply the wave front equation obtained in the Section 7.3.2 in some special cases: first in the formation of image by a plane dioptr and a spherical dioptr, and then in the

determination of the divergence factor in the plane wave scattering by an infinite circular cylinder and a sphere. These examples show two aspects of applications of the wave front equation. They help also to understand different aspects of the equations.

7.3.4.1. Image Formation by a Plane Diopter

Consider a plane diopter which separates two media 1 and 2 of refractive indices equal respectively to m_1 and m_2 (Fig. 7.8). A point object A is in the medium 1 distant OA from the refraction point O . The curvature radius of the diopter being infinity, the distance of the image A' from the refraction point, according to Eq. (7.33), is given by

$$\frac{m_2 \cos^2 \theta_r}{OA'} = \frac{m_1 \cos^2 \theta_i}{OA}. \quad (7.47)$$

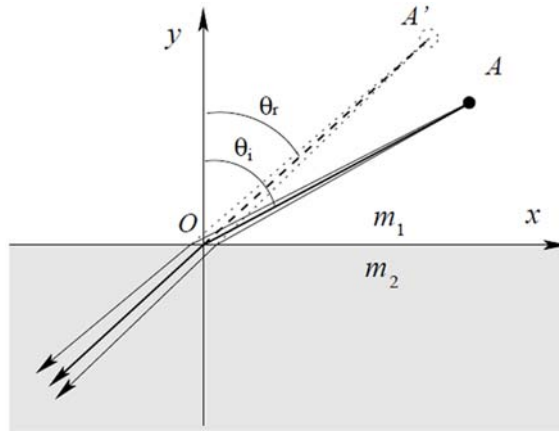


Fig. 7.8. Image formation by a plane diopter.

We note that \overline{OA} and $\overline{OA'}$ have the same sign. This implies that the object and its image are in the same side of the diopter. Suppose that the coordinates of the object point A are (x, y) . We can obtain the coordinates of the image point B according to Eq. (7.42)

$$x' = \frac{\cos^2 \theta_r}{\cos^2 \theta_i} x, \quad y' = \frac{m_2 \cos^2 \theta_r}{m_1 \cos^2 \theta_i} y. \quad (7.48)$$

Therefore, both the lateral and longitudinal distances between the image and the object $x' - x$ and $y' - y$ depend on the incident angle θ_i . When the incident angle θ_i is small, the common conjugation relation of a plane diopter $\frac{m_1}{OA} = \frac{m_2}{OA'}$ is recovered. And only in this case the lateral shift $x' - x$ tends to 0, i.e. $x' \approx x$.

7.3.4.2. Image Formation by a Spherical Diopter

Consider a spherical diopter of radius a and center C separating two media of refractive indices m_1 and m_2 (Fig. 7.9). A point source A is placed on an axis passing by C .

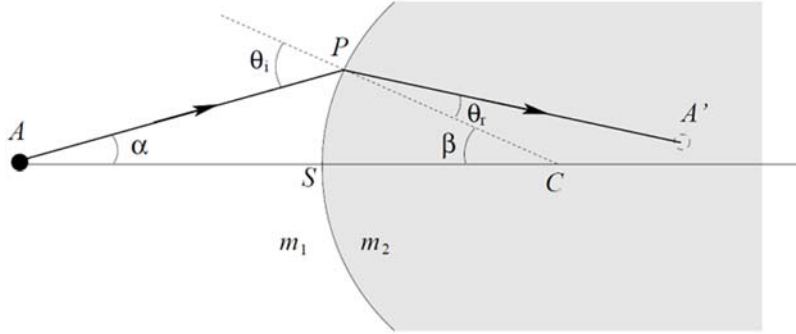


Fig. 7.9. Image formation by a spherical diopter.

In the paraxial case, i.e. for small θ_i , the commonly used conjugation relation

$$\frac{m_2}{SA'} = \frac{m_1}{SA} + \frac{m_2 - m_1}{SO}, \quad (7.49)$$

is found directly by either of Eq. (7.32) or Eq. (7.33). However, in general case, the two curvature radii of refracted wave are different. The image distance given by the wave front equation for the curvature radius in the incident plane (7.33) is

$$\frac{m_2 \cos^2 \theta_r}{PA'_2} = \frac{m_1 \cos^2 \theta_i}{PA} + \frac{m_2 \cos \theta_r - m_1 \cos \theta_i}{a}, \quad (7.50)$$

which indicates that the image is a focal line perpendicular to the incident plane at A'_2 . Yet, the image distance given by the second wave front equation is

$$\frac{m_2}{PA'_1} = \frac{m_1}{PA} + \frac{m_2 \cos \theta_r - m_1 \cos \theta_i}{a}, \quad (7.51)$$

which means that the image is a focal line in the incident plane at A'_1 . Consequently, the image formed by rays off - axis is deformed. This is the source of aberration. In fact, the wave front equation can be applied to any curved diopter, so it is a powerful tool to study the aberration of any curved surfaces. See Ref. [16] for detailed discussion on the aberration of imaging formation.

7.3.4.3. Divergence Factor of a Circular Cylinder

Consider an infinite circular cylinder of radius a illuminated perpendicularly by a plane wave. One of the principal directions of the surface of the cylinder is along the axis and the other is tangent to the surface and perpendicular the axis. The corresponding curvature radii are respectively ∞ and a . The two curvature radii of the plane wave are both infinite.

For the reflection, the normal component of the reflected wave vector is opposite to that of the incident wave $k_{rn} = -k_{in}$ or $\cos \theta_r = -\cos \theta_i$. From Eq. (7.33) it is ready to find the curvature radius in the incident plane of the reflected wave

$$R_{r1} = -\frac{a \cos \theta_i}{2}. \quad (7.52)$$

The negative sign signifies that the curvature center is on the other side of the emergent wave. The other curvature radius is ∞ .

For the refraction, the curvature radius of the refracted wave in the incident plane R_{r1} is found directly from Eq. (7.33)

$$\frac{m \cos^2 \theta_r}{R_{r1}} = \frac{m \cos \theta_r - m \cos \theta_i}{a}. \quad (7.53)$$

If the refractive index of the cylinder $m > 1$, the curvature radii are positive, i. e. the center of the refracted wave is in the same side of the cylinder axis. In the contrary, if $m < 1$ R_{r1} is negative, the wave center is in the opposite side of the cylinder axis.

From the curvature radii given above, we can deduce the divergence factor of a cylinder. For the reflection, the divergence factor is

$$D' = \frac{R_{r1}}{r - R_{r1}} = \frac{a}{r - a} \frac{\cos \theta_i}{2}, \quad (7.54)$$

and that of the refraction is

$$D' = \frac{R_{r1}}{R_{r2}} \frac{R_{r2}}{r - R_{r2}} = \frac{a}{r - a} \frac{m \cos \theta_i \cos \theta_r}{2(\cos \theta_i - m \cos \theta_r)}. \quad (7.55)$$

Naturally, in the far field the divergence factors given by geometrical optics are recovered as it should be.

7.3.4.4. Divergence Factor of a Sphere

Similar to the infinite cylinder, we can also demonstrate that the divergence factor given by GO can be found as a special case in VCRM. For example, the two principal curvature radii of the reflected wave can be obtained by the wave front equations (7.32) and (7.33)

$$R_{10} = -\frac{a \cos \theta_i}{2}, \quad R_{20} = -\frac{a}{2 \cos \theta_i}. \quad (7.56)$$

The convergence factor for reflection is therefore ($r \rightarrow \infty$) $D = \frac{1}{4}$.

We can also demonstrate that the divergence factor of a sphere given by GO for high order rays can be found from the formulism of VCRM. Though there is no difficulty in principle, the calculation is somewhat tedious and we will omit it here.

7.4. Applications of VCRM in Light Scattering

As a short review, we will give a very brief description of some remarkable results to show the power of VCRM. The detailed information may be found in the references therein.

7.4.1. Revisit of Airy Theory in Term of VCRM

We will apply VCRM to the scattering of a spherical particle. We are interested especially in the scattered intensity around the rainbow angles and compare the results of VCRM to the Airy theory since the latter is largely used in the particle granulometry.

We show first in Fig. 7.10 the scattering diagrams calculated with three different methods [17]. It is clear that the difference of the peak positions and the intensity maxima of supernumerary bows predicted by the Airy theory and the rigorous Debye theory [4] increases rapidly as function of the distance from the main peak of rainbow, especially for the second rainbow. Whereas the agreement between the results of VCRM and the Debye theory is very satisfactory except in the neighborhood of geometrical rainbow angle since the diffraction effect must be introduced in VCRM to remedy this flaw. The reason why the Airy theory differs from the rigorous theory is due to two approximations in that theory: 1). the cubic phase function is deduced from the relation of deviation angle in the vicinity of rainbow angle, but in the Airy integration, this variable is extended to infinity, and 2) the amplitude of the rays near the rainbow angle is assumed to be constant but this is certainly not true because the divergence factor tends to infinity at geometrical rainbow angle. In VCRM, the phase and the amplitude of each ray are calculated rigorously in the framework of ray model. The results are naturally better than that of Airy theory.

Furthermore, it is worth to point out that in VCRM the same procedure can be applied directly to calculate the intensity of the supernumerary bows of a non-spherical particle. Examples will be given in the last section.

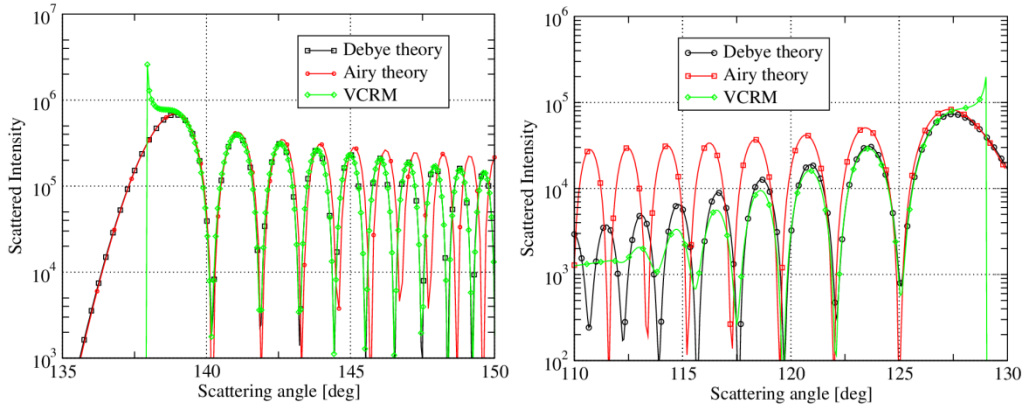


Fig. 7.10. Comparison of the scattered intensity calculated by Debye theory, Airy theory and VCRM for a spherical particle of water ($m = 1.333$) of radius $a = 50 \mu\text{m}$ for the first order (left figure $p = 2$) and the second order (right figure $p = 3$) of rainbow.

7.4.2. Scattering by an Elliptical Cylinder

The scattering of a plane wave by an infinite elliptical cylinder at normal incidence is the simplest case of scattering by non - spherical particle since the convergence and divergence occur in one plane. Only one scalar wave front equation (7.33) is necessary. As an example, we show in Fig. 7.11 the scattering diagrams of an elliptical cylinder of semi - axis $a = 50 \mu\text{m}$ along the incident direction and the other semi axis varies from $25 \mu\text{m}$ to $50 \mu\text{m}$. The refractive index of the cylinder is 1.33 and the incident wavelength is $\lambda = 0.6328 \mu\text{m}$.

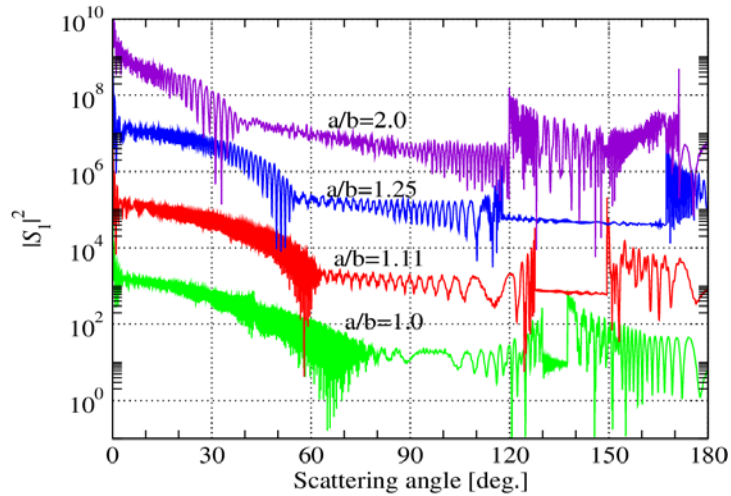


Fig. 7.11. Scattering diagrams of an elliptical cylinder illuminated by a plane wave for the different aspect ratios. The perpendicular polarization is chosen.

We find that when the aspect ratio a/b increases from 1 to 1.25, the first order rainbow goes to larger angle, the second order rainbow to the smaller angle and the Alexander region expands. When the aspect ratio is too big ($a/b = 2.0$ in the figure), the positions of the first rainbow (at 119.8°) and the second order rainbow (at 171.2°) are reversed.

VCRM is very easily to be extended to the scattering of a shaped beam. Fig. 7.12 shows the scattering diagrams of an elliptical cylinder at illuminated with a two dimensional Gaussian beam of different waist radius.

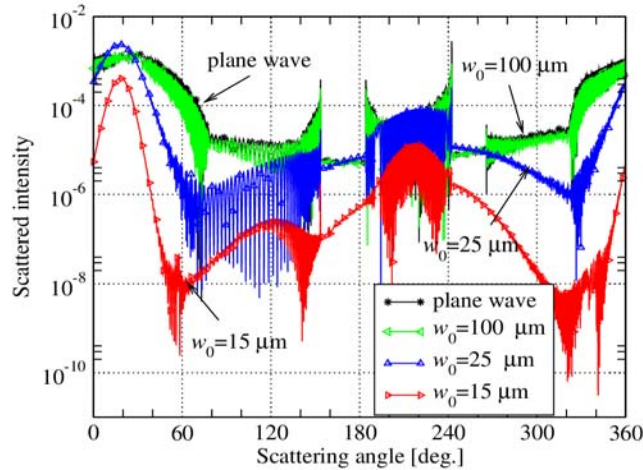


Fig. 7.12. Scattered diagrams of an elliptical cylinder of major radius $a = 50 \mu\text{m}$ and minor radius $b = 40 \mu\text{m}$ illuminated by plane wave and a two dimensional Gaussian beam of three different waist radius ($w_0 = 100, 25, 15 \mu\text{m}$). The incident beam is polarized along z direction and makes an angle $\theta_0 = 20^\circ$ with x axis. The center of the beam is located on the axis of the cylinder [13].

The two semi - axes of the elliptical cylinder are respectively $a = 50 \mu\text{m}$ and $b = 40 \mu\text{m}$ and the incident wave is the plane wave or a two dimensional Gaussian beam of waist radius $w_0 = 100 \mu\text{m}$, $25 \mu\text{m}$ or $15 \mu\text{m}$. The incident wave propagates in the direction x perpendicular to the axis of the cylinder z but makes an angle of 20° with the major axis of the elliptical section. We remark that the profile of the scattering diagrams are very different from those of the circular cylinder. The scattering diagrams are no longer symmetric, so they must be given in all directions (0 to 360°). The rainbow angles and the Alexander dark regions in the two sides of the scattering diagram (0° to 180° and 180° to 360°) are not symmetric neither. When a cylinder is illuminated by a two dimensional Gaussian beam of waist radius relatively small, the incident beam intensity at the impact position for rainbow is weak. For example, in the case $w_0 = 15 \mu\text{m}$, the rainbow phenomena are not visible in two sides relative to the incident direction (20 to 200° and 200° to 20°). If $w_0 = 25 \mu\text{m}$ the rainbow is much visible in the side of scattering angle smaller than 200° than in the other side.

7.4.3. Scattering of the Plane Wave by an Ellipsoidal Particle

An ellipsoidal particle is a simple but very interesting model of non spherical particle [18, 19]. By changing the three semi - axes and incident angle we can investigate the influence of the curvature of the particle surface in different direction. To illustrate the essential characteristics of the scattering we will limit ourselves to the scattering of a plane wave in a symmetric plane of the particle.

All the calculations presented in this section have been done by the software VCRMElI2D which can be download from our website [20]. A FORTRAN version is also available by requiring the authors. The results of VCRM and these codes have been validated by comparison with a rigorous numerical method [21, 22].

7.4.4. Software VCRMElI2D

The software VCRMElI2D is composed of two modules, one for ray tracing and the other for the calculation of scattering diagram.

In the module of “ray tracing”, apart from the properties of the particle, one can choose to illuminate whole or a portion of the particle with bundle of rays at a given angle. An example is shown in Fig. 7.13(a). This function permits to visualize the divergence and convergence of the wave in and out of the particle.

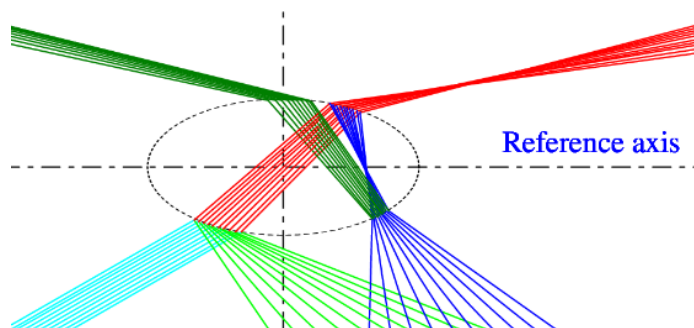
The module of “Scattering diagram” (Fig. 7.13(b)) permits to calculate the intensity of emergent rays for each order and the total scattering intensity. The diffraction in the forward direction can also be considered. Therefore, the interference is taken into account properly. The data of the scattering diagrams of each order and the total field are saved automatically in separated files. One can also choose to save all the properties of all the rays at each intersection point, including the coordinates of the point, the wave vectors, the Fresnel coefficients, the curvature radii of the diopter and the curvatures radii of the wave fronts of incident, reflected and refracted waves.

7.4.5. Hyperbolic Umbilic Foci of an Oblate Particle and Experimental Validation

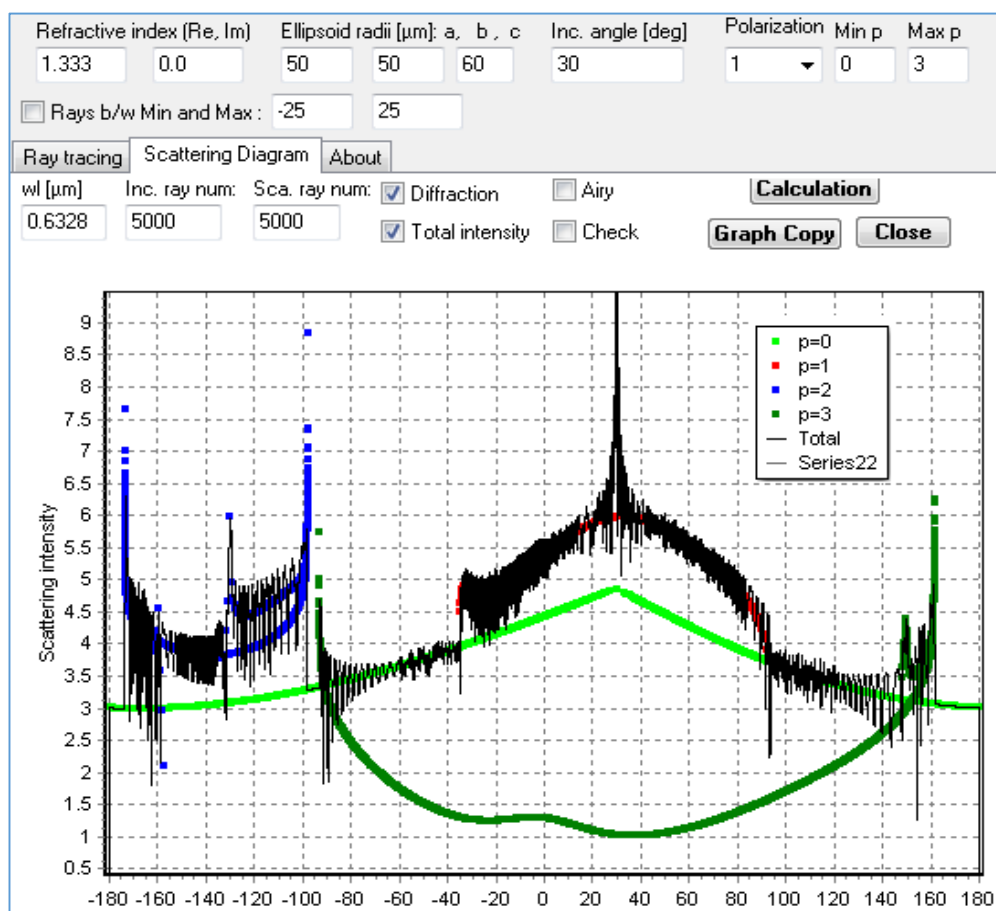
The hyperbolic umbilic foci or the hyperbolic umbilic diffraction catastrophe of an oblate particle is a very interesting phenomenon in the light scattering and attracted attention of many researchers. It is, in fact, a deformation of rainbow in an oblate particle. Nye [23] and Marston et al [24] have given explanation to their formation of the skeleton in term of geometrical optics. Fig. 7.14 shows the photographs of rainbow region scattering patterns and the explanation of Marston. But the geometrical optics does not permit to predict the fine structure of the fringes in the scattering patterns.

To investigate these phenomena in the framework of VCRM and validate our model, an experiment has been realized [25, 26]. A droplet of Di - Ethyl - Hexyl - Sebacat (DEHS) is levated using acoustic pressure. The red curves in Fig. 7.15 show the measured scattering intensities around the first rainbow angles. The blue curves presents the

scattering diagrams calculated with VCRMELI2D. It is clear that the VCRM predicts very well the fine structure in the rainbow region of an oblate.



(a) Ray tracing in a spheroidal particle.



(b) Calculation of the intensity of each order and the total scattering diagram.

Fig. 7.13. Illustration of the software VCRMELI2D.

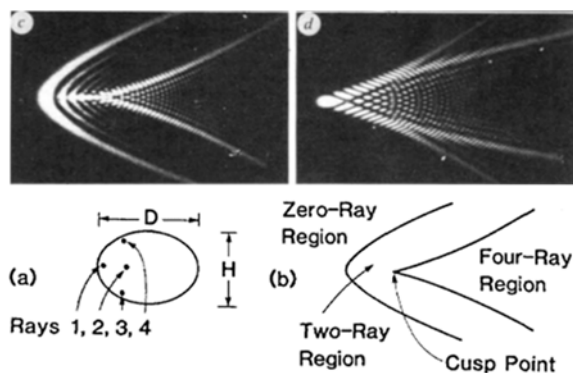


Fig. 7.14. Photographs of rainbow region scattering patterns (top) and explanation of Marston [24].

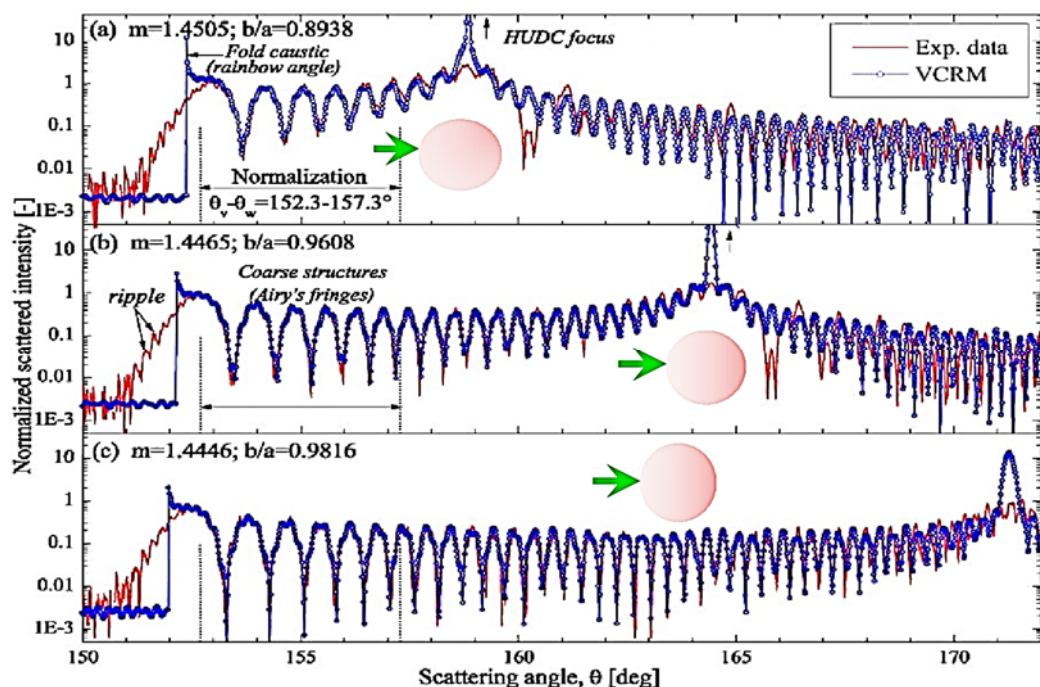


Fig. 7.15. Comparison of VCRM and experimental normalized equatorial scattering diagrams for the acoustically levitated droplets of Di - Ethyl - Hexyl - Sebacat [25].

7.4.6. Dependence of Two Rainbow Intensity Ratio on the Aspect Ratio of a Prolate Particle

It is known that the intensity of the rainbow of a spherical particle decreases rapidly with the increasing of the order. The intensity of the second rainbow is about one magnitude weaker than the first rainbow. However, for a prolate particle, though the rainbow angles remain the same as a spherical particle of radius equal to the minor axis, the ratio between

the rainbow intensities of different orders varies as function of its aspect ratio. This is because the convergence of the wave in the direction along the major axis. As examples, we show in Fig. 7.16 the intensity diagram near the first and the second rainbows calculated by VCRMII2D for a sphere and a prolate particle of two aspect ratios. The intensity ratio between the second rainbow and the first rainbow increases as function of the aspect ratio c/a . This property is very interesting to the characterization of non - spherical particle since it can be used to deduce the deformation of the particle [27].

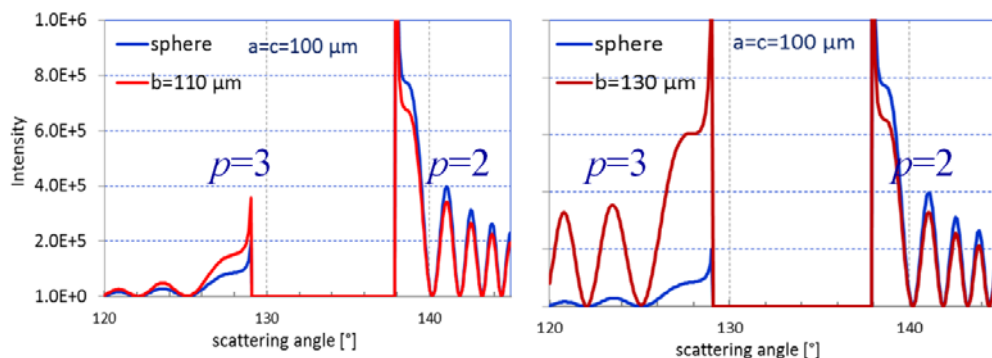


Fig. 7.16. Scattering intensities near the first and the second rainbows of a sphere and a prolate water droplet.

7.5. Conclusions

We have developed since a decade a novel model, called Vectorial Complex Ray Model (VCRM), to deal with the scattering of large non - spherical particle. The key originality of VCRM relies on the introduction of a new intrinsic property to describe the rays, i.e. the wave front curvature. This property, as well as four properties in the classical ray model (direction, amplitude, phase and polarization) evolves at each interaction of a ray with the particle surface and permits to evaluate precisely the amplitude and the phase of each ray so that to predict the scattering field with very good precision.

To ease the access of the new model, we begin with the simple cases of plane wave scattering by an infinite circular cylinder and a sphere. The essential concepts and the fundamental laws of VCRM are then presented along with simple applications to help the understanding.

Some remarkable results have been exemplified to illustrate the power of VCRM. We have shown that VCRM predicts much better the positions and the amplitudes of supernumerary bows than the Airy theory of rainbow, which has been largely used in the measurement of refractive index and size of spherical particle. The infinite elliptical cylinder is the simplest non - spherical particle. We have illustrated that its scattering diagram are very sensible to the aspect ratio and the waist radius of the incident Gaussian beam.

Finally, the free software VCRM2D has been presented and applied to the scattering of a spheroidal particle. It is shown that VCRM predicts very well the fine structure in the rainbow (called also the hyperbolic umbilic diffraction catastrophe or hyperbolic umbilic foci) of an oblate particle. The intensity ratio between different rainbows is sensible to the aspect ratio of a prolate particle. VCRM can quantify this ratio and therefore be applied to the characterization of the deformation of droplets.

Acknowledgements

This work was funded by the French National Research Agency (ANR) under grants AMO - COPS (ANR - 13 - BS09 - 0008 - 01).

References

- [1]. H. C. van de Hulst, Light Scattering by Small Particles, *Dover Publications*, 1957.
- [2]. A. Macke, M. Mishchenko, K. Muinonen, B. Carlson, Scattering of light by large nonspherical particles: ray-tracing approximation versus T-matrix method, *Opt. Lett.*, Vol. 20, 1995, pp. 1934-1936.
- [3]. M. I. Mishchenko, J. W. Hovenier, L. D. Travis, Light Scattering by Nonspherical Particles: Theory, Measurements and Applications, *Academic*, 2000.
- [4]. E. Hovenac, J. Lock, Assessing the contributions of surface waves and complex rays to far - field Mie scattering by use of the Debye series, *J. Opt. Soc. Am. A*, Vol. 9, 1992, pp. 781-795.
- [5]. P. Marston, Critical angle scattering by a bubble: physical - optics approximation and observations, *J. Opt. Soc. Am.*, Vol. 69, 1979, pp. 1205-1211.
- [6]. P. Yang, K. Liou, Geometric-optics-integral-equation method for light scattering by nonspherical ice crystals, *Appl. Opt.*, Vol. 35, 1996, pp. 6568-6584.
- [7]. F. Xu, K. Ren, X. Cai, Extension of geometrical-optics approximation to on-axis Gaussian beam scattering. I. By a spherical particle, *Appl. Opt.*, Vol. 45, 2006, pp. 4990-4999.
- [8]. F. Xu, K. Ren, X. Cai, J. Shen, Extension of geometrical-optics approximation to on-axis Gaussian beam scattering. II. By a spheroidal particle with end - on incidence, *Appl. Opt.*, Vol. 45, 2006, pp. 5000-5009.
- [9]. K. F. Ren, F. Onofri, C. Rozé, T. Girasole, Vectorial complex ray model and application to two - dimensional scattering of plane wave by a spheroidal particle, *Opt. Lett.*, Vol. 36, Issue 3, 2011, pp. 370-372.
- [10]. K. F. Ren, C. Rozé, T. Girasole, Scattering and transversal divergence of an ellipsoidal particle by using Vectorial Complex Ray Model, *J. Quant. Spectrosc. Radiat. Transfer*, Vol. 113, 2012, pp. 2419-2423.
- [11]. A. Ungut, G. Grehan, G. Gouesbet, Comparisons between geometrical optics and Lorenz - Mie theory, *Appl. Opt.*, Vol. 20, 1981, pp. 2911-2918.
- [12]. F. Xu, X. Cai, K. F. Ren, Geometrical - optics approximation of forward scattering by coated particles, *Appl. Opt.*, Vol. 43, Issue 9, 2004, pp. 1870-1879.
- [13]. K. Jiang, Theoretical study of light scattering by an elliptical cylinder, PhD Thesis, *University of Rouen*, France, 24 June 2013.
- [14]. Y. Yuan, Diffusion de la lumière par un objet irrégulier pour l'application à l'imagerie des sprays, PhD Thesis, *University of Rouen*, 29 March 2012.
- [15]. C. F. Bohren, D. R. Huffman, Absorption and Scattering of Light by Small Particles, *J. Wiley and Sons*, New York, 1983.
- [16]. M. Born, E. Wolf, Principles of Optics, 7th Ed., *Cambridge University Press*, 1999.

- [17]. K. F. Ren, Airy theory revisited and caustics in vectorial complex ray model, in *Proceedings of the 7th World Congress on Particle Technology (WCPT'14)*, Beijing, China, May 2014.
- [18]. J. Lock, Ray scattering by an arbitrarily oriented spheroid. I. Diffraction and specular reflection, *Appl. Opt.*, Vol. 35, 1996, pp. 500-514.
- [19]. J. Lock, Ray scattering by an arbitrarily oriented spheroid. II. Transmission and cross - polarization effects, *Appl. Opt.*, Vol. 35, 1996, pp. 515-531.
- [20]. AMOCOPS, <http://www.amocops.eu/>
- [21]. M. Yang, Y. Wu, X. Sheng, K. F. Ren, Comparison of scattering diagrams of large non - spherical particles calculated by VCRM and MLFMA, *J. Quant. Spectrosc. Radiat. Transfer*, Vol. 162, 2015, pp. 143-153.
- [22]. M. Yang, Computation of light scattering, radiation force, torque and stress of large non - spherical particles with multilevel fast multipole algorithm and Vectorial Complex Ray Model, PhD Thesis, *University of Rouen*, 9 December 2014.
- [23]. J. F. Nye, Rainbow scattering from spheroidal drops - an explanation of the hyperbolic umbilic foci, *Nature*, Vol. 312, Issue 5994, 1994, pp. 531-532.
- [24]. P. L. Marston, E. H. Trinh, Hyperbolic umbilic diffraction catastrophe and rainbow scattering from spheroidal drops, *Nature*, Vol. 312, Issue 5994, 1984, pp. 529-531.
- [25]. M. P. L. Sentis, F. R. A. Onofri, L. Méès, S. Radev, Scattering of light by large bubbles: coupling of geometrical and physical optics approximations, *J. Quant. Spect. Rad. Trans.*, Vol. 170, 2016, pp. 8-18.
- [26]. K. F. Ren, F. R. A. Onofri, M. Yang, X. Sheng, The fine structure in the scattering diagrams of large ellipsoidal particle predicted by Vectorial Complex Ray Model, in *Proceedings of the 9th International Symposium on Measurement Techniques for Multiphase Flows (ISMTMF'15)*, Sapporo, Japan, Sept. 2015.
- [27]. K. F. Ren, C. Roze, S. Idlahcen, Z. Ma, Observation of exotic scattering patterns of suspended droplets and their theoretical prediction by Vectorial Complex Ray Model, in *Proceedings of the 27th Annual Conference on Liquid Atomization and Spray Systems (ILASS-Europe'16)*, Brighton, United Kingdom, Sept. 2016.

Chapter 8

Magneto-Optical Effects Arising from Coupling of Magnetic and Dielectric Properties for Colloidal Particle System

Jian Li and Xiangshen Meng

8.1. Introduction

Magneto-optical effects, which are generally independent of intensity of light as linear optical effects [1, 2] and related to the strong field-responsive properties of magnetically controlled colloids, such as ferrofluids and some liquid crystals, have been intensively studied. However, the origins of these magneto-optical effects are different. For ferrofluids based on ferromagnetic/ferrimagnetic nanoparticles, the magneto-optical effects observed are regarded generally as resulting from structural transformation on chain-like aggregates produced by interparticle interactions, rather than the properties of individual particles [3-6], though a few reports disagree [7, 8]. Indeed Skibin et al. [9] and Scholten [10] have already proven theoretically that models based on long range orientation of permanent dipoles of single particles would never be able to account for the commonly observed high values of birefringence and dichroism in ferrofluids. Moreover, the light scattering by linear chains of magnetic nanoparticles in ferrofluids have been studied [11, 12]. Also, for liquid crystals, the additional magneto-optical effects result from the orientation of the anisotropic building blocks in the magnetic field [13-15]. In addition, there are two reports describing magneto-optical and light-effect experiments performed with ferrofluid-doped lyotropic liquid crystals in which chain formation with and without the application of an external magnetic field is discussed [16, 17].

In our previous work, magneto-optical effects of the colloids based on weakly magnetic goethite nanoparticles have been found [18, 19]. In this work, we show a mechanism of magneto-optical effects, which arise from the coupling of the magnetic and dielectric properties of colloidal particles rather than structural transformation in the colloid or long-ranged spatial order of the particles. Also, it is proved that for magnetic colloids based on

spherical nanoparticles having cubic crystal structure, such as are employed in most of the experiments on ferrofluids, the reaction of single particles could not results in the magneto-optical effects.

8.2. Theoretical Framework

8.2.1. Origin of Magneto-Optical Effects for Colloidal Particle System

So called magneto-optical effects or field-induced optical effects mean that the electromagnetic properties of the media change in an external magnetic field, so that the propagation behavior of light in the media also changes. The well-known magneto-optical effects are independent of intensity of light and linear optical effects. Therefore, such optical effects are related to the media's dielectric tensor $[\epsilon]$, conductivity tensor $[\sigma]$ and permeability tensor $[\mu]$. Generally, the optical effects take place in the high-frequency regime, as the visible region of the electromagnetic spectrum, so that the action of ϵ involves σ , and $\mu \approx 1$. As a consequence, field-induced optical effects can be discussed in terms of an effective dielectric constant [20]. For colloids consisting of nanoparticles and a carrier liquid, the relationship between the electric displacement vector D and the electrically polarized vector P is

$$D = [C]E = [\epsilon_0]E + P, \quad P = P_p + P_c, \quad (8.1)$$

where E is the external electric field, $[C]$ is the dielectric tensor of the colloid and $[\epsilon_0] = \epsilon_0 \delta_{ij}$, $\delta_{ij} = 1 (i = j)$, $\delta_{ij} = 0 (i \neq j)$, ϵ_0 is the dielectric constant in free space, and P_p and P_c are electrically polarized vectors of the colloid particles and carrier liquid, respectively. The colloids can be viewed as individual nanoparticles dispersed in the continuous phase of carrier liquid, so P_p and P_c , which depend linearly on optical field (E) in linear optics, can be described as

$$P_p = \frac{1}{V} \sum_{i=1}^N V_p ([\epsilon_p]E - [\epsilon_0]E) = \Phi (<[\epsilon_p]> - [\epsilon_0])E, \quad (8.2)$$

where V_p and V are the volume of a particle and the colloid, respectively, N is the number of particles and $[\epsilon_p] = [\epsilon(\omega)_p]$ is the dielectric tensor of a particle and in relation to the field frequency ω , $\Phi = NV_p/V$ is the volume fraction of the particle system in the colloid, $<[\epsilon_p]>$ means the statistical average of the tensors of all particles and

$$P_c = \frac{1}{V} \int_0^{V_c} ([\epsilon_c]E - [\epsilon_0]E) dV = (1 - \Phi)([\epsilon_c] - [\epsilon_0])E, \quad (8.3)$$

where V_c is the volume of the carrier liquid, $(1 - \Phi) = V_c/V$ is the volume fraction of the carrier liquid in the colloid, and $[\epsilon_c]$ is the dielectric tensor of the carrier liquid, which is isotropic, i.e. $[\epsilon_c]_{ij} = \epsilon_c \delta_{ij}$. Substituting Eqs. (8.2) and (8.3) into (8.1), the dielectric tensor of the colloid $[C]$ is given by

$$[\epsilon] = \Phi \langle [\epsilon_p] \rangle + (1 - \Phi) [\epsilon_c]. \quad (8.4)$$

Eq. (8.4) shows that $\langle [\epsilon_p] \rangle$ characterizes the optical action of colloids when the carrier liquid is isotropic and its magnetic effect can be neglected. Electron microscope observations have shown that nonsphericity of particles in ferrofluids is usually very small, so that the shape anisotropy of individual particles is very small. In the present work, only spherical particles are considered. For simplicity, we denote $\langle [\epsilon_p] \rangle$ as $[\epsilon]$.

8.2.2. Magneto-Dielectric Properties of Crystals and Colloids

Since magnetic perturbation have little effect on the optical properties of the medium, for a crystal having a spontaneous magnetization M , the dielectric constants of ϵ_{ij} can be expressed in the form

$$\epsilon_{ij}(M) = \epsilon_{ij}^0 + f_{ijl}^{(1)} M_l + f_{ijlm}^{(2)} M_l M_m + f_{ijlmn}^{(3)} M_l M_m M_n + \dots, \quad (8.5)$$

where ϵ_{ij}^0 is the dielectric tensor of a crystal in the paramagnetic state (for which $M = 0$), $f_{ijl}^{(1)}$ is a third-rank tensor which governs magneto-optical effects that are linear in respect of the magnetization, $f_{ijlm}^{(2)}$ is a fourth-rank tensor describing the quadratic effects, etc. [21]. The generalized principle of symmetry of the kinetic coefficients, i.e. Onsager relation, relates the components ϵ_{ij} and ϵ_{ji} as $\epsilon_{ij}(M) = \epsilon_{ji}(-M)$, lead to the realization that some components are odd in M , some even in M . Thus, the requirement of reality of ϵ_{ij} - tensor for a transparent medium is replaced with the requirement of its hermiticity: $\epsilon_{ij} = \epsilon_{ji}^*$, and $[\epsilon]$ can be written as [22, 23]

$$[\epsilon] = \underbrace{\begin{bmatrix} \epsilon'_{11} & \epsilon'_{12} & \epsilon'_{13} \\ \epsilon'_{12} & \epsilon'_{22} & \epsilon'_{23} \\ \epsilon'_{13} & \epsilon'_{23} & \epsilon'_{33} \end{bmatrix}}_{\substack{\text{symmetric} \\ \text{in } M}} + i \underbrace{\begin{bmatrix} 0 & \epsilon''_{12} & \epsilon''_{13} \\ -\epsilon''_{12} & 0 & \epsilon''_{23} \\ -\epsilon''_{13} & -\epsilon''_{23} & 0 \end{bmatrix}}_{\substack{\text{antisymmetric} \\ \text{in } M}}. \quad (8.6)$$

For M or H in the Z -direction and a crystal of lower than uniaxial symmetry, symmetry considerations lead to the $[\epsilon]$ of the following form

$$[\epsilon] = \begin{bmatrix} \epsilon_{11} & 0 & 0 \\ 0 & \epsilon_{22} & 0 \\ 0 & 0 & \epsilon_{33} \end{bmatrix} + i \begin{bmatrix} 0 & \epsilon_{12} & 0 \\ -\epsilon_{12} & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad (8.7)$$

where ϵ_{11} , ϵ_{22} and ϵ_{33} are even functions of either M or H , and ϵ_{12} is an odd function [20]. In the description of Eq. (8.7), it is considered that like every symmetrical tensor of rank two, the tensor ϵ_{ij} can be brought to diagonal form by a suitable choice of the coordinate

axes. Only considering the influence of terms linear in magnetization M , ϵ_{ii} is independent on M and ϵ_{12} is linear function of the M .

For spherical particles, it can be assumed that particles have identical intrinsic dielectric properties $[\epsilon_p]$ as Eq. (8.7) described and intrinsic magnetic moments m parallel to a particular crystal axis. When an external magnetic field H is applied, the moment m inside the particle interacts with the external field via the potential $U_{m-H} = -\mu_0 m \cdot H$, and orienting torque acting on the moment $T^m = \mu_0 m \times H$, where μ_0 is the permeability of free space [24]. Thus, the moment may follow Brownian rotation, accompanied by bulk rotational diffusion of the particle within the carrier liquid [25], and favor the orientation parallel to the direction of the magnetic field. Nevertheless, thermal motion tends to destroy the orientation. We set up a Cartesian coordinate system (X, Y, Z) representing the laboratory and a Cartesian coordinate system (ζ, η, ξ) along the axes of the particle, as shown in Fig. 8.1. The external magnetic field is directed along the Z axis, and the moment fixed inside the particle m is in the direction of the ξ axis. By transforming $[\epsilon_p]$ from the coordinate system (ζ, η, ξ) to the coordinate system (X, Y, Z) , the dielectric tensor of the particle in the coordinate system (X, Y, Z) $[\epsilon_p]_{tr}$ can be obtained, as

$$[\epsilon_p]_{tr} = A[\epsilon_p]A^+, \quad (8.8)$$

where A is the transforming matrix of (ζ, η, ξ) to (X, Y, Z) . This expression is given by

$$A = \begin{bmatrix} \cos \psi \cos \phi - \cos \theta \sin \phi \sin \psi, & -\cos \psi \sin \phi - \cos \theta \cos \phi \sin \psi, & \sin \psi \sin \theta \\ \sin \psi \cos \phi + \cos \theta \sin \phi \cos \psi, & -\sin \psi \sin \phi + \cos \theta \cos \phi \cos \psi, & -\cos \psi \sin \theta \\ \sin \theta \sin \phi, & \sin \theta \cos \phi, & \cos \theta \end{bmatrix}. \quad (8.9)$$

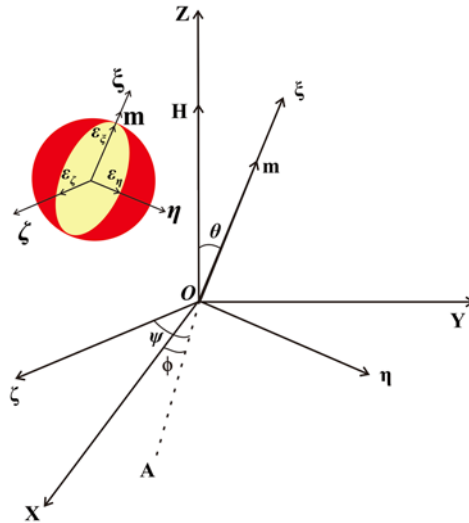


Fig. 8.1. The relationship between the (X, Y, Z) coordinate system and the (ζ, η, ξ) coordinate system.

A^+ is the adjoint matrix, and for the rotational transformation, $A_{ij}^+ = A_{ji}$. Assuming that a depolarizing field can be neglected in spherical magnetic particles, the dielectric tensor of the particles in the (X, Y, Z) coordinate system $[\epsilon]$ can be obtained by statistical averaging the dielectric tensor of the particle $[\epsilon_p]_{tr}$ with respect to the Euler angles ψ, ϕ and θ , i.e.

$$[\epsilon] = \left\langle [\epsilon_p]_{tr} \right\rangle_{\partial\psi\phi} = \frac{\int_0^{2\pi} d\psi \int_0^{2\pi} d\phi \int_0^\pi d(\cos\theta) [\epsilon_p]_{tr} \exp(\mu_0 m H \cos\theta / k_B T)}{\int_0^{2\pi} d\psi \int_0^{2\pi} d\phi \int_0^\pi d(\cos\theta) \exp(\mu_0 m H \cos\theta / k_B T)}, \quad (8.10)$$

where k_B is the Boltzmann constant and T is the absolute temperature. Performing the averaging calculation for all components of the matrix $[\epsilon_p]_{tr}$, the dielectric tensor $[\epsilon]$ is finally obtained as

$$[\epsilon] = \begin{bmatrix} \epsilon_{11} & 0 & 0 \\ 0 & \epsilon_{22} & 0 \\ 0 & 0 & \epsilon_{33} \end{bmatrix} + i \begin{bmatrix} 0 & \epsilon_{12} & 0 \\ -\epsilon_{12} & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad (8.11)$$

where

$$\begin{aligned} \epsilon_{11} &= \epsilon_{22} = (\epsilon_{11} + \epsilon_{22})(1 - L(\alpha)/\alpha)/2 + \epsilon_{33} L(\alpha)/\alpha, \\ \epsilon_{33} &= (\epsilon_{11} + \epsilon_{22})L(\alpha)/\alpha + \epsilon_{33}(1 - 2L(\alpha)/\alpha), \\ \epsilon_{12} &= \epsilon_{12} L(\alpha), \end{aligned}$$

and $\alpha = \mu_0 m H / k_B T$, $L(\alpha) = \coth(\alpha) - 1/\alpha$ is the Langevin function.

For colloidal system based on such crystal particles, it can be assumed that their intrinsic optical properties are unaffected by intrinsic magnetic properties and by an external magnetic field, since the magnetic permeability of the particles can be assumed to be unity using at optical frequency and magnetic perturbation have little effect on the optical proportion [3, 21]. Accordingly, we can consider only the first term in Eq. (8.5). Thus, Eq. (8.11) is simply described as

$$[\epsilon] = \begin{bmatrix} \epsilon_X & 0 & 0 \\ 0 & \epsilon_Y & 0 \\ 0 & 0 & \epsilon_Z \end{bmatrix}, \quad (8.12)$$

where

$$\begin{aligned} \epsilon_X &= \epsilon_Y = (\epsilon_\zeta + \epsilon_\eta)(1 - L(\alpha)/\alpha)/2 + \epsilon_\xi L(\alpha)/\alpha, \\ \epsilon_Z &= (\epsilon_\zeta + \epsilon_\eta)L(\alpha)/\alpha + \epsilon_\xi(1 - 2L(\alpha)/\alpha), \\ \epsilon_{ij} &= 0 (i \neq j), \end{aligned}$$

and $\epsilon_\zeta = \epsilon_{11}$, $\epsilon_\eta = \epsilon_{22}$ and $\epsilon_\xi = \epsilon_{33}$ are known as the principal dielectric constants, in whose respect, all crystals fall into three types, cubic, uniaxial and biaxial [26]. Thus, for a

particle having an intrinsic magnetic moment m parallel to a particular crystal axis, it can be regarded as a dielectric ellipsoid with intrinsic moment along the same axis, with semi-axis ε_ζ , ε_η and ε_ξ , as shown in Fig. 8.1. In addition, colloids are composed of the ellipsoids and the isotropic carrier liquid. The magneto-optical effects are determined by the orientation of the ellipsoids in the presence of an external magnetic field. Eq. (8.12) shows that the dielectric constant of the ellipsoid system can still be described with three principle values, $C_X (= C_{11})$, $C_Y (= C_{22})$ and $C_Z (= C_{33})$. From Eq. (8.12), we give the main results as follows:

- 1) When $H = 0$, $L(\alpha)/\alpha = 1/3$ and $C_X = C_Y = C_Z = (\varepsilon_\zeta + \varepsilon_\eta + \varepsilon_\xi)/3$. This means that for the ellipsoid system of any crystal type, the dielectric constant is completely isotropic in the absence of a magnetic field.
- 2) For a cubic crystal, since $\varepsilon_\zeta = \varepsilon_\eta = \varepsilon_\xi = \varepsilon$, which means that the anisotropic dielectric ellipsoid reverts an isotropic dielectric sphere, Eq. (8.12) shows that $C_X = C_Y = C_Z = \varepsilon$. Therefore for the system of a cubic crystal, the dielectric constant remains isotropic at optical frequency, independent of the magnetic field.
- 3) Under an externally applied magnetic field, beside cubic crystals, the dielectric constant of the ellipsoid system is anisotropic like that of the uniaxial crystal, i.e. $C_X = C_Y \neq C_Z$.

For the colloidal particle system, the colloidal particle dielectric constant is in relation to the carrier liquid [26]. Considering the inner and outer electric field of the colloidal particle in the coordinate system (ζ, η, ξ) , the following relation hold between inner electric field $E_k^{(i)}$ and outer electric field E :

$$(1 - n_k) \varepsilon_c E_k^{(i)} + n_k D_k^{(i)} = \varepsilon_c E_k, \quad k = \zeta, \eta, \xi, \quad (8.13)$$

where $E_k^{(i)}$ and E_k are the k -th components of $E^{(i)}$ and E , n_k is the electric depolarization factors of the particle, D_k is the k -th components of the electric displacement vector and ε_c is the dielectric constant of the carrier liquid. The inner electric field $E_k^{(i)}$ can be described as [26]

$$E_k^{(i)} = \frac{E_k}{1 + \left(\frac{\varepsilon_k^{(i)}}{\varepsilon_c} - 1 \right) n_k}, \quad (8.14)$$

where $\varepsilon_k^{(i)}$ is intrinsic dielectric constant of the particle, as Eq. (8.7) shown. Substituting the Eq. (8.13) into (8.14), $D_k^{(i)}$ can be obtained as:

$$D_k^{(i)} = \frac{\varepsilon_c \varepsilon_k^{(i)}}{\varepsilon_c (1 - n_k) + \varepsilon_k^{(i)} n_k} E_k. \quad (8.15)$$

Substituting $D_k = \varepsilon_k E_k$, the dielectric constant of the colloidal particle is given as

$$\varepsilon_k = \frac{\varepsilon_c \varepsilon_k^{(i)}}{\varepsilon_c (1 - n_k) + \varepsilon_k^{(i)} n_k}. \quad (8.16)$$

For spherical particle, $n_\zeta = n_\eta = n_\xi = 1/3$. Thus, Eq. (8.16) can be written as:

$$\varepsilon_k = \frac{3\varepsilon_c \varepsilon_k^{(i)}}{2\varepsilon_c + \varepsilon_k^{(i)}}. \quad (8.17)$$

Eqs. (8.16) and (8.17) show that the dielectric constant of the colloidal particle in Eq. (8.12) depends on not only the intrinsic dielectric constant of the particle $\varepsilon_k^{(i)}$, but also the dielectric constant of the carrier liquid ε_c .

8.2.3. Magneto-Optical Birefringence and Dichroism in Colloids

The dielectric constants of the particles are, in general, complex to describe the attenuation taking place when light propagates in the medium and can be written [27] as: $\varepsilon_\zeta = \varepsilon'_\zeta - i\varepsilon''_\zeta$, $\varepsilon_\eta = \varepsilon'_\eta - i\varepsilon''_\eta$, $\varepsilon_\xi = \varepsilon'_\xi - i\varepsilon''_\xi$, where the minus sign is used to make the observed ε''_ζ , ε''_η and ε''_ξ positive. The dielectric constants of the particle system can be written as

$$\begin{aligned} \mathbf{\epsilon}_X &= \mathbf{\epsilon}_Y = (\varepsilon'_\zeta + \varepsilon'_\eta)(1 - L(\alpha)/\alpha)/2 + \varepsilon'_\xi L(\alpha)/\alpha \\ &\quad - i[(\varepsilon''_\zeta + \varepsilon''_\eta)(1 - L(\alpha)/\alpha)/2 + \varepsilon''_\xi L(\alpha)/\alpha] \\ &= \mathbf{\epsilon}'_X(\mathbf{\epsilon}'_Y) - i\mathbf{\epsilon}''_X(\mathbf{\epsilon}''_Y), \\ \mathbf{\epsilon}_Z &= (\varepsilon'_\zeta + \varepsilon'_\eta)L(\alpha)/\alpha + \varepsilon'_\xi(1 - 2L(\alpha)/\alpha) \\ &\quad - i[(\varepsilon''_\zeta + \varepsilon''_\eta)L(\alpha)/\alpha + \varepsilon''_\xi(1 - 2L(\alpha)/\alpha)] \\ &= \mathbf{\epsilon}'_Z - i\mathbf{\epsilon}''_Z. \end{aligned} \quad (8.18)$$

After the application and removal of a magnetic field, the particles or ellipsoids with intrinsic moments return to equilibrium via two different mechanisms, namely Brownian rotation and Néel rotation [28]. Brownian rotation is connected with rotation of the particle itself within the carrier liquid, and has a characteristic relaxation time $\tau_B = 3V'\eta'/k_B T$, where V' is the hydrodynamic volume of the particle and η' is the dynamic viscosity of the carrier liquid. The Néel rotation is simply the rotation of the intrinsic moments, with a relaxation time $\tau_N = \tau_0 \exp(KV/k_B T)$, where τ_0 is usually approximated as 10^{-9} s, K is the anisotropy constant and V is the volume of the magnetic particle core. For the orientation of the dielectric ellipsoids by a magnetic field, only Brownian rotation is expected, and the Néel rotation does not produce any relaxation. The dominant re-orientation process of the magnetic moment is determined by the shortest characteristic time [29]. Thus, in Eq. (8.18), the parameter α should be multiplied formally by a weight factor $f = \tau_N/(\tau_N + \tau_B)$, to characterize the orienting power of the Brownian rotation. When $\tau_B \ll \tau_N$, $f = 1$ and the magnetically induced orientation is at its most effective. Such particles can be regarded as rigid dielectric ellipsoids. For magnetic nanoparticles, some local spins can be pinned to the surface [30] or the magnetic moments of the particles can

remain fixed with respect to the particles, due to the large anisotropy constant of the magnetic material itself [31]. This will prevent the moment from Néel rotating, so that the effective f value may be larger than the calculated value.

Eq. (8.18) also shows that the dielectric constants perpendicular to the field $\epsilon_{\perp} (= \epsilon_x, \epsilon_y)$ and parallel to the field $\epsilon_{\parallel} (= \epsilon_z)$ have different variations with the field strength. The refractive indices (n_{\perp}, n_{\parallel}) and absorption coefficients ($\kappa_{\perp}, \kappa_{\parallel}$) of the particle system in the two directions can be described as [26]

$$\begin{aligned} n_{\perp} &= \left[(\sqrt{\epsilon_x'^2 + \epsilon_x''^2} + \epsilon_x')/2 \right]^{1/2}, \\ \kappa_{\perp} &= \left[(\sqrt{\epsilon_x'^2 + \epsilon_x''^2} - \epsilon_x')/2 \right]^{1/2}, \end{aligned} \quad (8.19)$$

and

$$\begin{aligned} n_{\parallel} &= \left[(\sqrt{\epsilon_z'^2 + \epsilon_z''^2} + \epsilon_z')/2 \right]^{1/2}, \\ \kappa_{\parallel} &= \left[(\sqrt{\epsilon_z'^2 + \epsilon_z''^2} - \epsilon_z')/2 \right]^{1/2}. \end{aligned} \quad (8.20)$$

Clearly, n_{\perp}, n_{\parallel} and $\kappa_{\perp}, \kappa_{\parallel}$ should also have different variations with field strength. Such differences can lead to magneto-optical effects. For example, the difference between n_{\perp} and n_{\parallel} will cause birefringence, and the difference between κ_{\perp} and κ_{\parallel} will cause dichroism. The magneto-optical effects are a coupling action of both magnetic and dielectric properties. It is worth noting that the results in Eqs. (8.19) and (8.20) are independent of any interparticle interaction (as with ferrofluids) and the shape of the particles (such as liquid crystals).

Neglecting the effect of the carrier liquid, the birefringence Δn^* and the dichroism $\Delta \kappa^*$ of the magnetic colloid are written as

$$\Delta n^* = \Phi \delta n^* = \Phi (n_{\parallel} - n_{\perp}), \quad (8.21)$$

and

$$\Delta \kappa^* = \Phi \delta \kappa^* = \Phi (\kappa_{\parallel} - \kappa_{\perp}), \quad (8.22)$$

where $\delta n^* (= n_{\parallel} - n_{\perp})$, $\delta \kappa^* (= \kappa_{\parallel} - \kappa_{\perp})$ are called the reduced birefringence and reduced dichroism of the colloid. Generally, the dielectric constant ϵ for a solid is very difficult to calculate theoretically [26], so that the ϵ , which characterize the coupling action of both magnetic and dielectric properties for the particles system, are similarly difficult. In practice, optical anisotropy measurements were used to determine the magneto-dielectric anisotropy effects [27]. Moreover, from Eqs. (8.21) and (8.22), it is known that the coupling action can be revealed qualitatively by magneto-optical experiments. In other words, the coupling of the magnetic and dielectric properties of colloidal particles can produce magneto-optical effects.

8.3. Experiment Description and Results [32]

8.3.1. Sample Description

Goethite, i.e. α -FeOOH, has an orthorhombic crystal structure, in which the three principal values of the dielectric tensor, ϵ_ζ , ϵ_η and ϵ_ξ , are different and the positions of the principal axes coincide with the three crystal axes within the crystallographic unit cell [33]. Goethite is an antiferromagnetic material whose sublattice magnetization lies essentially along the $[0\ 1\ 0]$ direction [34]. Such nanoparticles, where the bulk is antiferromagnetic, can exhibit superparamagnetism or weak ferromagnetism, since net spins can exist on their surface [35]. In addition, when considering Fig. 8.1, Z and Y, and ξ and η should be exchanged and for Eq. (8.6), the Z should be replaced with Y, since the direction of the particle moments is along the η axis. The nanoparticles were prepared by chemical precipitation. Transmission electron microscopy (TEM) observations showed that the particles are spherical, and the diameter corresponding to the average volume [36] is about 8 nm. To make comparisons, spherical ZnFe_2O_4 nanoparticles and CoFe_2O_4 nanoparticles were produced by co-precipitation, and their diameters are about 5 nm and 12 nm, respectively. Typical TEM pictures are shown in Fig. 8.2 ZnFe_2O_4 and CoFe_2O_4 both have cubic crystal structures, and the bulk materials are anti-ferromagnetic and ferrimagnetic, respectively. Magnetization curves of these three nanoparticles were measured using a vibrating sample magnetometer (VSM) at room temperature, as shown in Fig. 8.3. Clearly, the magnetization of the α -FeOOH nanoparticles is the weakest of the three particles. The magnetization of the α -FeOOH nanoparticles is so weaker that the dielectric properties of a system based on such particles is very suitable to description using Eq. (8.12). For the ZnFe_2O_4 and CoFe_2O_4 nanoparticles, the coupling constants λ [37] are about 3.16×10^{-1} and 4.13, respectively. Therefore, only the CoFe_2O_4 nanoparticle system is able to form a chain-like aggregate structure by magnetic inter- action, because its coupling constant λ is larger than 2 [38]. The three aqueous colloids were synthesized by Massart's method [39], in which these ferrite nanoparticles were treated in $\text{Fe}(\text{NO}_3)_3$ solution and dispersed in nitric acid aqueous of $\text{pH} \approx 3$. The viscosity of the carrier liquid η' is 0.911 mPa. The anisotropy constant for α -FeOOH is $K = 6 \times 10^4 \text{ J/m}^3$ [40]. Therefore, for the α -FeOOH colloids, τ_B and τ_N are calculated at $T = 300 \text{ K}$ to be approximately $1.88 \times 10^{-7} \text{ s}$ and $6.16 \times 10^{-8} \text{ s}$, respectively, and $f = 0.25$.

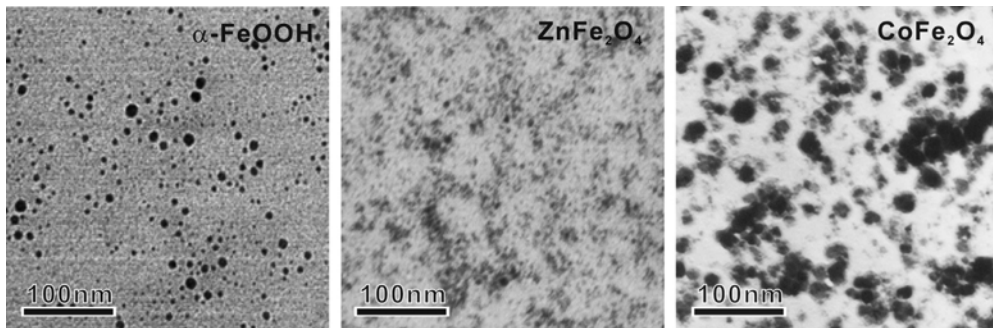


Fig. 8.2. Typical TEM pictures of α -FeOOH, ZnFe_2O_4 and CoFe_2O_4 colloidal particles.

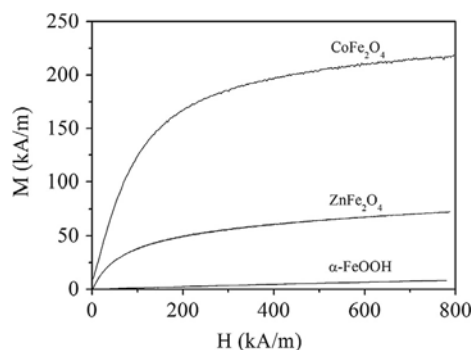


Fig. 8.3. Magnetization curves of the colloidal particles at room temperature.

8.3.2. The Optical Experiments and Results

8.3.2.1. The Optical Effect of Light Beam Perpendicular to Magnetic Field

The colloids were sealed into rectangular glass cells to form colloidal films of a thickness of about 0.3 mm. The incident light, from a He–Ne laser with wavelength 632.8 nm, was normal to both the applied magnetic field and the colloidal films. Since the measurement of the absolute transmission of the light can be greatly influenced by surface cleanliness of the glass of the sample [41], the normalized transmission was used to characterize the magneto-optical effect. This normalized transmission, or relative transmission coefficient, is defined as $T = (I_a/I_0)/(I_b/I_0)$, where I_0 is the intensity of the incident light and I_a (I_b) is the intensity of transmitted light after (before) the magnetic field is applied. Fig. 8.4 shows the measured results at $H = 400 \text{ kA}/4 \pi \mu\text{m}$ for the three colloids, when the polarization of the incident light is perpendicular and parallel to the direction of the magnetic field.

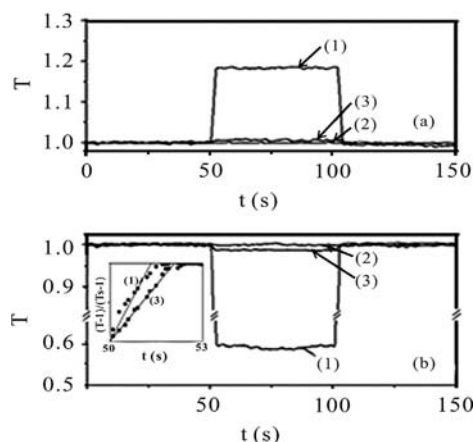


Fig. 8.4. The relative transmission intensity of (1) the α -FeOOH colloid; (2) the ZnFe_2O_4 colloid, and (3) the CoFe_2O_4 colloid. The external magnetic field $H = 400 \text{ kA}/4 \pi \mu\text{m}$ is applied over a period 50–100 s. The polarization of the incident light is (a) perpendicular, and (b) parallel to the direction of the magnetic field.

These results show that in the magnetic field, the colloid based on ZnFe_2O_4 nanoparticles of cubic crystal type with weak magnetism ($\lambda = 3.16 \times 10^{-1} < 2$) is optically isotropic, but the colloid based on $\alpha\text{-FeOOH}$ nanoparticles of orthorhombic crystal type, with weaker magnetism than the ZnFe_2O_4 particles, is optically anisotropic. Because the magnetization of the $\alpha\text{-FeOOH}$ particles is far weaker than that of the ZnFe_2O_4 particles, it is concluded that the $\alpha\text{-FeOOH}$ nanoparticles in the colloid cannot form any aggregates as a result of magnetic interaction between the particles to produce the optical effect, and the magneto-dichroism of the $\alpha\text{-FeOOH}$ colloid clearly arises from the orientation of the dielectric ellipsoids. It can also be seen in Fig. 8.4 that the dichroism of the $\alpha\text{-FeOOH}$ colloid is even more apparent than that of the CoFe_2O_4 ferrofluid.

For comparison of the switching times Δt of both the CoFe_2O_4 and the $\alpha\text{-FeOOH}$ colloids, the curves of $(T-1)/(T_s-1)$ vs. t were drawn from $T = 1$ to $T = T_s$ (see the inset in Fig. 8.4), where T_s is the static relative transmission intensity after the application of the magnetic field. From the curves, the switching times Δt were estimated to be about 2.1 s and 1.3 s for the CoFe_2O_4 and the $\alpha\text{-FeOOH}$ colloids, respectively. For the $\alpha\text{-FeOOH}$ colloid, the estimated switching time Δt , which is close to the roughly estimated Brownian relaxation time of 0.52 s (obtained by measurement of the magnetization of a colloid based on magnetic iron oxide nanoparticles [42]), is much longer than the calculated Brownian relaxation time τ_B , and the difference may arise from the influence of the distribution of particle sizes, the hydrodynamic interaction between particles on the Brownian relaxation mechanism [28, 43] and the delay time of field switching [44]. Even though the estimated switching times are only approximate, it is known from these values that after application of an external magnetic field, the response of the magneto-optical effects of the $\alpha\text{-FeOOH}$ colloid will be faster than that of the CoFe_2O_4 colloid.

For the $\alpha\text{-FeOOH}$ colloid, the angular distribution of the relative intensity of the transmitted light T was examined by orienting the polarization direction of the incident light P at different angles θ corresponding to the magnetic field H (see the inset in Fig. 8.5). The results are shown in Fig. 8.5.

It was found that the distribution of T vs. θ is a quasi-ellipse whose short axis is along the direction of the magnetic field, i.e. the minimum T_{\min} is at $\theta = 0^\circ$ or 180° , and the long axis is normal to the magnetic field, i.e. the maximum T_{\max} is at $\theta = 90^\circ$. In addition, using circularly polarized light to illuminate the $\alpha\text{-FeOOH}$ colloid sample, the value of T was measured at different angles parallel to the magnetic field with an analyzer polarizer, in zero field and in a magnetic field of $400 \text{ kA}/4 \pi \text{ m}$. The results are also shown in Fig. 8.5. Clearly, in the absence of a magnetic field, the $\alpha\text{-FeOOH}$ colloid is optically isotropic and the transmitted light is still circularly polarized, i.e. T is constant at different θ . When the external magnetic field was applied, the colloid became optically anisotropic and the transmitted light was elliptically polarized, with its intensity exhibiting a quasi-elliptical distribution similar to the result found when linearly polarized light illuminates the sample. However, the short axis was not along the direction of the magnetic field, deviating by an angle $\Delta\theta$ of about 20° (see Fig. 8.5). This characterizes the magneto-dichroism as well as the birefringence and can be explained as follows.

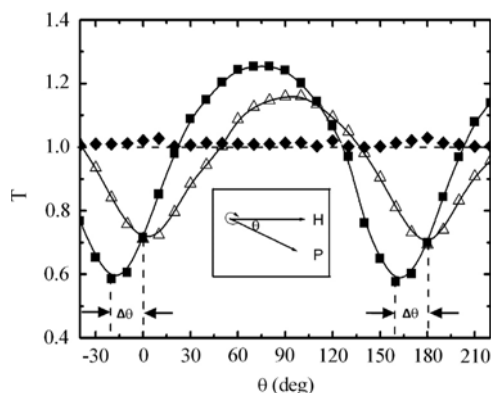


Fig. 8.5. The angular distribution of the relative transmitted intensity T of linearly and circularly polarized light illuminating the α -FeOOH colloid. $T = I_\theta/I_0$, where I_0 is the intensity of transmitted polarized light at the angle θ between the polarization direction of the incident light \mathbf{P} (illuminating with linearly polarized light) or the polarization direction of the analyzer \mathbf{P} (illuminating with circularly polarized light) and the external magnetic field \mathbf{H} (as shown in the inset). I_0 is the intensity before the field is applied. Δ -illuminating with linearly polarized light at $H = 400 \text{ kA}/4 \pi \text{ m}$. \diamond , \blacksquare -illuminating with circularly polarized light at $H = 0$ and at $H = 400 \text{ kA}/4 \pi \text{ m}$, respectively.

The incident circularly polarized light can be divided into two beams of linearly polarized light with the same amplitudes and a $\pi/2$ phase difference, their electric vectors being perpendicular (E_\perp) and parallel (E_\parallel) to the field, i.e. $E_\perp = A_\perp \cos(\omega t)$ and $E_\parallel = A_\parallel \cos(\omega t + \pi/2)$. When the circularly polarized light is transmitted through the α -FeOOH colloidal sample under the influence of a magnetic field, the amplitudes of the two beams of polarized light will become different due to the dichroism and their phase difference will be increased by an amount $\Delta\tau$ due to the birefringence [45]. Thus, the transmitted light can be described as $E_\perp = A_\perp \cos(\omega t)$ and $E_\parallel = A_\parallel \cos(\omega t + \pi/2 - \tau)$, where $A_\perp \neq A_\parallel$ and $\Delta\tau > 0$. Accordingly, the transmitted light is an elliptically polarized beam whose short axis deviates from the direction of the applied magnetic field.

For a α -FeOOH colloid with $\Phi = 0.15 \%$, the relationships between δn^* , $\delta \kappa^*$ and the magnetic field H were obtained using circularly polarized light, as shown in Fig. 8.6. Clearly, the δn^* and $\delta \kappa^*$ increase with the magnetic field.

8.3.2.2. The Optical Effect of Light Beam Parallel to Magnetic Field

In order to prove that the optical effects of α -FeOOH colloids are related to the coupling of both magnetic and dielectric properties, rather than the aggregation of the colloidal particles, the behavior of light transmitted parallel to the field direction was measured. The results are shown in Fig. 8.7, from which it can be seen that for CoFe_2O_4 ferrofluids, the transmittance exhibited a nonlinear relaxation process which is attributed to motion of the chains [46], but for α -FeOOH colloids, the transmittance was clearly enhanced. Clearly, the behavior of α -FeOOH colloids results from the dielectric properties of the colloidal particles and can be explained as follows.

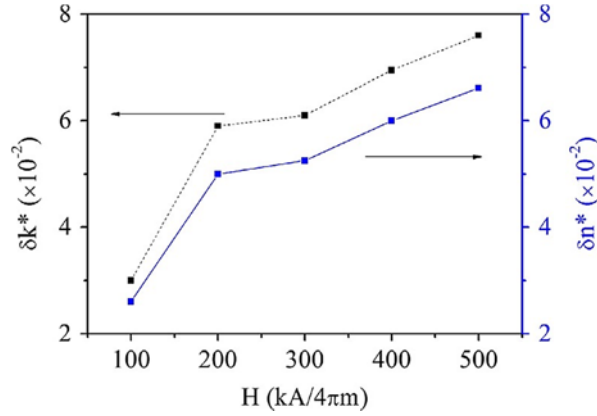


Fig. 8.6. The relationships between δn^* , $\delta \kappa^*$ and magnetic field H for the α -FeOOH colloid with $\Phi = 0.15\%$.

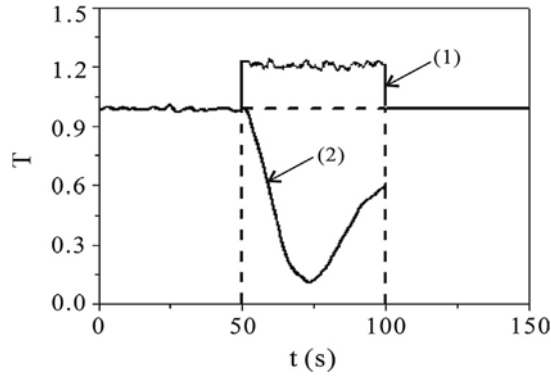


Fig. 8.7. The relative transmission intensity of (1) the α -FeOOH colloid and (2) the CoFe_2O_4 colloid. The direction of the incident light is parallel to the magnetic field. The external magnetic field $H (= 300 \text{ kA}/4 \pi m)$ is applied over a period 50–100 s.

The α -FeOOH particle size parameter is $\Gamma = (2\pi d/\lambda) < 1$, so the particles can be regarded as absorbing Rayleigh scatters [47], in which the scattering cross-section σ_{scat} is much less than the absorption section σ_{abs} and the transmittance depends mainly on σ_{abs} [48].

The Rayleigh absorption cross-section is

$$\sigma_{\text{abs}} = -\frac{8\pi d^3}{\lambda} \text{Im} \left(\frac{\tilde{n}_p^2 - \tilde{n}_c^2}{\tilde{n}_p^2 + 2\tilde{n}_c^2} \right), \quad (8.23)$$

where \tilde{n}_c , \tilde{n}_p are the complex refractive indices of both the carrier liquid and the particles, and described by $\tilde{n}_c^2 = \epsilon'_c - i\epsilon''_c$ and $\tilde{n}_p^2 = \epsilon'_p - i\epsilon''_p$, respectively, and Im indicates the imaginary part. For the aqueous carrier liquid, the imaginary part of the dielectric constant may be neglected, i.e. $\tilde{n}_c^2 = n_c^2 = \epsilon'_c = \epsilon_c$.

While the light propagates along the direction of magnetization, it is always polarized perpendicular to the magnetization and its absorption should thus be described by the dielectric constant perpendicular to the direction of an external field. For a moderate field, $L(\alpha) = \alpha/3 - \alpha^3/45$, the field is defined as being in the Y direction since the magnetization of α -FeOOH lies essentially along [0 1 0] (the η direction). Thus, the dielectric constant perpendicular to the field can be described from Eq. (8.18) by

$$\begin{aligned}\epsilon_p &= \epsilon_x (= \epsilon_z) = \epsilon'_x - i\epsilon''_x \\ &= (\epsilon'_\zeta + \epsilon'_\eta) \left(\frac{2}{3} + \frac{\alpha^2}{45} \right) / 2 + \epsilon'_\eta \left(\frac{1}{3} - \frac{\alpha^2}{45} \right) \\ &\quad - i \left[(\epsilon''_\zeta + \epsilon''_\eta) \left(\frac{2}{3} + \frac{\alpha^2}{45} \right) / 2 + \epsilon''_\eta \left(\frac{1}{3} - \frac{\alpha^2}{45} \right) \right].\end{aligned}\quad (8.24)$$

Consequently, when the light is transmitted through the α -FeOOH colloidal film along the direction of the applied magnetic field, the absorption cross-section can be written as

$$\begin{aligned}\sigma_{abs} &= \frac{8\pi d^3}{\lambda} \times \frac{3\epsilon''_p \epsilon_c}{(\epsilon'_p + 2\epsilon_c)^2 + \epsilon''_p^2} = \frac{8\pi d^3}{\lambda} \times \frac{3\epsilon''_p \epsilon_c}{(\epsilon'_x + 2\epsilon_c)^2 + \epsilon''_x^2} = \frac{8\pi d^3}{\lambda} \\ &\quad \times \frac{\epsilon_c \left[\epsilon''_\zeta + \epsilon''_\eta + \epsilon''_\zeta + \left(\alpha^2 (\epsilon''_\zeta + \epsilon''_\eta - 2\epsilon''_\eta) / 30 \right) \right]}{\left[(1/3)(\epsilon'_\zeta + \epsilon'_\eta + \epsilon'_\zeta) + \left(\alpha^2 (\epsilon'_\zeta + \epsilon'_\eta - 2\epsilon'_\eta) / 90 \right) + 2\epsilon_c \right]^2 + \left[(1/3)(\epsilon''_\zeta + \epsilon''_\eta + \epsilon''_\zeta) + \left(\alpha^2 (\epsilon''_\zeta + \epsilon''_\eta - 2\epsilon''_\eta) / 90 \right) \right]^2}.\end{aligned}\quad (8.25)$$

Although the details of the dielectric constant of the α -FeOOH particle are unknown so that σ_{abs} is difficult to be calculated theoretically, Eq. (8.25) shows that the σ_{abs} is reduced with increasing magnetic field because, in the equation, the numerator and denominator increase with $\alpha^2(\propto H^2)$ and $\alpha^4(\propto H^4)$, respectively.

In addition, for a cubic crystal, $\epsilon_\zeta = \epsilon_\eta = \epsilon_\xi = \epsilon$, i.e. $\epsilon'_\zeta = \epsilon'_\eta = \epsilon'_\xi = \epsilon'$ and $\epsilon''_\zeta = \epsilon''_\eta = \epsilon''_\xi = \epsilon''$. Thus, according to Eq. (8.23), the absorption cross-section of a magnetic colloidal system based on cubic crystal nanoparticles, along the direction of the magnetic field, can be written as

$$\sigma_{abs} = \frac{8\pi d^3}{\lambda} \times \frac{3\epsilon_c \epsilon''}{(\epsilon' + 2\epsilon_c)^2 + \epsilon''^2}.\quad (8.26)$$

Obviously, in a model of single particles, the absorption cross-section of such a system is independent of the magnetic field. So, it can be further determined that for ferrofluids based on spherical CoFe_2O_4 nanoparticles, with cubic crystal structure, the behavior of the transmitted light can result only from the kinetics of aggregation and de-aggregation while switching the field on and off [49] rather than the orientation of individual particles.

8.4. Conclusions

In linear optics, optical effects are independent of intensity of incident light, and dielectric constant can be used to describe the propagation behavior of light in the media. Both theoretical and experimental results indicate that the colloids based on non-cubic crystal nanoparticles with weak intrinsic moments are equivalent to a dielectric ellipsoid system and exhibit optical uniaxial anisotropy due to the orientational order of the ellipsoids by a magnetic field. Such mechanism of field-induced optical effects is different from the well-known magneto-optical effect, which is proportional to the magnetization or to the magnetic field strength applied to a magnetic nanoparticles, film, etc. Instead of observing of a magneto-optical effect, the present effects are just results observing a collective optical effect exclusively coming from the optical anisotropy of the colloidal particles, which has a non-cubic crystal structure. Consequently, being magnetic at the same time, all the colloidal particles align its anisotropic axis with respect to the magnetic field and contribute to the optical change, whereas this does not happen with the other magnetic colloid particles that have cubic crystal structure. Since there is no structural influence on the orientation of the dielectric ellipsoids by a magnetic field, this relaxation time could be far shorter than those for ferrofluids and liquid crystals.

Until now, the physical origin of the magneto-optical effects in ferrofluids remained controversial and two opposing theories have been put forward to explain the linear magneto-induced anisotropy [50]. On the one hand, this phenomenon could result from orientation by the applied magnetic field of single particles, or of pre-existing aggregates, and on the other hand, it might be due to a cooperative effect brought about by induced aggregation under such a field. The theory presented in this work shows that for magnetic colloid based on spherical nanoparticles with cubic crystal structure, such as are employed in most of the experiments on ferrofluids, the magneto-optical effects at optical frequency could not be due to the reaction of single particles. The transition from orientation disorder to order of the dielectric ellipsoids could also take place in ferrofluids based on non-cubic ferromagnetic /ferrimagnetic nanoparticles, but for these ferrofluids similar effects could be easily concealed by the magneto-optical effects of field-induced aggregates.

In addition, the results corresponding to CoNi_{1-x} nanoparticles show that smaller nanoparticles result in lower dielectric constant due to quantum size effects [51]. However, comparing experimental and theoretical results based on bulk properties, α -FeOOH nanoparticles have refractive index approximately 10^3 times larger than that of the bulk; according, dielectric constant could be approximately 10^6 times larger in the optical-frequency regime [19]. This better dielectric properties of α -FeOOH nanoparticles are worth to be investigated further.

Acknowledgement

This work has been supported by the National Nature Science Foundation Project of PR China.

References

- [1]. S. L. Gómez, F. L. S. Cuppo, A. M. Figueiredo Neto, T. Kosa, M. Muramatsu, R. J. Horowicz, Z-scan measurement of the nonlinear refractive indices of micellar lyotropic liquid crystals with and without the ferrofluid doping, *Phys. Rev. E*, Vol. 59, 1999, pp. 3059-3063.
- [2]. J. F. Chen, X. F. Chen, S. L. Pu, Z. Y. Di, Y. X. Xia, Realization of optical limiting with a magnetic fluid film, *Opt. Commun.*, Vol. 276, 2007, pp. 268-271.
- [3]. M. Xu, P. J. Ridler, Linear dichroism and birefringence effects in magnetic fluids, *J. Appl. Phys.*, Vol. 82, 1997, pp. 326-332.
- [4]. K. Butter, P. Bomans, P. Frederik, G. Vroege, A. Philipse, Direct observation of dipolar chains in iron ferrofluids by cryogenic electron microscopy, *Nat. Mater.* 2, Vol. 33, 2003, pp. 88-91.
- [5]. S. Taketomi, Aggregation of magnetic fluids under an external field: micelle formation: a review, *Jordan J. Phys.*, Vol. 4, 2001, pp. 1-37.
- [6]. J. Li, D. Li, The Bidispersed Model System and Binary System for Ferrofluids, *Phys. Int.* Vol. 3, 2012, pp. 28-43.
- [7]. E. Hasmonay, E. Dubois, J.-C. Bacri, R. Perzynski, Y. L. Raikher, V. Stepanov, Static magneto-optical birefringence of size-sorted $\gamma\text{-Fe}_2\text{O}_3$ nanoparticles, *Eur. Phys. J. B*, Vol. 5, 1998, pp. 859-867.
- [8]. D. Jamon, F. Donatini, J. Monin, M. Rasa, V. Socoliuc, O. Filip, D. Bica, V. Sofonea, Concentration dependence of magnetisation and magneto-optical effects in a ferrofluid with double layer stabilized particles, *J. Magn. Magn. Mater.*, Vol. 201, 1999, pp. 174-177.
- [9]. Y. N. Skibin, V. Chekanov, Y. L. Raizer, Birefringence in a ferromagnetic liquid, *Zh. Eksp. Teor. Fiz.*, Vol. 72, 1977, pp. 949-855.
- [10]. P. Scholten, The origin of magnetic birefringence and dichroism in magnetic fluids, *IEEE Trans. on Magn.*, Vol. 16, 1980, pp. 221-225.
- [11]. J. M. Laskar, S. Brojabasi, B. Raj, J. Philip, Comparison of light scattering from self assembled array of nanoparticle chains with cylinders, *Opt. Commun.*, Vol. 285, 2012, pp. 1242-1247.
- [12]. J. Philip, J. M. Laskar, Optical Properties and Applications of Ferrofluids-A Review, *J. Nanofluids*, Vol. 1, 2012, pp. 3-20.
- [13]. T. Stinson, J. Litster, Pretransitional phenomena in the isotropic phase of a nematic liquid crystal, *Phys. Rev. Lett.*, Vol. 25, 1970, p. 503-506.
- [14]. B. J. Lemaire, P. Davidson, J. Ferré, J. P. Jamet, P. Panine, I. Dozov, J. P. Jolivet, Outstanding magnetic properties of nematic suspensions of Goethite ($\alpha\text{-FeOOH}$) nanorods, *Phys. Rev. Lett.*, Vol. 88, 2002, 125507.
- [15]. J. C. P. Gabriel, P. Davidson, New trends in colloidal liquid crystals based on mineral moieties, *Adv. Mater.*, Vol. 12, 2000, pp. 9-20.
- [16]. P. Fernandes, H. Mukai, I. Laczowski, Magneto-optical effect in lyotropic liquid crystal doped with ferrofluid, *J. Magn. Magn. Mater.*, Vol. 289, 2005, pp. 115-117.
- [17]. H. Mukai, S. M. Shibli, P. R. G. Fernandes, Orientational order studies by magneto-optical and light-effects in lyotropic liquid crystal, *J. Mol. Liq.*, Vol. 135, 2007, pp. 53-56.
- [18]. J. Li, A. Wang, Y. Lin, X. Liu, J. Fu, L. Lin, L. Chen, The increase of the light transparency induced by a magnetic field for the colloid film based on $\alpha\text{-FeOOH}$ nanoparticles, *Opt. Mater. Express*, Vol. 2, 2012, pp. 1760-1767.
- [19]. J. Li, X. Qiu, Y. Lin, L. Chen, X. Liu, D. Li, Large magneto-optical birefringence of colloidal suspensions of $\alpha\text{-FeOOH}$ goethite nanocrystallites, *Chem. Phys. Lett.*, Vol. 590, 2013, pp. 165-168.
- [20]. P. Pershan, Magneto-optical effects, *J. Appl. Phys.*, Vol. 38, 1967, pp. 1482-1490.
- [21]. G. A. Smolenskiĭ, R. V. Pisarev, I. G. Siniĭ, Birefringence of light in magnetically ordered crystals, *Sov. Phys. Usp.* Vol. 18, 1976, pp. 410-429.

- [22]. A. K. Zvezdin, V. A. Kotov, Modern Magneto-optics and Magneto-optical Materials, *IOP Publishing Ltd.*, 1997.
- [23]. J. F. Dillon, Magneto-optics and its uses, *J. Magn. Magn. Mater.* Vol. 31-34, 1983, pp. 1-9.
- [24]. A. Satoh, Y. Sakuda, Rheology and orientational distributions of rodlike particles with magnetic moment normal to the particle axis for semi-dense dispersions (analysis by means of mean field approximation), *J. Colloid Interface Sci.*, Vol. 308, 2007, pp. 532-541.
- [25]. S. Chung, A. Hoffmann, S. Bader, C. Liu, B. Kay, L. Makowski, L. Chen, Biological sensors based on Brownian relaxation of magnetic nanoparticles, *Appl. Phys. Lett.*, Vol. 85, 2004, pp. 2971-2973.
- [26]. L. D. Landau, E. M. Lifšic, L. P. Pitaevskii, Electrodynamics of Continuous Media, *Butterworth-Heinemann Ltd.*, 1995.
- [27]. N. A. Yusuf, I. O. Abu-Aljarayesh, Magneto-optical and magneto-dielectric anisotropy effects in magnetic fluids, *Jordan J. Phys.*, Vol. 2, 2009, pp. 1-46.
- [28]. R. Kötitz, P. Fannin, L. Trahms, Time domain study of Brownian and Néel relaxation in ferrofluids, *J. Magn. Magn. Mater.*, Vol. 149, 1995, pp. 42-46.
- [29]. E. Hasmonay, E. Dubois, S. Neveu, J.-C. Bacri, R. Perzynski, Alternating magneto-birefringence of ionic ferrofluids in crossed fields, *Eur. Phys. J. B*, Vol. 21, 2001, pp. 19-29.
- [30]. Y.-W. Du, J. Wu, H.-X. Lu, T.-X. Wang, Z.-Q. Qiu, H. Tang, J. Walker, Magnetic properties of fine iron particles, *J. Appl. Phys.*, Vol. 61, 1987, pp. 3314-3316.
- [31]. S. Neveu-Prin, F.A. Tourinho, J.-C. Bacri, R. Perzynski, Magnetic birefringence of cobalt ferrite ferrofluids, *Colloids and Surfaces A: Physicochemical and Engineering Aspects*, Vol. 80, 1993, pp. 1-10.
- [32]. L.-L. Chen, J. Li, X.-Y. Qiu, Y.-Q. Lin, X.-D. Liu, H. Mao, J. Fu, Magneto-optical effect in a system of colloidal particle having anisotropic dielectric property, *Opt. Comm.*, Vol. 316, 2014, pp. 146-151.
- [33]. M. Born, E. Wolf, Chapter 15, in Principles of Optics, *Cambridge University Press*, Cambridge, UK, 1999.
- [34]. D. E. Madsen, L. Cervera-Gontard, T. Kasama, R. E. Dunin-Borkowski, C. Koch, M. F. Hansen, C. Frandsen, S. Mørup, Magnetic fluctuations in nanosized goethite (α -FeOOH) grains, *J. Phys. Condens. Matter*, Vol. 21, 2009, 016007.
- [35]. W. J. Schuele, V. D. Deetscreek, Appearance of a weak ferromagnetism in fine particles of antiferromagnetic materials, *J. Appl. Phys.*, Vol. 33, 1962, pp. 1136-1138.
- [36]. C. G. Granqvist, R. A. Buhrman, Ultrafine metal particles, *J. Appl. Phys.*, Vol. 47, 1976, pp. 2200-2219.
- [37]. S. Han, J. Li, R. Gao, T. Zhang, B. Wen, The modification effect in magnetization behaviors for CoFe_2O_4 - $\text{p-NiFe}_2\text{O}_4$ binary ferrofluids, *Appl. Phys. A*, Vol. 98, 2010, pp. 179-186.
- [38]. T.-Z. Zhang, J. Li, H. Miao, Q.-M. Zhang, J. Fu, B.-C. Wen, Enhancement of the field modulation of light transmission through films of binary ferrofluids, *Phys. Rev. E*, Vol. 82, 2010, 021403.
- [39]. R. Massart, Preparation of aqueous magnetic liquids in alkaline and acidic media, *IEEE Trans. Magn.*, Vol. 17, 1981, pp. 1247-1248.
- [40]. J. M. D. Coey, A. Barry, J. Broto, H. Rakoto, S. Brennan, W.N. Mussel, A. Collomb, D. Fruchart, Spin flop in goethite, *J. Phys. Condens. Matter*, Vol. 7, 1995, pp. 759-768.
- [41]. S. Taketomi, M. Ukita, M. Mizukami, H. Miyajima, S. Chikazumi, Magneto-optical effects of magnetic fluid, *J. Phys. Soc. Jpn.*, Vol. 56, 1987, pp. 3362-3374.
- [42]. R. Kötitz, W. Weitschies, L. Trahms, W. Semmler, Investigation of Brownian and Néel relaxation in magnetic fluids, *J. Magn. Magn. Mater.*, Vol. 201, 1999, pp. 102-104.
- [43]. J. Zhang, C. Boyd, W. Luo, Two mechanisms and a scaling relation for dynamics in ferrofluids, *Phys. Rev. Lett.*, Vol. 77, 1996, pp. 390-393.
- [44]. J. Li, X. Qiu, Y. Lin, X. Liu, J. Fu, H. Miao, Q. Zhang, T. Zhang, Oscillatory-like relaxation behavior of light transmitted through ferrofluids, *Appl. Opt.*, Vol. 50, 2011, pp. 5780-5787.

- [45]. J. Fu, J. Li, Y. Lin, X. Liu, H. Miao, L. Lin, Study of magneto-optical effects in γ -Fe₂O₃/ZnFe₂O₄ nanoparticle ferrofluids, using circularly polarized light, *Sci. China Phys. Mech. Astron.*, Vol. 55, 2012, pp. 1404-1411.
- [46]. J. Li, Y. Huang, Q. Li, Coordinated chain notion resulting in intensity variation of light transmitted through ferrofluid film, *Phys. Lett. A*, Vol. 372, 2008, pp. 6952-6955.
- [47]. D. Lacoste, F. Donatini, S. Neveu, J. Serughetti, B. Van Tiggelen, Photonic Hall effect in ferrofluids: Theory and experiments, *Phys. Rev. E*, Vol. 62, 2000, pp. 3934-3943.
- [48]. C. M. Soreff, Scattering and absorption of light by particles and aggregates, in *Handbook of Surface and Colloidal Chemistry*, CRC Press, Boca Raton, Florida, 1997.
- [49]. J. M. Laskar, J. Philip, B. Raj, Experimental investigation of magnetic-field-induced aggregation kinetics in nonaqueous ferrofluids, *Phys. Rev. E*, Vol. 82, 2010, 021402.
- [50]. D. Jamon, F. Donatini, A. Siblani, F. Royer, R. Perzynski, V. Cabuil, S. Neveu, Experimental investigation on the magneto-optic effects of ferrofluids via dynamic measurements, *J. Magn. Mater.*, Vol. 321, 2009, pp. 1148-1154.
- [51]. S. Hussain, I. J. Youngs, J. J. Ford, The electromagnetic properties of nanoparticle colloids at radio and microwave frequencies, *J. Phys. D: Appl. Phys.* Vol. 40, 2007, pp. 5331-5337.

Chapter 9

General Overview of Coherent X-Ray Diffraction Imaging and Ptychography and Their Developments and Applications

Xiaowen Shi

9.1. Coherent X-Ray Diffraction Imaging

9.1.1. Overview and Introduction

Coherent X-ray diffraction imaging (CXDI) is a lensless imaging technique that utilizes both absorption and phase contrasts for studying objects of all length-scales, including nano-crystals and micron-sized biological chromosomes. Unlike most of other real-space imaging techniques that usually require aggressive sample sectioning to thin them down to a few nano-meters, which may result to alteration of sample original internal structures; CXDI is a non-destructive technique that often requires minimum amount of sample preparation and in principle, it is a diffraction-limited technique, which sets its ultimate spatial resolution to 1 nm to 1 Å in soft and hard X-rays regimes respectively. Conventional real-space X-ray microscopy methods are limited by the X-ray probe size that is determined by the availability of nano-size focusing optics. The smallest focused X-ray probe size is usually limited by the manufacture of high-quality X-ray mirrors and Electron Beam lithography (EBL). Currently, X-rays mirrors can produce high-intensity focused X-ray to around 1 μm and diffraction optics such as Fresnel Zone Plate (FZP) can focus X-ray to size of around 20 -30 nm. CXDI is expected to offer at least a factor of three better spatial resolution than real-space Transmission X-ray Microscopy (TXM), in both soft and hard X-rays. Fig. 9.1 illustrates some typical CXDI diffraction patterns.

CXDI requires sufficient degree of coherence of X-rays, both in longitudinal and transverse directions of synchrotron X-rays because the technique heavily relies on constructive interference of coherent diffraction intensities to be detected on X-ray detector downstream to collect oversampled data. The general definitions of longitudinal

and transverse coherence lengths within the framework of a multi-dimensional coherence function are as follows:

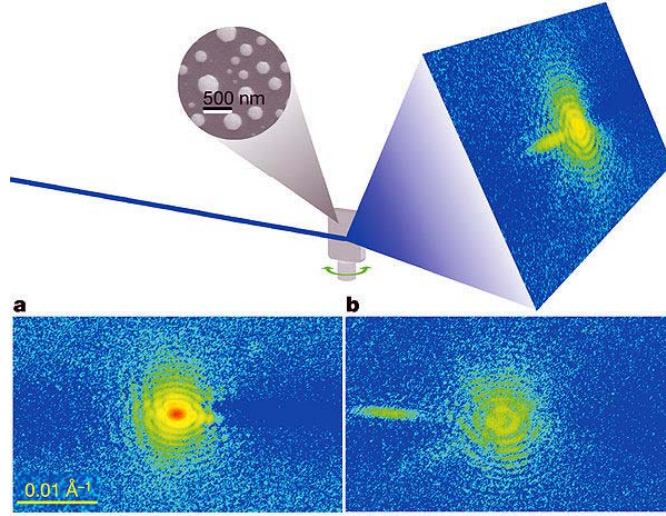


Fig. 9.1. Image is from [1] Typical CXDI experimental measurements of Pb nanoparticles, a and b images are two frames of the 3D rocking data frame collection, with scale-bar shown here representing the reciprocal-space lattice vector.

$$\xi_L = \frac{\lambda^2}{2\Delta\lambda}, \quad (9.1)$$

$$\xi_T = \frac{\lambda R}{2\sigma_{x,y}}, \quad (9.2)$$

where λ is the wavelength of incident X-ray and $\Delta\lambda$ is the difference in wavelength between different incident waves, and $\sigma_{x,y}$ is the finite source sizes of accelerated electrons from which synchrotron X-rays are produced, R being the distance the incident X-ray travels and it is considered that $R \gg \sigma_{x,y}$. The distance, at which two incident X-ray waves are just totally out of phase, i.e. out-of-phase by radians of, defines coherence length $\frac{\pi}{2}$.

The mutual coherence function (MCF) of the electric field is generally defined as:

$$\Gamma(\rho_1, \rho_2, \tau) = \langle E(\rho_1, t) E^*(\rho_2, t + \tau) \rangle. \quad (9.3)$$

And the normalized mutual coherence function is usually described as the follows:

$$\gamma(\rho_1, \rho_2, \tau) = \frac{\langle E(\rho_1, t) E^*(\rho_2, t + \tau) \rangle}{\sqrt{\langle I(\rho_1, t) \rangle \langle I(\rho_2, t) \rangle}}, \quad (9.4)$$

ρ_1 and ρ_2 are the position vectors of two points at a particular wave front in which two points are positioned. $E(\rho_1, t)$ and $E^*(\rho_2, t + \tau)$ are electromagnetic fields at these two points which are separated by a specific time interval of τ , after some time t in the initial starting point of measurement. $\langle I(\rho_1, t) \rangle$ and $\langle I(\rho_2, t) \rangle$ are expectation values (statistically average) of intensity of electromagnetic field at these two points in the wave front. The value $\gamma(\rho_1, \rho_2, \tau)$ is defined to be between 0 and 1, with 0 being complete incoherence and 1 being perfect coherence, which means any values of $\gamma(\rho_1, \rho_2, \tau)$ being between 0 and 1 indicate that the X-ray source is somewhat partially coherent, and this is the case for real-life synchrotron radiation X-ray sources that are available for measurement.

The “Phase problem” has been an extensive research field since the middle of the last century, scientists from various research fields such as applied mathematics and astronomy have been actively involved in solving “Phase problem” from various approaches. Due to the difficulties of solving the “Phase problem” directly, CXDI has promised to mathematically determine the phase solutions by using the theorem based on oversampling criterion. In CXDI, if recorded diffraction data satisfy oversampling conditions [2, 3], they contain enough information to solve the phase problem in most cases, however, cases that contain exceptional elementary symmetries are relatively non-trivial to have general solutions. In theory, CXDI should record Oversampled measured data, which means the measured data should be recorded at a sampling frequency that is at least twice the maximum frequency of the reciprocal-space information, that is the usually defined as the sampling frequency of the reciprocal-space lattice point in a crystal is a sample is considered to be a single crystal structure. Sayre proposed the theorem that to ensure successful data reconstructions; at least a factor of 2 oversampling data has to be obtained⁴ owing to the fact that both absorption and phase image are to be reconstructed. In practice, the general consensus is that the minimum oversampling ratios required are $\sqrt{2}$ and $\sqrt[3]{2}$ for 2D and 3D coherent diffraction intensities respectively.

9.1.2. Bragg CXDI

Bragg CXDI utilises the concept of Bragg diffraction of single crystals to study the structures of the crystals. Fig. 9.2 shows a typical Bragg diffraction in crystals.

Typically crystal shape functions are complex, and they represent 3D electronic structures within the crystals, where the phase component measures displacements fields inside the finite-sized crystals and can be converted into measured phases by dot product of Fourier space Q vector of Bragg reflection and the displacements at specific location. The displacements fields measure amount of deviation of crystal lattices from their ideal lattice

points. This concept is reviewed by Robinson [2] and can be quantitatively calculated by the following equation:

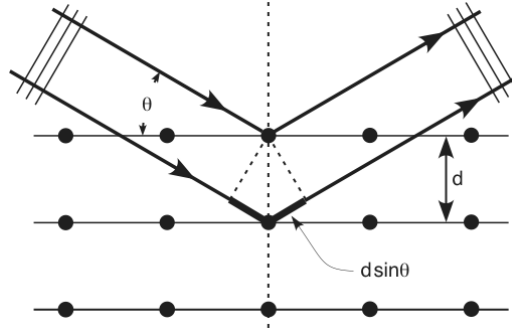


Fig. 9.2. Illustration of Bragg's law of diffraction, Image is from Michael Hadmack, <http://photonicswiki.org>

$$\vec{k}_f \odot \vec{u} - \vec{k}_i \odot \vec{u} = \vec{Q} \odot \vec{u} \quad (9.5)$$

A schematic diagram described the total phase shifts of X-rays relative to that of the ideal perfect crystal in Bragg CXDI by using equation 9.5. Fig. 9.3 illustrates how lattice displacements fields are calculated and analysed.

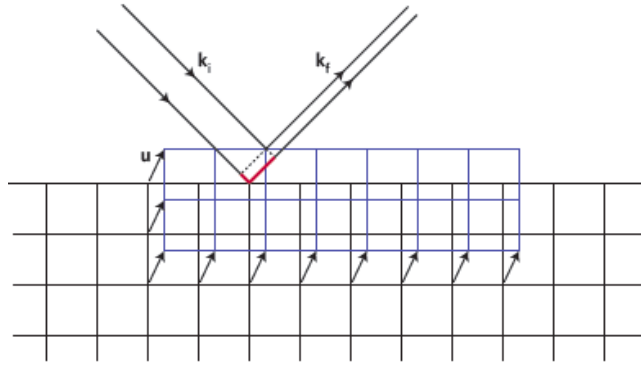


Fig. 9.3. The red lines indicated correspond to the total phase shifts of X-rays relative to that of the crystal specimen, in Bragg reflection geometry, which can be calculated

$$\text{by } \vec{k}_f \odot \vec{u} - \vec{k}_i \odot \vec{u} = \vec{Q} \odot \vec{u} \text{ . Image is from [2].}$$

An exemplary Bragg CXDI reconstruction on single crystal is illustrated in Fig. 9.4.

In Bragg CXDI, to obtain there-dimensional Bragg coherent diffraction intensities, one needs to rock the third dimension so that 3D coherent patterns are being measured. The

3D Bragg CXDI can measure lattice displacements or even quantitatively, the full (three-dimensional) strain and stress tensors of single crystals if CXDI measurements are performed on three different Bragg reflections with none coplanar with each other. An example of reconstruction of full tensor is reported by Newton [5]. The typical 3D Bragg CXDI rocking dimension can be of along the plane perpendicular to the K_f , which is the direction of sample exit wave. Recent studies⁶ show that the rocking direction can also be of along the X-ray energies instead of rocking sample stage; that is, one can rock the X-ray energies around Bragg energy as long as oversampling in the energy rocking direction is satisfied.

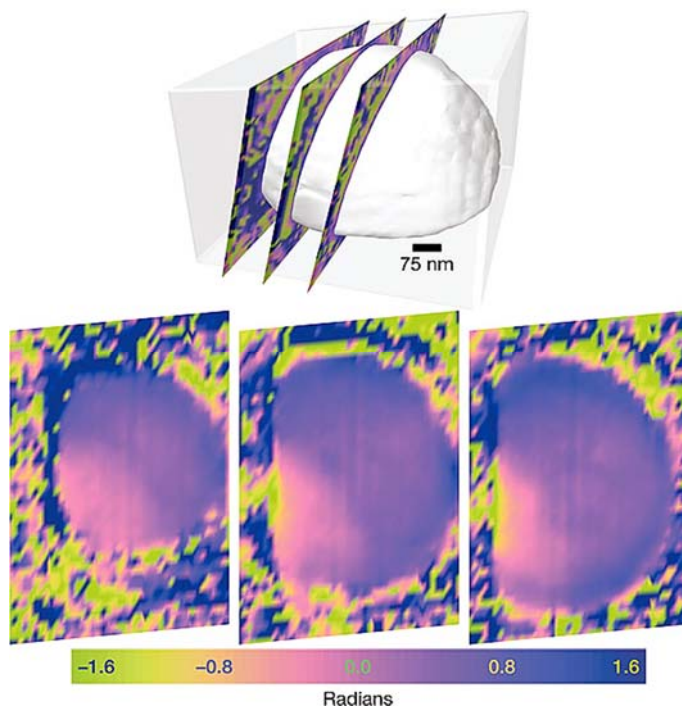


Fig. 9.4. Image is from [1, 2] reconstructed amplitude and phase of typical diffraction intensities of Pb nanoparticles by using iterative algorithms.

Bragg CXDI is a novel technique for investigation of strained single crystals when the displacements fields are relatively less than 2π that is less than a lattice unit cell. When the relative phase shifts exceeds 2π , reconstruction algorithms normally fail to converge, especially when objects under study have high degree of phase curvature in real-space. Shi has illustrated⁷ this problem by simulations on highly strained objects with both linear phase shifts and high levels of phase curvatures. The simulations show that reconstructions work well for linear phase shifted objects, while objects with high phase curvatures generally do not converge with conventional Error-Reduction (ER) [8] and Hybrid-Input-Output (HIO) [9, 10] algorithms. Figs. 9.5-9.7 show how CXDI fails data reconstructions when object phase structures are of high curvatures.

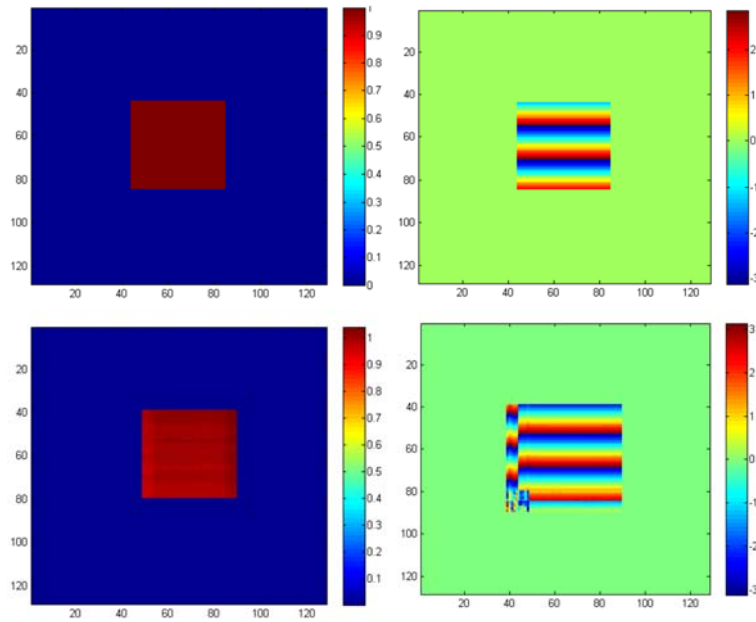


Fig. 9.5. Top Panel: Simulated amplitude and phase; where the support is 2D box with 4 pixels bigger than the original object on each sides of the box. Bottom Panel: Reconstructed amplitude and phase by using ER (500 iterations) + HIO (500 iterations) + ER (500 iterations). Figure is from Xiaowen Shi's PhD Thesis, University College London, UK, 2013.

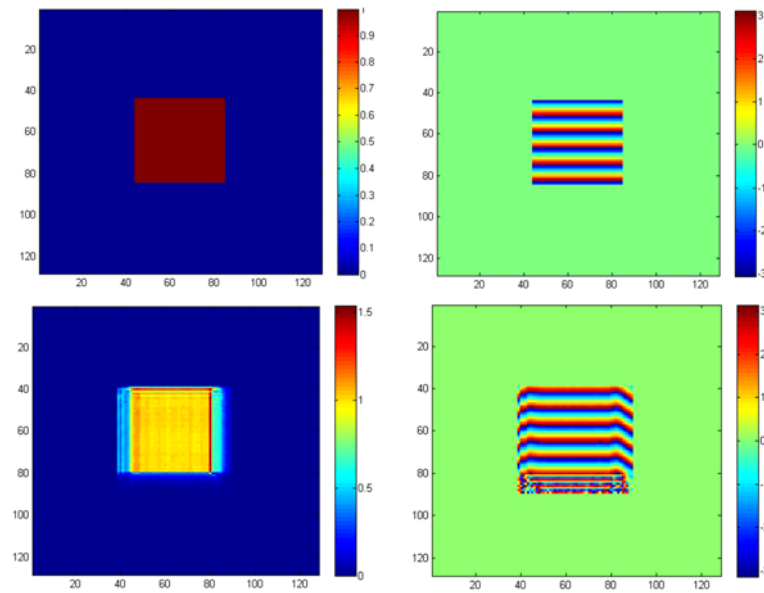


Fig. 9.6. Top Panel: Simulated amplitude and phase; where the support is 2D box with 4 pixels bigger than the original object on each sides of the box. Bottom Panel: Reconstructed amplitude and phase by using ER (500 iterations) + HIO (500 iterations) + ER (500 iterations). Figure is from Xiaowen Shi's PhD Thesis, University College London, UK, 2013.

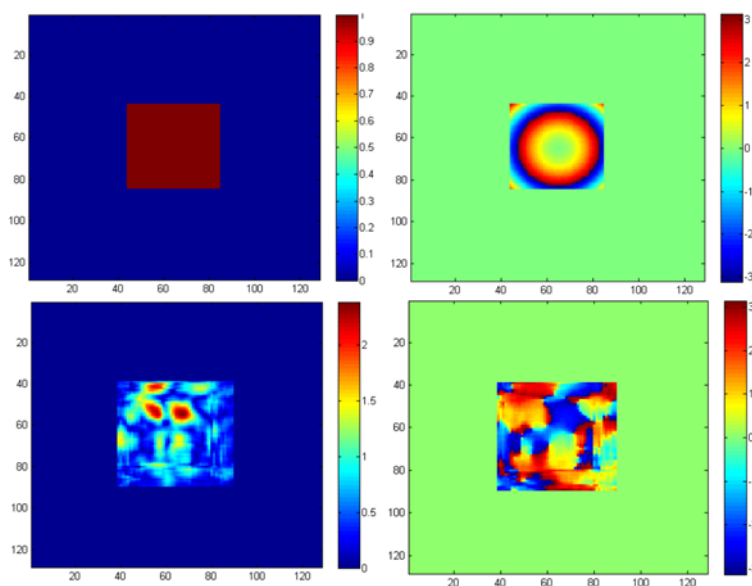


Fig. 9.7. Top Panel: Simulated amplitude and phase of a strong phase structure in 2D; where the support is 2D box with 2 pixels bigger than the original object on each sides of the box. Bottom Panel: Reconstructed amplitude and phase by using ER (500 iterations) + HIO (500 iterations) + ER (500 iterations). Figure is from Xiaowen Shi's PhD Thesis, University College London, UK, 2013.

CXDI in both forward and Bragg geometries are overall techniques with limited sample sizes, i.e. the illumination X-ray probe has to be equal or bigger than the sample structure. In general, researchers do structural characterizations on extended samples. The constraints in CXDI are support constraint and modulus constraints in real-space and Fourier-space respectively [11]. The real-space constraint in CXDI sometimes can be difficult to satisfy due to the extended nature of samples in materials sciences, as a result, a new class of real-space constraints are required for extended sample studies. Rodenburg [12] combined both Scanning Transmission X-ray Microscopy (STXM) and CXDI to have successfully reconstructed extended samples, and the new technique is named Ptychography. The following diagram illustrates how the new real-space constraint called “overlap constraint” is used in Ptychography iterative algorithms.

9.2. Ptychography

9.2.1. Overview and Introduction

Ptychography is an imaging ideology that was originally invented by Walter Hoppe [13] in 1960s, primarily aiming to study especially the phase problem by recording overlapping Bragg reflections in samples of scientific interests, especially in crystals. The word “Ptycho” is thought to be derived from the Greek word “Ptycho” relates to the folding of the diffractions via convolution of the Fourier transform of an illumination function in the

plane of the systems of interests. After the intense developments of Coherent X-ray diffraction imaging (CXDI) [2, 3, 11] has been under intense developments for over a decade, Rodenburg [12] pioneered the advancement of merging both Scanning Transmission X-ray Microscopy (STXM) and CXDI, which is ptychography in modern age, to study arbitrarily large extended samples, ranging from crystals to amorphous materials. The following diagram illustrates how the new real-space constraint that is replaced by the “overlap constraint”, is usually used in Ptychography iterative algorithms for data reconstructions. Fig. 9.8 illustrates how initial ptychographic iterative algorithms operate.

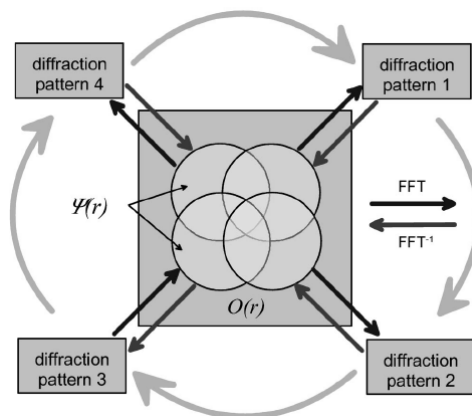


Fig. 9.8. Illustration of the earliest ptychographical iterative engine. The update of object and probe function is in a sequential order. Image is from [12].

Detailed ptychographic iterative algorithms can be found in various literatures [12, 14-16], the conventional algorithms used in ptychography are Extended Ptychographic Iterative Engine (E-PIE) [15], difference-map [14, 16, 17] and RAAR [18], although new advanced ptychographic algorithms [19] are being developed and have been proven to have significant improvements over reconstruction performances.

9.2.2. Forward Ptychography

X-ray ptychography is usually divided into two geometries, forward and Bragg geometries. In forward geometry, sample, X-ray optics and X-ray detectors are co-linear, and depending on the Fresnel numbers, there are two divisions within forward geometry Ptychography, near field [20] and far field. 2D Ptychography can be trivially extended to Ptychography-Tomography [21], where tomographic reconstructions on series of 2D ptychographic projections that are reconstructed using conventional iterative algorithms for tomography reconstructions. Advanced tomographic reconstructions algorithm such as model-based iterative Reconstruction (MBIR) [22] has been demonstrated to have reduced artefacts and noises in reconstructed images using ptychography-Tomography. At present, the record spatial resolution for soft X-ray ptychography is 5 nm [23], where in hard X-ray regime, the resolution record is 16 nm in 3D [24] and roughly 20 nm in 2D.

In materials sciences application, such as in thin film magnetic domain sample, the record resolution is around 10 nm [25] in soft X-rays. One can also utilize Near-edge X-ray refraction fine structure microscopy recently developed in Advanced Light Source [26] to study elemental mapping of nano-battery materials using soft X-ray spectroscopy-ptychography. Materials scientists have been heavily investigating structural, physical and chemical properties of high-performance battery systems by forward geometry ptychography, recent advancements [27] have attracted enormous attentions in advanced battery research community worldwide.

Ptychography with upsampling method [54, 55] has been reported to be successful with hard x-rays, researchers have experimentally demonstrated that oversampling condition can be relaxed because of the nature of over-redundancies in ptychographic datasets. Recent studies show that introduction of diffuser in hard x-ray ptychography has marginally improved ptychographic reconstructions on test samples [56].

9.2.3. Bragg Ptychography

Ptychography in Bragg geometry, has its overwhelming advantage over conventional Bragg CXDI due to the fact that more reliable real-space overlap constraints are incorporated into the iterative algorithms. Not only the Bragg ptychographic data reconstruction convergence is more likely to happen, samples to be used in Bragg ptychography can be of extended structures with arbitrarily large scale in general.

9.2.3.1. Bragg Ptychography on Single Crystals

Various reports on single crystal structures using Bragg Projection Ptychography have been published, which include SiGe-on-SOI structure [28], Ferroelectric Thin Films [29], Nanoscale Semiconductor Heterostructures [30], and Single InGaN/GaN Core-Shell Nanowire [31]. For 3D geometry, various studies have been conducted, notably of which include 2D displacements fields retrieval from Silicon nanowires on a Silicon- insulator wafer [32], structure of an extended ZnO crystal [33], silicon-on-insulator nanostructure [34], an extended InP nanostructured layer bonded onto silicon [35] and nanoscale lattice behaviour and strain fields by using 3D Bragg projection ptychography (3DBPP) [36]. Fig. 9.9 shows reconstructions of a single InGaN/GaN Core-Shell nanowire at various Bragg rocking angles.

9.2.3.2. Bragg Ptychography on Thin Film Phase Domains

Bragg projection ptychography on niobium phase domains has been recently demonstrated [37] with spatial resolution of about 37 nm. In order to satisfy Projection approximation, the thin film thickness has been much smaller than the X-ray probe size on sample [38]. In projection geometry, where the X-ray footprint is relatively large due to small incident angle, this condition can be satisfied when thin film samples are of thickness of several hundreds of nanometer or less [38]. Further studies should be focused on quantitative analyses of thin film domain structures using Bragg projection

ptychography, namely the rotations and vibrations of random phase domains. Domains in other complex electronic systems such as those of orbital-order (OO), magnetic Skyrmion topological state or charge-order (CO) nature should also be looked at in great details by Bragg projection ptychography, because of its unique ability of high resolution phase contrast imaging in Bragg geometry. Fig. 9.10 shows reconstructions of random domain structures with Bragg projection ptychography in Nb thin film.

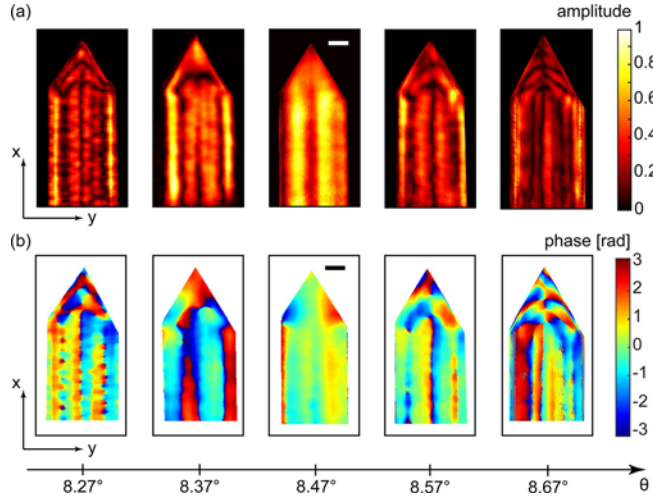


Fig. 9.9. Results of 2D XBP reconstructions. Amplitudes (a), and phases (b) of the NW reconstruction for five angular positions of the rocking curve scan (see inset in Fig. 9.1 (a)). Scale bars are equal to 100 nm. Image is from [31].

9.2.4. Partial Coherence and Multimodes in CXDI Diffraction Imaging

9.2.4.1. Transverse Partial Coherence

Conventional CXDI algorithms assume that the illuminating X-rays are fully coherent, both in longitudinal and transverse directions. The modulus constraint in reciprocal-space simply replaces the calculated square root of the diffraction intensities with the square root of the measured intensities. While this assumption is generally valid, it does not take into account of the fact that in most synchrotron X-ray beamlines, the coherent X-rays are never fully coherent. For transverse partial coherence, experimental CXDI data can be corrected mathematically when partially coherent X-rays are used. For transverse partially coherent illumination, one can consider the complex coherence function in real-space $\mu(\Delta x, \Delta y)$ can be approximated as two-dimensional Gaussian function as follows:

$$\mu(\Delta x, \Delta y) = \mu(r) = \exp \left[- \left((\Delta x)^2 + (\Delta y)^2 \right) / 2\sigma^2 \right] . \quad (9.6)$$

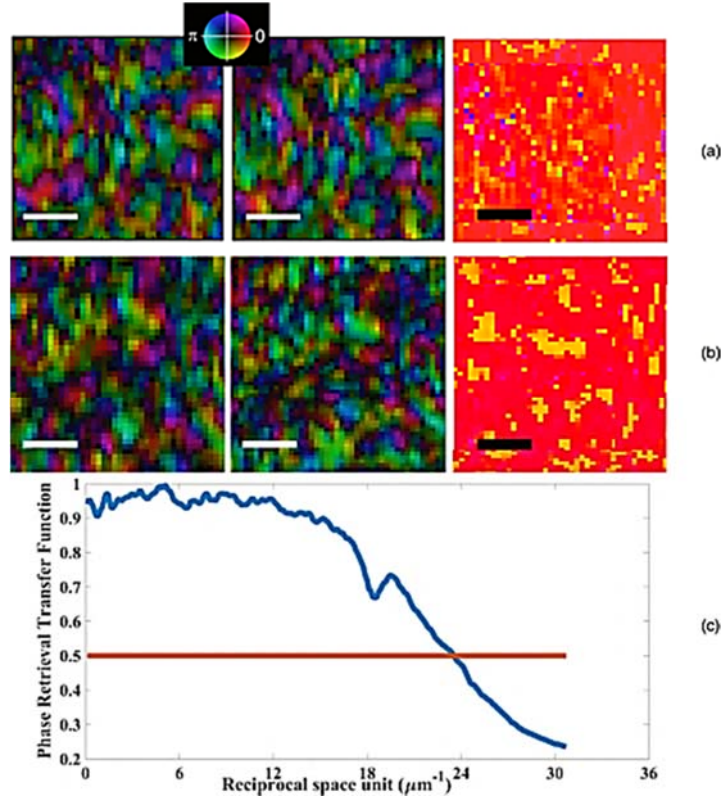


Fig. 9.10. (a) Hue rendering of a central region of objects with suppressed phase ramps, according to Eq. 9.4 and phase difference map of the two reconstructions with random initial starts. The difference map image was performed on the image shown. (b) Same for a different sample area. Both white and black scale bars are 200 nm. (c) Phase retrieval transfer function of the data used in this study. Figure from [37].

The measured partially coherent diffraction intensities can be described by taking into account of the complex coherence function as follows:

$$I_{pc}(q) = I_{fc}(q) \otimes \hat{\gamma}(q), \quad (9.7)$$

where $I_{fc}(q)$ is the fully-coherent diffraction intensities for ideal case when the illuminating X-rays are ideally coherent, $\hat{\gamma}(q)$ is the Fourier transform of the complex coherence function, and \otimes represents convolution.

The resulting reciprocal-space modulus constraint is modified to be as follows:

$$\hat{\psi}'(q) = \hat{\psi}(q) \frac{\sqrt{I_m(q)}}{\sqrt{|\hat{\psi}(q)|^2 \otimes \hat{\gamma}(q)}}. \quad (9.8)$$

While this Richardson-Lucy (RL) algorithm has been widely adopted [39-41] for transverse partial coherence correction, others consider partial coherence of X-rays as electromagnetic illumination of multiple orthogonal modes [42]. By taking into account the multi-modal approach, CXDI reconstructions are significantly improved.

Although partial coherence can usually be corrected in measured data, some argue that there is a fundamental limit [43] beyond which the measured CXDI data would no longer be useful. There has also been claims that simple partial coherence correction might degrade data reconstructions for highly-complex objects [44] due to the fact that an underdetermination of the phase problem in the case of partially coherent illumination.

9.2.4.2. Longitudinal Partial Coherence

Synchrotron X-rays that are generated from insertion devices always exhibit certain degrees of longitudinal partial coherence. In the case that X-rays monochromators are absent, illuminating X-rays are considered to be broadband or pink beams due to the nature of the spectra composed with multiple wavelengths. When measuring CXDI data using broadband or pink X-ray beams, partial coherence due to longitudinal coherence should be incorporated into the iterative algorithms to accommodate the effects of multiple wavelengths X-rays spectra used in study. Abbey proposed CXDI algorithms that include electromagnetic wave spectra as a function of energy to reconstruct measured CXDI data by using broadband X-rays produced by an insertion device [45], others used similar approach to optimise iterative algorithms to take into account multiple wavelength illumination [46]. Clark has elaborated [47] the general descriptions of mutual optical intensity (MOI) for both X-ray beam and sample. The two MOI can be described by the same principles, and they are mutually equivalent mathematically. Their mathematical formulations are as follows:

$$J_B(r_1, r_2) = \sum_{n=1}^N \eta_n B_n(r_1) B_n^*(r_2), \quad (9.9)$$

$$J_s(r_1, r_2) = \sum_{m=1}^M \eta_m S_m(r_1) S_m^*(r_2). \quad (9.10)$$

Batey have also studied [48] multiple wavelengths in CXDI, and their results have also demonstrated the feasibility of utilizing multiple wavelengths of X-rays to efficiently reconstruct CXDI data. Fig. 9.11 shows how probe modes reconstructions look like as a function of sample vibration modes in experiments.

9.2.5. Resolution Limit and Inverse 4th Power Law (Flux and Resolution)

CXDI is a diffraction imaging technique; it has overwhelming advantages for sample characterizations due to the ability to achieve sub-atomic spatial resolution, quantitative phase-contrast imaging, and study single dislocation/defects in single crystal samples. The

inherent limitation in general is the ultimate ability of scattering of samples under study, and the ability to reliably measure all reciprocal-Q range (especially the high reciprocal-Q scattering signal) X-ray diffraction intensities on high-performance X-ray detectors. The immediate limitation lies in the fact that the minimum flux of required incident X-rays is inverse proportional to the 4th power of the best obtainable spatial resolution [49] according to the consensus of CXDI community worldwide. This suggests that we are only able to achieve diffraction-limited spatial resolution by using diffraction-limited 3rd generation synchrotron sources. Other limitations might include trade-off between spatial resolution and field of view of samples. Also, to measure large field of view of samples, more measurement time is required.

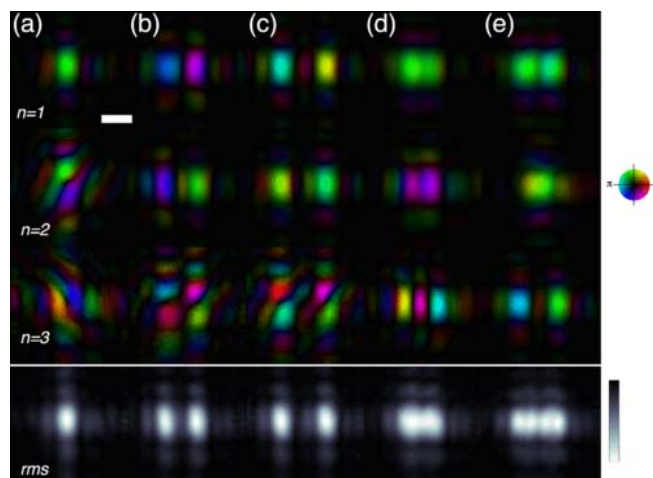


Fig. 9.11. The first three ($n = 1, 2, 3$) recovered probe modes for the cases of (a) a static sample, (b) a square wave with amplitudes $1 \mu\text{m}$ and (c) $1.5 \mu\text{m}$, and (d) a sine wave with amplitudes $1 \mu\text{m}$ and (e) $1.5 \mu\text{m}$. Also shown is the square root of the summed intensity (rms, right column, linear scale) from five probe modes for the cases listed previously. The rms shows clearly how the probe forms can be manipulated using a vibrating sample. The scale bar is $1 \mu\text{m}$. Figure comes from [47].

Although the way forward is to massively increase X-ray doses on samples, soft materials, such as biological specimens are particularly prone to radiation damage, and they are usually relatively weak scatterers. In principle, hard X-rays can be used to study structures of biological samples using phase contrast, however the weak scattering nature of these specimens results to unsatisfactory phase contrast images. Soft X-ray diffraction imaging is an alternative way of studying absorption contrast on biological specimens [Robinson submitted (2017)].

To estimate the minimum amount of X-ray doses required to achieve particular spatial resolution, one can consider theorem by Howells [49] by using concept of Rose criterion [50]. The Rose criterion states that to reliably distinguish of a structure in background noises, the signals generating from the structure has to be at least 5 times larger than the root-mean-square (RMS) of that of the surrounding background when measurement takes place.

$$D = \frac{\mu P h \nu}{\varepsilon \sigma_s} = \frac{\mu P h \nu}{\varepsilon} \frac{1}{r_e^2 \lambda^2 |\tilde{\rho}_r|^2 d^4}, \quad (9.11)$$

where $h\nu$ is the energy of incident X-ray, σ_s is the coherent scattering cross-section of 3D unit volume, called 3D voxel of experimental specimens, ε is the density of object, μ is the intensity absorption coefficient, P is the number of X-rays scattered into the detector from a given voxel (three-dimensional pixel), r_e is the classical radius of an electron, $\tilde{\rho}_r$ is defined to be the complex electron density, and d is the size of single voxel $d \times d \times d$ in each dimension.

The cross-sections of coherent X-ray scattering, which is proportional to the fourth power of pixel resolution is expressed as follows;

$$\sigma_s = r_e^2 \lambda^2 |\tilde{\rho}_r|^2 d^4. \quad (9.12)$$

The requirement for minimum imaging dose is defined as $N_0 = \frac{P}{\sigma_s}$, where N_0 is the total number of incident X-ray per unit area. This estimation leads to the total number N_0 .

9.3. Future Developments Aspects of Single-Shot Coherent Diffraction Imaging

9.3.1. Diffract-and-Destroy Serial Femtosecond Nanocrystallography in 4th Generation Free Electron Laser Facilities

Investigation of biological systems such as crystalized protein with ultra-fast femto-second intense coherent X-ray pulses has not been feasible until recently, a 4th generation X-ray free electron laser at Stanford, California, United States has made researchers' dreams become reality. As we have discussed in Section 9.2.5 in this chapter, high-resolution X-ray imaging has been rather limited by the maximum amount of X-ray radiation doses that a typical biological specimen, such as a protein nanocrystal can tolerate, using existing methods, such as conventional CXDI or the scanning version of the technique, ptychography. Therefore, in order to achieve high spatial resolution imaging on samples of this category, one might consider using coherent imaging techniques that utilize single-shot ideology. A notably example of such techniques that has been under intense investigation is the diffract-and-destroy or alternatively named as serial femtosecond nanocrystallography (SFX). The first success of SFX has been carried out at FLASH soft-X-ray free-electron laser, Deutsches Elektronen-Synchrotron DESY in Hamburg, Germany, lead by Prof. Henry Chapman [51]. Research teams lead by Prof. John Spence from Arizona State University has pioneered this technique by measuring SFX data at the world's first hard X-rays free electron laser facility, the Linear Coherent Light Source (LCLS) located in Stanford, California [52].

9.3.2. Single-Shot Coherent Modulation Imaging (CMI) for Materials Sciences Application

Prof. Fucai Zhang has developed another strong contender for single-shot coherent imaging, that is the Coherent modulation imaging (CMI) technique [53], and his group has reported their first successful measurement recently. CMI is an advanced imaging technique that involves incorporation of a random phase modulator co-linear to the X-ray propagation direction in forward coherent imaging geometry. High contrast and resolution images are reconstructed on extended samples with arbitrarily large field-of-view.

9.4. Conclusion and Future Outlooks

Overall, CXDI and ptychography are of the newest emerging high-resolution X-ray diffraction imaging techniques that require sophisticated iterative algorithms, high-performance computing power and most importantly, diffraction-limited synchrotron and 4th generation X-ray free electron laser facilities. The advancements and explorations of the techniques during the past decade have been highly encouraging and fruitful. The immediate future work should focus on advanced characterizations of next-generation and nanotechnology materials using CXDI and ptychography to facilitate general scientific discoveries using high spatial and temporal resolution X-ray diffraction imaging in broad range of materials, including biomedical, hard condensed matter systems and spintronics nano-devices.

Acknowledgements

Xiaowen Shi acknowledge partial supported by the U.S. Department of Energy, Office of Basic Energy Sciences, Division of Materials Science and Engineering, under Grant No. DE- FG02-11ER46831.

Xiaowen Shi gratefully acknowledges financial support from physical sciences division and Beamline I13 Coherence branch of Diamond Light Source, Didcot, Oxfordshire, United Kingdom.

References

- [1]. M. F. Pfeifer, G. J. Williams, I. A. Vartanyants, R. Harder, I. K. Robinson, Three-dimensional mapping of a deformation field inside a nanocrystal, *Nature*, Vol. 442, 2006, pp. 63-66.
- [2]. I. Robinson, R. Harder, Coherent X-ray diffraction imaging of strain at the nanoscale, *Nat. Mater.*, Vol. 8, 2009, pp. 291-298.
- [3]. P. C. Jianwei Miao, K. Janos, D. Sayre, Extending the methodology of X-ray crystallography to allow imaging of micrometre-sized non-crystalline specimens, *Nature*, Vol. 400, 1999, pp. 342-344.
- [4]. D. Sayre, Some implications of a theorem due to Shannon, *Acta Cryst.*, Vol. 5, 1952, p. 843.
- [5]. M. C. Newton, S. J. Leake, R. Harder, I. K. Robinson, Three-dimensional imaging of strain in a single ZnO nanorod, *Nature Materials*, Vol. 9, 2009, pp. 120-124.

- [6]. W. Cha, et al., Three dimensional variable-wavelength X-Ray Bragg coherent diffraction imaging, *Physical Review Letters*, Vol. 117, 2016, 225501.
- [7]. X. Shi, Coherent X-ray diffraction imaging and ptychography on silicon-on-insulator nanostructures, PhD Thesis, Department of Physics and Astronomy, *University College*, London, 2013.
- [8]. R. W. Gerchberg, W. O. Saxton, A practical algorithm for the determination of phase from image and diffraction plane pictures, *Optik*, Vol. 35, 1972, pp. 237-246.
- [9]. J. R. Fienup, Phase retrieval algorithms-a comparison, *Applied Optics*, Vol. 21, 1982, pp. 2758-2769.
- [10]. J. R. Fienup, C. C. Wackerman, Phase-retrieval stagnation problems and solutions, *Optical Society of America*, Vol. 3, 1986, pp. 1897-1907.
- [11]. G. Xiong, O. Moutanabbir, M. Reiche, R. Harder, I. Robinson, Coherent X-Ray diffraction imaging and characterization of strain in silicon-on-insulator nanostructures, *Advanced Materials*, Vol. 26, 2014, pp. 7747-7763.
- [12]. J. Rodenburg, et al., Hard-X-Ray lensless imaging of extended objects, *Physical Review Letters*, Vol. 98, 2007, 034801.
- [13]. W. Hoppe, Beugung im inhomogenen Primärstrahlwellenfeld. I. Prinzip einer Phasenmessung von Elektronenbeugungsinterferenzen, *Acta Crystallographica Section A*, Vol. 25, Issue 4, 1969, p. 495.
- [14]. P. Thibault, et al., High-resolution scanning X-ray diffraction microscopy, *Science*, Vol. 321, 2008, pp. 379-382.
- [15]. A. M. Maiden, J. M. Rodenburg, An improved ptychographical phase retrieval algorithm for diffractive imaging, *Ultramicroscopy*, Vol. 109, 2009, pp. 1256-1262.
- [16]. P. Thibault, M. Dierolf, O. Bunk, A. Menzel, F. Pfeiffer, Probe retrieval in ptychographic coherent diffractive imaging, *Ultramicroscopy*, Vol. 109, 2009, pp. 338-343.
- [17]. V. Elser, Phase retrieval by iterated projections, *J. Opt. Soc. Am. A*, Vol. 20, 2003, pp. 40-55.
- [18]. D. R. Luke, Relaxed averaged alternating reflections for diffraction imaging, *Inverse Problems*, Vol. 21, 2005, pp. 37-50.
- [19]. S. Marchesini, Y.-C. Tu, H.-T. Wu, Alternating projection, ptychographic imaging and phase synchronization, *Applied and Computational Harmonic Analysis*, Vol. 41, 2016, pp. 815-851.
- [20]. M. Stockmar, et al., Near-field ptychography: phase retrieval for inline holography using a structured illumination, *Scientific Reports*, Vol. 3, 2013, p. 1927.
- [21]. M. Dierolf, et al., Ptychographic X-ray computed tomography at the nanoscale, *Nature*, Vol. 467, 2010, pp. 436-439.
- [22]. S. V. Venkatakrishnan, et al., Robust X-ray phase ptycho-tomography, *IEEE Signal Processing Letters*, Vol. 23, 2016, pp. 944-948.
- [23]. D. A. Shapiro, et al., Chemical composition mapping with nanometre resolution by soft X-ray microscopy, *Nat. Photon.*, Vol. 8, 2014, pp. 765-769.
- [24]. M. Holler, et al., X-ray ptychographic computed tomography at 16 nm isotropic 3D resolution, *Scientific Reports*, Vol. 4, 2014, 3857.
- [25]. X. Shi, et al., Soft X-ray ptychography studies of nanoscale magnetic and structural correlations in thin SmCo₅ films, *Applied Physics Letters*, Vol. 108, 2016, 094103.
- [26]. M. Farmand, et al., Near-edge X-ray refraction fine structure microscopy, *Applied Physics Letters*, Vol. 110, 2017, 063101.
- [27]. A. M. Wise, et al., Nanoscale chemical imaging of an individual catalyst particle with soft X-ray ptychography, *ACS Catalysis*, Vol. 6, 2016, pp. 2178-2181.
- [28]. S. O. Hruszkewycz, M. V. Holt, C. E. Murray, J. Bruley, J. Holt, A. Tripathi, O. G. Shpyrko, I. McNulty, M. J. Highland, P. H. Fuoss, Quantitative nanoscale imaging of lattice distortions in epitaxial semiconductor heterostructures using nanofocused X-ray Bragg projection ptychography, *Nano Letters*, Vol. 12, 2012, pp. 5148-5154.

- [29]. S. O. Hruszkewycz, et al., Imaging local polarization in ferroelectric thin films by coherent X-Ray Bragg projection ptychography, *Physical Review Letters*, Vol. 110, 2013, 177601.
- [30]. M. V. Holt, et al., Strain imaging of nanoscale semiconductor heterostructures with X-Ray Bragg projection ptychography, *Physical Review Letters*, Vol. 112, 2014, 165502.
- [31]. D. Dzhigaev, et al., X-ray Bragg ptychography on a single InGaN/GaN core-shell nanowire, *ACS Nano*, Vol. 11, 2017, pp. 6605-6611.
- [32]. P. Godard, et al., Three-dimensional high-resolution quantitative microscopy of extended crystals, *Nature Communications*, Vol. 2, 2011, 568.
- [33]. X. Huang, R. Harder, S. Leake, J. Clark, I. Robinson, Three-dimensional Bragg coherent diffraction imaging of an extended ZnO crystal, *J. Appl. Cryst.*, Vol. 45, 2012, pp. 778-784.
- [34]. V. Chamard, M. Allain, P. Godard, A. Talneau, G. Patriarche, M. Burghammer, Strain in a silicon-on-insulator nanostructure revealed by 3D X-ray Bragg ptychography, *Scientific Reports*, Vol. 5, 2015, 9827.
- [35]. A. I. Pateras, et al., Nondestructive three-dimensional imaging of crystal strain and rotations in an extended bonded semiconductor heterostructure, *Physical Review B*, Vol. 92, 2015, 205305.
- [36]. S. O. Hruszkewycz, et al., High-resolution three-dimensional structural microscopy by single-angle Bragg ptychography, *Nat. Mater.*, Vol. 16, 2017, pp. 244-251.
- [37]. N. Burdet, et al., Bragg projection ptychography on niobium phase domains, *Physical Review B*, Vol. 96, 2017, 014109.
- [38]. S. O. Hruszkewycz, et al., Structural sensitivity of X-ray Bragg projection ptychography to domain patterns in epitaxial thin films, *Physical Review A*, Vol. 94, 2016, 043803.
- [39]. J. N. Clark, X. Huang, R. Harder, I. K. Robinson, High-resolution three-dimensional partially coherent diffraction imaging, *Nature Communications*, Vol. 3, 2012, 993.
- [40]. N. Burdet, et al., Evaluation of partial coherence correction in X-ray ptychography, *Optics Express*, Vol. 23, 2015, pp. 5452-5467.
- [41]. L. W. Whitehead, et al., Diffractive imaging using partially coherent X rays, *Physical Review Letters*, Vol. 103, 2009, 243902.
- [42]. P. Thibault, A. Menzel, Reconstructing state mixtures from diffraction measurements, *Nature*, Vol. 494, 2013, pp. 68-71.
- [43]. B. Chen, et al., Diffraction imaging: The limits of partial coherence, *Physical Review B*, Vol. 86, 2012, 235401.
- [44]. D. H. Parks, X. Shi, S. D. Kevan, Partially coherent X-ray diffractive imaging of complex objects, *Physical Review A*, Vol. 89, 2014, 063824.
- [45]. B. Abbey, L. W. Whitehead, H. M. Quiney, D. J. Vine, G. A. Cadenazzi, C. A. Henderson, K. A. Nugent, E. Balaur, C. T. Putkunz, A. G. Peele, G. J. Williams, I. McNulty, Lensless imaging using broadband X-ray sources, *Nature Photonics*, Vol. 5, 2011, pp. 420-424.
- [46]. B. Chen, et al., Multiple wavelength diffractive imaging, *Physical Review A*, Vol. 79, 2009, 023809.
- [47]. J. N. Clark, X. Huang, R. J. Harder, I. K. Robinson, Dynamic imaging using ptychography, *Physical Review Letters*, Vol. 112, 2014, 113901.
- [48]. D. J. Batey, D. Claus, J. M. Rodenburg, Information multiplexing in ptychography, *Ultramicroscopy*, Vol. 138, 2014, pp. 13-21.
- [49]. M. R. Howells, et al., An assessment of the resolution limitation due to radiation-damage in X-ray diffraction microscopy, *J. of Electron Spectroscopy and Related Phenomena*, Vol. 170, 2009, pp. 4-12.
- [50]. A. Rose, Television pickup tubes and the problem of vision, in *Advances in Electronics*. Vol. 1 (L. Marton, Ed.), *Academic Press*, New York, 1948, pp. 131-166.
- [51]. H. N. Chapman, et al., Femtosecond diffractive imaging with a soft-X-ray free-electron laser. *Nature Physics*, Vol. 2, 2006, pp. 839-843.

- [52]. J. Spence, X-ray imaging: Ultrafast diffract-and-destroy movies, *Nat. Photon.*, Vol. 2, 2008, pp. 390-391.
- [53]. F. Zhang, et al., Phase retrieval by coherent modulation imaging, *Nature Communications*, Vol. 7, 2016, 13367.
- [54]. Edo, T. B. et al., Sampling in x-ray ptychography, *Physical Review A*, 87, 2013, 053850.
- [55]. Batey, D. J. et al., Reciprocal-space up-sampling from real-space oversampling in x-ray ptychography, *Physical Review A*, 89, 2014, 043812.
- [56]. Peng, L. et al., Multiple mode x-ray ptychography using a lens and a fixed diffuser optic, *Journal of Optics*, 18, 2016, 054008.

Chapter 10

The Optical Anderson Localization in Three-Dimensional Percolation System

**Gennadiy Burlak, Alfredo Díaz-de-Anda,
Erika Martínez-Sanchez and Álvaro Zamudio-Lara**

10.1. Introduction

Disordered photonic materials can diffuse and localize light through random multiple scattering, offering opportunities to study mesoscopic phenomena, control light-matter interactions, and provide new strategies for photonic applications [1]. Wave transport in disordered systems is a fast developed topic, one of that is the optical wave in random dielectrics [2]. In disordered optical materials, the multiple scattering of light and the interferences between propagating waves lead to the formation of electromagnetic modes with varying spatial extent, depending on scattering strength, structural correlations, and dimensionality [3] of the system [4, 5]. This combination leads to a series of interesting physical effects and also creates large potential for new disorder-based optical applications [6]. It is well-known that at the optical Anderson localization (OAL) the counter-propagating waves form closed loops; interference along these loops lead to randomly shaped standing-wave patterns confining the light [6, 7]. In one and two dimensions, the chance of coming back to the same region a necessary requirement to form a loop is much higher than in three dimensions. For this reason, observing OAL in three dimensions is extremely difficult for light waves [6]. Three-dimensional disordered structures have been studied recently for investigation of complex optical phenomena, including light localization [6-11], and random lasing [12-14]. In most lasing random materials, the intensity is spread throughout the sample and, in general, there are several lasing modes. In certain cases interference of different modes can lead to light localization [15, 16].

The study for Anderson transition in 3-D optical systems still has not been conclusive despite considerable efforts. The localization transition may be difficult to reach for the light waves due to various effects in dense disordered media required to achieve strong

scattering (see e.g. [17] and references therein). The experimental observation of OAL [18] just below the Anderson transition in an open 3D disordered medium shows strong fluctuations of the wave function that leads to nontrivial length-scale dependence of the intensity distribution (multifractality). Such behavior can be specified by varying the system size and quantified deeper by using the generalized inverse participation ratios (IPR).

Various schemes can be proposed to study OAL, e.g. the use a coherently prepared three-level atomic medium provides a disordered scheme for realizing the Anderson localization (see e.g. [19] and references therein).

Other perspective 3D disordered systems where the optical transport was studied belong to percolating crystals. In such materials the optical transparency assisted by disordered porous clusters was observed [20]. Also it was studied some interesting properties of optical nanoemitters incorporated into 3D spanning cluster [21]. In supercritical state, the field intensity is large enough to produce a dynamic high-density coherent field. For material with small losses the long-term coherence arises in the area close to the percolation threshold. It is found also the random lasing assisted by nanoemitters incorporated into such a disordered structure [22]. In this system the spanning cluster produces a global percolation that results in qualitatively modification of its spatial properties. One can argue that already in a vicinity of the percolating phase transition the fractal dimension of such system $D_H = 2.54$ considerably differs from the dimension of the embedded space $D = 3$ (multifractality). One of important question is whether the optical Anderson localization [23] can still be archived for a non-integer dimension case with a fractal (Hausdorff) dimension of $D_H < 3$, where the strong randomness for properties of the system is expected.

In this chapter we study the optical localization associated with the structure of three dimensional (3D) disordered percolation system, where the percolating clusters are filled by active media composed by light noncoherent emitters. In such a non-uniformly spatial structure the radiating and scattering of field occur by incoherent way. We have studied the range of parameters where the wave localization can take place. The Fermat principle and Monte Carlo approach are applied to characterize the optimal optical path and interconnection between the radiating emitters. This allows considering the average free path of light as a mean that in a simplest case is a ratio of the length of path to number of emitters. This leads to formulating of condition when the localization can occur (similarly to Ioffe-Regel criterion) for a percolating model. Our 3D simulations allow calculating of such free path to study the situation where such condition can be satisfied. FDTD numerical simulations have shown that in such a system the localized optical modes arise closely to the percolation threshold. We also studied the properties of associated field inverse participating ratio (IPR) and found that mean free path, IPR and the percolation order parameter show well-pronounced critical behavior near the percolating threshold. Considered in this chapter random percolating system is specially challenging due to its spatial inhomogeneity and also multifractality in 3D.

The chapter is organized as follows. In Section 10.2 we formulate the main equations. In Section 10.3 we present the numerical results on the field distribution generated by the emitters in percolating medium. Section 10.4 contains the approach and formulation the conditions of optical localization in percolating system. In Section 10.5 we study the properties of the inverse participation ratio (IPR) for percolating systems. Section 10.6 contains the results of our numerical simulations of the Anderson localization, and the last sections contain the discussion and conclusions.

10.2. Basic Equations

We study the integral emission of electromagnetic energy from a cubical sample $(x, y, z) \in [0, l_0]$. The output flux of energy can be written as

$$I = \oint_s (K \cdot n) dS = I_x + I_y + I_z, \quad (10.1)$$

where K is the Poynting vector, n is the normal unit vector to the surface S of cube, and $I_{x,y,z}$ indicate the fluxes from two faces of the cube perpendicular to a particular direction. To find the emission from the system we solve numerically the equation that couples the polarization density P , the electric field E , and occupations of the levels of emitters. In the case of uncoupled emitters this equation is [24]

$$\frac{\partial^2 P}{\partial t^2} + \Delta \omega_a \frac{\partial P}{\partial t} + \omega_a^2 P = \frac{6\pi\epsilon_0 c^3}{\tau_{21}\omega_a^2} (N_1 - N_2) E. \quad (10.2)$$

Here $\Delta \omega_a = \tau_{21}^{-1} + 2T_2^{-1}$, where T_2 is the mean time between dephasing events, τ_{21} is the decay time from the second atomic level to the first one, and ω_a is the frequency of radiation. The electric and magnetic fields, E and H , and the current $j = \partial P / \partial t$ are found from the Maxwell equations, together with the equations for the densities $N_i(r, t)$ of atoms residing in i -th level. In the case of four level laser $i = 0, 1, 2, 3$ these rate equations read (see [25] and references therein)

$$\frac{\partial N_3}{\partial t} = A_r N_0 - \frac{N_3}{\tau_{32}}, \quad \frac{\partial N_2}{\partial t} = \frac{N_3(t)}{\tau_{32}} + \frac{(j \cdot E)}{\hbar \omega_a} - \frac{N_2}{\tau_{21}}, \quad (10.3a)$$

$$\frac{\partial N_1}{\partial t} = \frac{N_2(t)}{\tau_{21}} - \frac{(j \cdot E)}{\hbar \omega_a} - \frac{N_1}{\tau_{10}}, \quad \frac{\partial N_0}{\partial t} = \frac{N_1}{\tau_{10}} - A_r N_0. \quad (10.3b)$$

An external source excites emitters from the ground level ($i = 0$) to third level at a certain rate A_r , which is proportional to the pumping intensity in experiments. After a short lifetime τ_{32} , the emitters transfer nonradiatively to the second level. The second level and

the first level are the upper and the lower lasing levels, respectively. Emitters can decay from the upper to the lower level by both spontaneous and stimulated emission, and $(j \cdot E) / \hbar \omega_a$ is the stimulated radiation rate. Finally, emitters can decay nonradiatively from the first level back to the ground level. The lifetimes and energies of upper and lower lasing levels are τ_{21} , E_2 and τ_{10} , E_1 , respectively. The individual frequency of radiation of each emitter is then $\omega_a = (E_2 - E_1) / \hbar$.

Below we consider the situation when incipient percolating cluster is completely filled with light sources. Such cluster (simple cubic lattice) is shown in Fig. 10.1, where all internal uncoupled clusters have been omitted. We indicate that the percolation cluster in Fig. 10.1 has a typical dendrite shape that, however, depends on the actual random sampling. Re-running simulation with another random seed value will lead to percolation cluster with somewhat different geometry, which will also have similar sponge structure. The cluster is grown in the x -direction indicated by the arrow in Fig. 10.1, which provides substantially asymmetric structure. The equations (10.2)-(10.3) are solved with initial conditions that correspond to inversion of occupations of the emitters levels and some random seed for the electromagnetic field that simulates the noise in the system [24, 26].

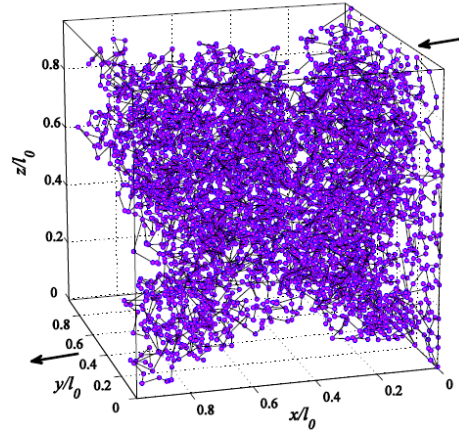


Fig. 10.1. (Color online) Typical spatial structure of the incipient percolating cluster near the percolation threshold at $p = 0.32$ in the cube $l_0 \times l_0 \times l_0$, where $l_0 = 10^{-4} \text{ m}$. The cluster is shown for $50 \times 50 \times 50$ numerical grid. Only clusters coupled to the spanning cluster are shown, while all the internal clusters unconnected to the entry side are not displayed. In this configuration considerable quantity of the emitters are incorporated closely to the entry side (indicated by incoming arrow) of the crystal. The solid line connects the nodes jointed with the use of the variational Fermat's principle.

The structure of field in disordered three-dimensional (3D) active media relies strongly on the huge amount of computational resources needed when dealing with 3D problems.

The simpler, low-dimensional models (1D and 2D) are quite unsatisfactory for many important reasons, including the critical nature of the random lasing in 3D [27]. Since the classical diffusion model does not correctly describe the propagation of photons in a scattering medium with non-uniformly distributed gain or loss [28] apply FDTD technique [29] for our numerical simulations.

In calculations we considered the gain medium with parameters close GaN powder, similar to Ref. [30]. The lasing frequency ω_a is $2\pi \times 3 \times 10^{13}$ Hz, the lifetimes are $\tau_{32} = 0.3$ ps, $\tau_{10} = 1.6$ ps, $\tau_{21} = 16.6$ ps, and the dephasing time is $T_2 = 0.0218$ ps. The percolating cluster has been generated inside the cube of $l_0 = 1 \mu\text{m}$ edge having L^3 nodes, with $L = 75$ and $L = 100$ that was sufficient to simulate the percolating structure [20, 21]. Each node is supposed to indicate the position of many emitters, with the total concentration of emitters inside the percolation cluster being $N = N_0 + N_1 + N_2 + N_3 = 3.3 \times 10^{24} \text{ m}^{-3}$. The initial (at $t = 0$) values of densities $N_0(0) = 0.001 N$, $N_1(0) = 0.002 N$, $N_2(0) = 0.002 N$, and $N_3(0) = 0.995 N$ are used. The dielectric permittivity of a host material is $n = 2.2$ that is close to the typical values for ceramics $\text{Lu}_3\text{Al}_5\text{O}_{12}$, SrTiO_3 , ZrO_2 , see review [31]. The results of simulations are obtained for cw pumping given by $A_r = 10^7 \text{ s}^{-1}$. At this pumping all the simulations show the formation a well-defined lasing for $t > t_s$, and we refer t_s as the lasing start time in what follows. To simulate the noise in our system the initial seed for the electromagnetic field has been created with random phases at each node.

10.3. Lasing in Percolating System with Incorporated Emitters

To simulate the laser medium we consider the situation when the incipient percolating cluster is completely filled with light sources. Such percolating cluster is shown in Fig. 10.1, where all internal uncoupled sites have been omitted. We indicate that the percolation cluster in Fig. 10.1 has a typical dendrite shape that, however, depends on the actual random sampling. Re-running simulation with another random seed value will lead to percolation cluster with somewhat different geometry, which will also have similar sponge structure. The cluster is grown in the x -direction indicated by the arrow in Fig. 10.1, which provides substantially asymmetric structure. The equations (1-3) are solved with initial conditions that correspond to inversion of occupations of the emitters levels and some random seed for the electromagnetic field that simulates the noise in the system [24, 26].

The theory of lasing from disordered three-dimensional (3D) active media relies strongly on the huge amount of computational resources needed when dealing with 3D problems. The simpler, low-dimensional models (1D and 2D) are quite unsatisfactory for many important reasons, including the critical nature of the random lasing in 3D [27]. Since the classical diffusion model does not correctly describe the propagation of photons in a scattering medium with non-uniformly distributed gain or loss [28], we apply FDTD technique for our numerical simulations. Briefly, the strategy of our simulations consists

in the following: (i) Calculating the geometry of the spanning percolating cluster, (ii) Calculating the photon field E generated by emitters incorporated in spanning cluster with the use of FDTD technique [21, 29], and (iii) Solution of nonlinear dynamic coupled equations for field, polarization P and the occupation numbers N_i for all the emitters in 3D system. The Runge-Kutta algorithm with automatic control of the time stepping has been applied for the latter.

In calculations we considered the gain medium with parameters close to GaN powder, similar to Ref. [30]. The lasing frequency ω_a is $2\pi \times 3 \times 10^{13}$ Hz, the lifetimes are $\tau_{32} = 0.3$ ps, $\tau_{10} = 1.6$ ps, $\tau_{21} = 16.6$ ps, and the dephasing time is $T_2 = 0.0218$ ps. The percolating cluster has been generated inside the cube of $l_0 = 1 \mu\text{m}$ edge having L^3 nodes, with $L = 75$ and $L = 100$ that was sufficient to simulate the percolating structure [20, 21]. Each node is supposed to indicate the position of many emitters, with the total concentration of emitters inside the percolation cluster being $N_t = N_0 + N_1 + N_2 + N_3 = 3.3 \times 10^{24} \text{ m}^{-3}$. The initial (at $t = 0$) values of densities $N_0(0) = 0.001 N$, $N_1(0) = 0.002 N$, $N_2(0) = 0.002 N$, and $N_3(0) = 0.995 N$ are used. The dielectric permittivity of a host material is $n = 2.2$ that is close to the typical values for ceramics $\text{Lu}_3\text{Al}_5\text{O}_{12}$, SrTiO_3 , ZrO_2 , see review [31]. The results of simulations shown in figures 2-6 are obtained for cw pumping given by $A_r = 10^7 \text{ s}^{-1}$. At this pumping all the simulations show the formation a well-defined lasing for $t > t_s$, and we refer t_s as the lasing start time in what follows. To simulate the noise in our system the initial seed for the electromagnetic field has been created with random phases at each node.

Fig. 10.2 displays the dynamics of field and polarization as functions of time for emitters incorporated in a percolating cluster with occupation probability $p = 0.32$ in simple cubic lattice, which is close to the critical percolation value $p_c \approx 0.3116$ (see [32]). Panels (a) and (b) show the flux of energy I_x in the growth direction (see Eq. (10.1)), while panels (c), (d) display the polarizations $P_x(t)$ corresponding to the deepest (with respect of the entry surface) emitter in the percolating cluster. Comparing the panels (a) and (b) in Fig. 10.2 we observe that the lasing start time t_s increases with the system size (L^3), or more precisely with the number of emitters incorporated in the cluster. This is due to the fact that increase of the system size increases the volume of the phase synchronization (number of emitters) in a percolating medium. Fig. 10.2 shows that the lasing begins with the spiking oscillations having strongly modulated shape, but after some time the amplitude of oscillations decreases. Note that the phenomenon of 'laser spiking' shown in Fig. 10.2 has been observed at fast turn-on of some lasers [33]. Normally such oscillations occur when the gain changes rapidly due to the fluctuations of pumping. In our case, however, such oscillations appear due to considerably inhomogeneous distribution of emitters in the percolating medium. The other important feature of lasing from the

percolating media is that the energy flux at output side $I_x(L)$ it is about twice more than the flux at input side $I_x(0)$ indicated by incoming arrow in Fig. 10.1. The total emission along the growth direction is also more than twice stronger than the emission from the lateral sides. The presence of considerable directionality emission is presumably due to the substantial anisotropy of the incipient percolation cluster seen in Fig. 10.1.

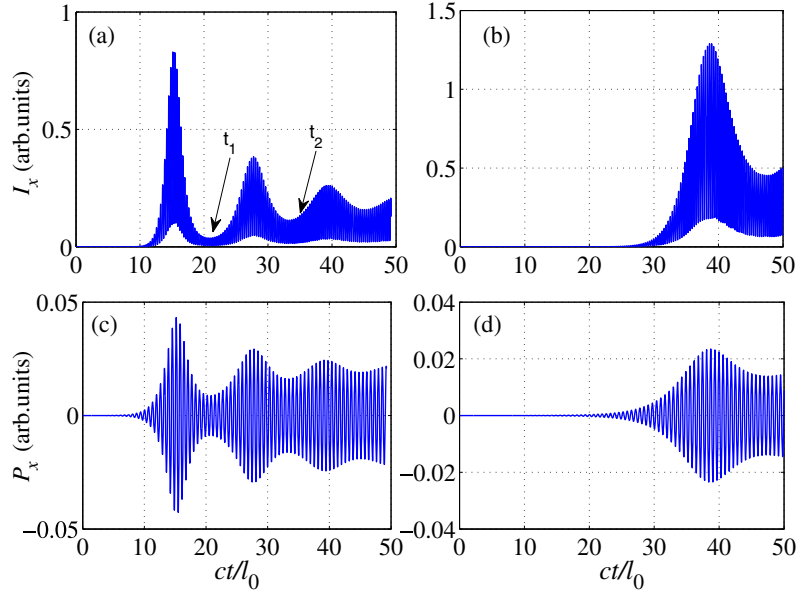


Fig. 10.2. Showing the formation of laser generation. Panels (a) and (b) demonstrate the flux I_x in the x -direction of system. Panels (c) and (d) show the polarization $P_x(t)$ of the deepest emitting node in the cluster. In panels (a, c) and (b, d) the numerical grid with $L=75$ and $L=100$ has been used. The lasing start times are $t_s = \frac{12l_0}{c} = 4$ ps and $t_s = \frac{28l_0}{c} = 9.3$ ps for $L=75$ and 100, respectively.

To study the details of this lasing we investigated the spectra of the field output. The Fig. 10.3 displays the spectral structure of the output shown in Fig. 10.2 in panels (a) and (b) for $L=75$ and $L=100$, respectively. From Fig. 10.3 we observe that the spectrum has rather indented shape beyond the central line. Note that this line is placed closely to the single emitter frequency ω_q but does not coincide with it exactly, and it shows substantial narrowing of the emission spectrum. The shift in the emission frequency appears because the surrounding media, in which the percolating cluster resides, serves as an effective cavity helping to keep the radiation inside the cluster and to form the lasing effect. As a result, at larger times after the excitation start, $t \gg t_s$, great number of emitters become synchronized to contribute into the stimulating radiation. Contrary to the distributed random lasers, we observe that only one mode dominates at large times and this mode extends through the whole light emitted media, while all other modes are suppressed.

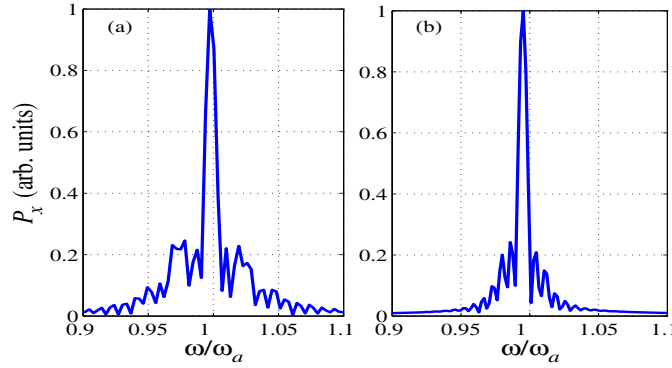


Fig. 10.3. The frequency spectrum of the field shown in Fig. 10.2. Panels (a) and (b) show the amplitudes of emission spectra for cases $L=75$ and $L=100$, respectively. An overall narrowing of the frequency spectrum is observed.

Fig. 10.4 (a, b) exhibits the dynamics of population numbers N_i (for the deepest emitter), where the levels 1 and 2 correspond to the laser levels. The emission intensity averaged over the total observation time, $\bar{I}_i(t) = t^{-1} \int_0^t I_i(\tau) d\tau$, is shown in Fig. 10.4, panels (c, d), again for the cases $L=75$ and $L=100$, respectively. The fluxes of the emission energy through the growth direction, \bar{I}_x , and the lateral sides, $\bar{I}_{y,z}$, demonstrate considerably anisotropic emission at least in considered case of the incipient percolating cluster.

Fig. 10.5 shows the details of finite-difference time-domain (FDTD) calculation of distribution of the field intensity in the central intersection of 3D sample in a spanning percolating cluster for systems of different sizes. One can observe fluctuating, nonuniform, and likely speckle patterns. They are characterized by presence of “hot spots” of typical sub-micron sizes. We note that such patterns are generic near-field imaging feature for various 2D systems near percolation [34, 35].

Fig. 10.5(a-c) displays the complete FDTD simulations of 3D spatial structure of the field amplitude E_x generated by emitters incorporated in percolating cluster at different times after the pumping start. Initially, before the lasing start time $t_s = 4 \text{ ps}$, the emission appears from small-scaled uncoupled domains which give rise to weak incoherent radiation with random phases, see Fig. 10.5 (a). However, latter the radiation from nearest patterns is being synchronized that leads to formation of macroscopically large areas (patterns) of field.

In every cluster the partial lasing state will be established and this leads to overall increase of coherent emission. For longer times, the other behavior of field is observed. From Fig. 10.5(b) one can see that the lasing modes are confined to 3D areas around the localization centers. Each mode has its own specific frequency and corresponds to a peak in the radiated spectrum inside the system (see Fig. 10.3). From Fig. 10.5 (b) we observe the random field structure where the coupled coherent field patterns are arisen.

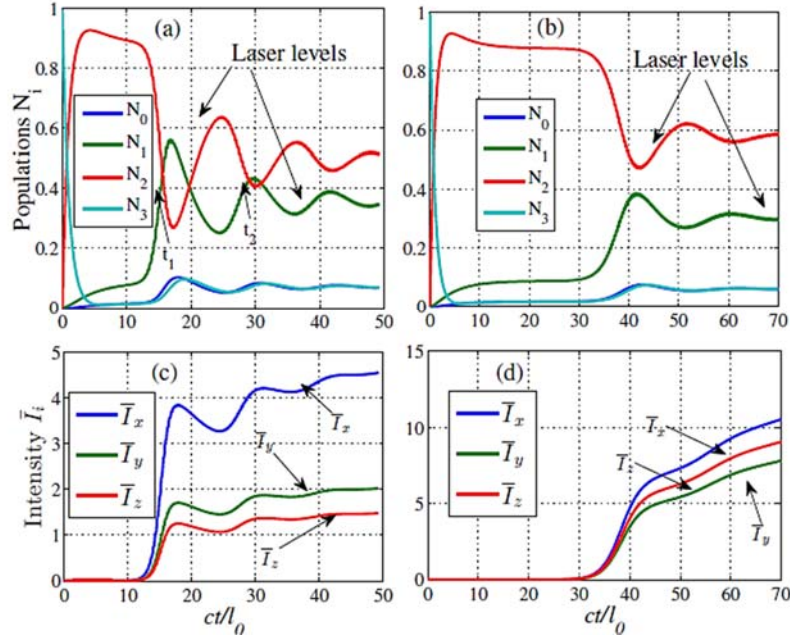


Fig. 10.4. (a, b) Dynamics of population numbers N_i ($i = 0, 1, 2, 3$) where $N_{1,2}$ correspond to the lasing levels. The times $t_{1,2}$ of minima of the field output in Fig. 10.2(a) correspond to the intersections of lasing levels. (c, d) Integrated over the observation time t intensities $\bar{I}_{x,y,z}$ of radiation emitted along corresponding directions. See details in text.

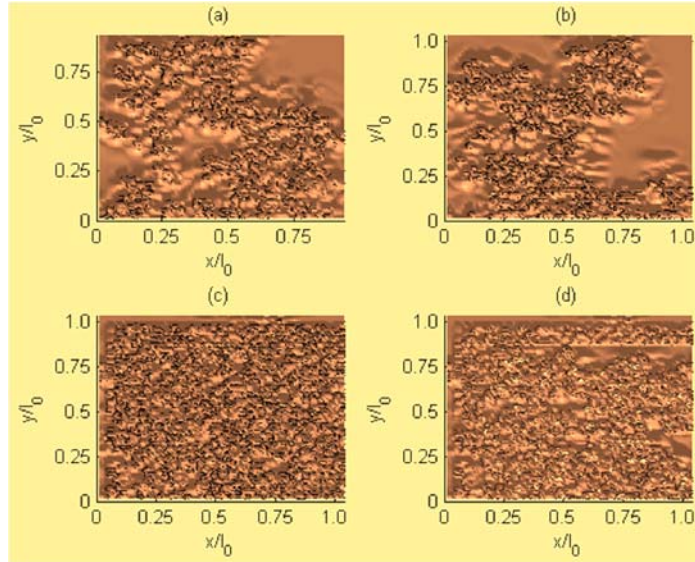


Fig. 10.5. Finite-difference time-domain calculation of the spatial distribution of the field intensity (in central intersection of sample) in 3D percolating cluster for different system sizes with (a) $L = 75$; (b) $L = 83$; (c) $L = 82$, and (d) $L = 80$.

As a result of nonlinear dynamical evolution, the areas with strong field are extended due to the merge of small-scaled structures and reconnecting of the field patterns. A coherent lasing occurs in this percolating system. Once the lasing has started, the gain no longer depends on the pumping rate but is controlled by the losses in the system. From Fig. 10.5 (c) we observe that the field is concentrated in a number of 3D spatially interconnected laser modes which practically are localized in the central part of the percolating medium.

One can see from Figs. 10.5 (a-c) that the all optical 3D modes are strongly mixed and it is difficult to separate a dominated mode among others. To get insight into the structure of the lasing mode, in the next section we will apply the Fermat variation principle that allows finding the interconnecting trajectories between the emitters.

10.3.1. The Case of Small Time

First we restrict our attention by a time t that does not exceed the critical time t_c of lasing generation, such that $t < t_c$. In this case the field amplitude E is small enough and in Eqs. (10.2)-(10.3) the nonlinear terms are negligible small.

10.3.2. Numerical 3D Simulations

In what follows we simulate the 3D radiated field distribution by FDTD technique for about indicated parameters. We will concentrate in values of the occupation probability p_c closely to percolating threshold when the incipient cluster spans all the system in first time. Since the complicity to analyze the general shape of 3D field (see Fig. 10.6), we first consider the structure of radiated field in central intersection of the system. Fig. 10.7 shows typical field distribution (in central intersection) $|E_x|$ in the percolating system with $p = 0.316$ nearly p_c for cub with $L = 100$. From Fig. 10.7 we observe that the field structure consists of two different shapes. First one corresponds to the field of disordered emitters that is depicted in Fig. 10.7 as small color point-like squares in position of emitters. Such points correspond to radiated emitters that are placed an irregular cluster and its have an expected field shape. However one can see from Fig. 10.7 yet different field spots with localized smooth shape are observed in a free-source area. For instance, one of localized mode is situated in Fig. 10.7 around of $x = 20, y = 40$ and it exhibits the amplitude of localized field. From Fig. 10.7 we observe that other localized modes are also generated beyond the source area (percolating cluster), but with lesser amplitudes. Clear that such spots correspond to optical localized modes.

Fig. 10.8 exhibits the field distribution $|E_x|$ in the percolating system with $p = 0.32$ for cub with $L = 75$ in different planes: panel (a) $z = 19$ and panel (b) $z = 32$. Again we observe various spots corresponding extended localized modes with quite large amplitudes in a source-free area.

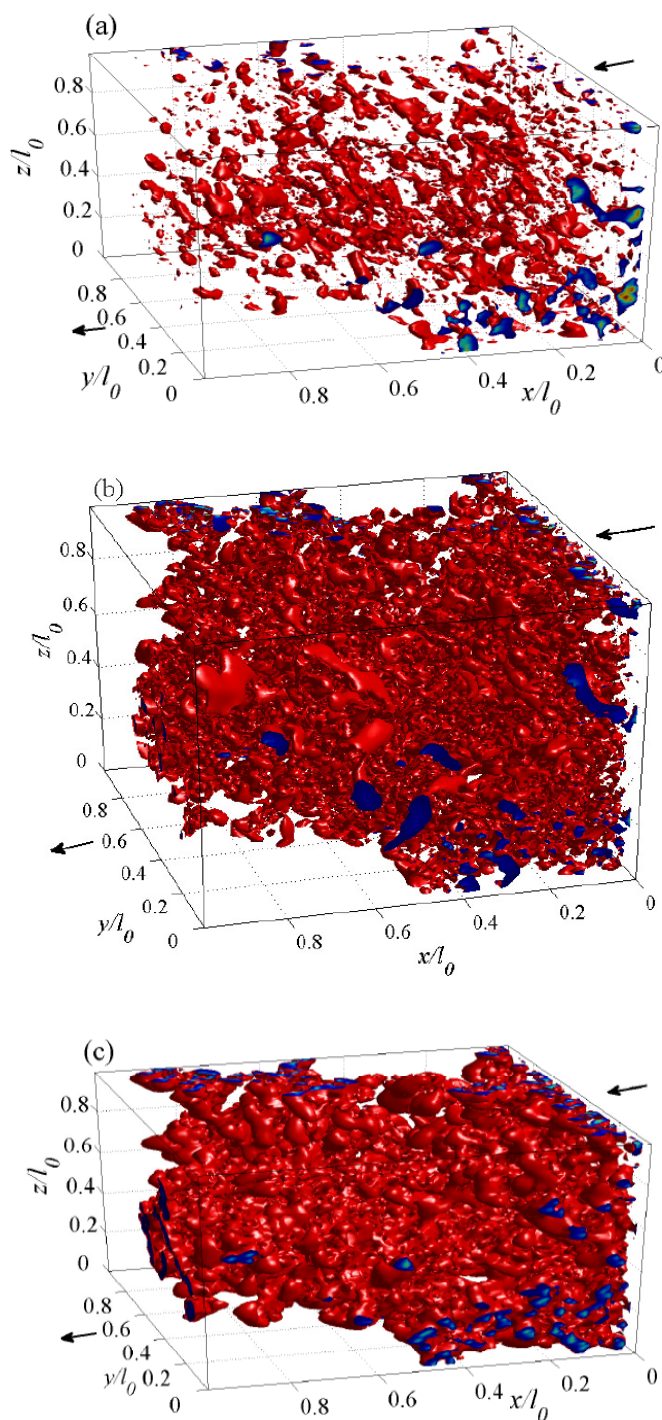


Fig. 10.6. The FDTD simulations of 3D spatial structure of the fields amplitude $|E_x|$, generated by emitters incorporated in percolating clustered (a) below the lasing regime, (b) at the lasing start time t_s and (c) for well-developed lasing.

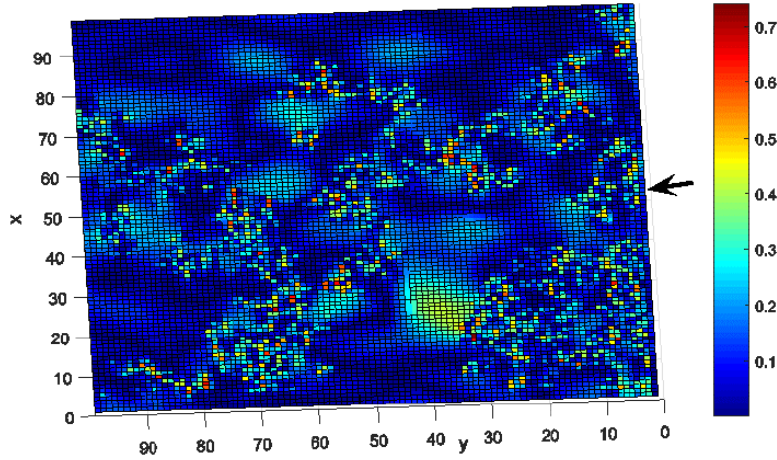


Fig. 10.7. (Color on line.) The typical field distribution (in central intersection) $|E_x|$ in the percolating system with $p = 0.316$ nearly p_c for cub with $L = 100$. In this case the number of emitters N is about 10^5 . Small color squares display the radiated field in position of emitters inside the cluster. However the spot around $x=20, y=40$ exhibits the amplitude of localized mode beyond the cluster. Other modes are generated too, but with lesser amplitudes.

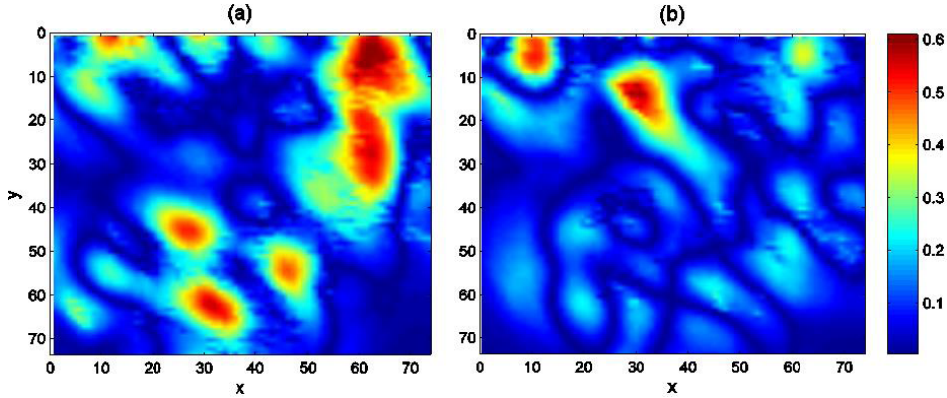


Fig. 10.8. (Color on line.) The field distribution $|E_x|$ in the percolating system with $p=0.32$ for cub with $L = 75$ in planes (a) $z = 19$ and (b) $z = 32$. We observe various spots corresponding to localized modes with quite large amplitudes in a source-free area.

More details we observe from Fig. 10.9 that shows the colorized isosurface of 3D field amplitude $|E_x|$. In this case the field is generated by emitters incorporated in a percolating cluster with probability p slightly below the percolating phase transition $p = 0.317 < p_c = 0.318$. In Fig. 10.9 the smooth green inclusions are seen clearly and they occupy rather large domains. We observe from Fig. 10.9 that such localized modes are distributed not only in area the central intersection (shown in Fig. 10.7) but also in various parts of 3D sample.

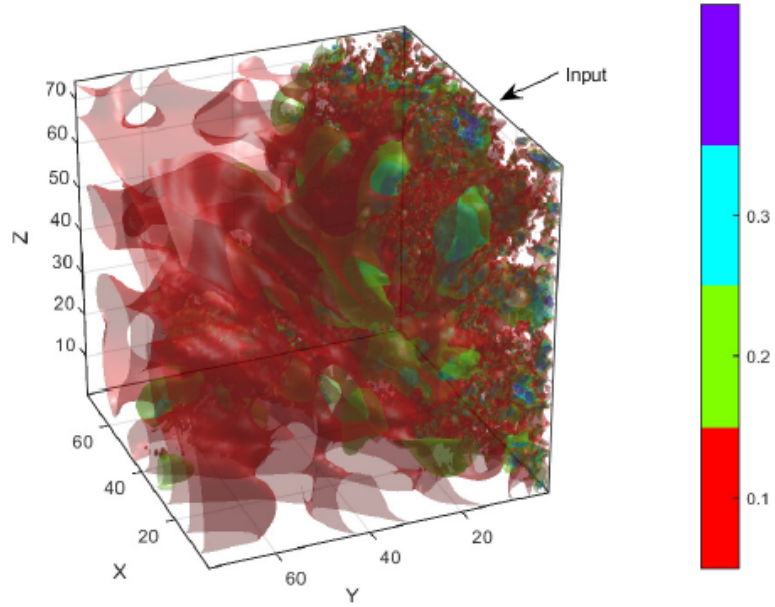


Fig. 10.9. (Color on line.) The field amplitude $|E_x|$, generated by emitters incorporated in a percolating cluster with probability p slightly below the percolating phase transition at $p = 0.317 < p_c = 0.318$.

Fig. 10.10 displays with more details the structure of some 3D localized field that are shown in Fig. 10.9. We observe from Fig. 10.10 that the field localized structures situated beyond the source area and normally have the ellipsoidal shape.

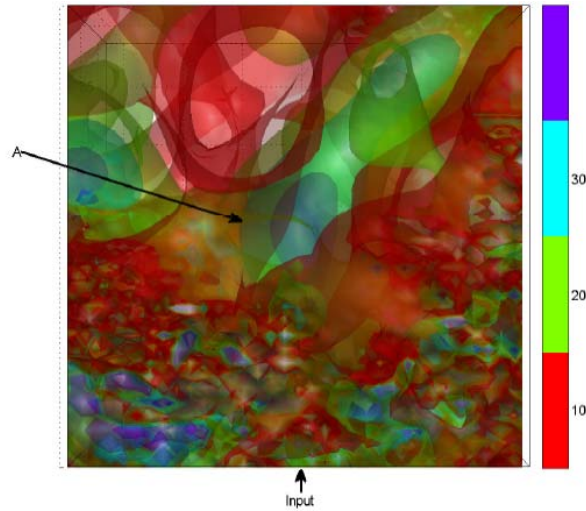


Fig. 10.10. (Color on line.) The same as in Fig. 10.9 but with detailing view of the central part that exhibits the localized Anderson mode indicated by symbol A .

10.4. Condition of Optical Localization in Percolating System

The mentioned above probability of a light wave returning to the same point (closed loops) depends on the scattering mean free path l_{av} . It is expected strong localization to occur in three dimensions when $l_{av} < \lambda / 2\pi$, which is known as the Ioffe- Regel criterion [36].

The optical field structures in 3D disordered system shown in the Figs. 10.7-10.10 (for time closely to the beginning of lasing) are generated by a significantly non-uniform distribution of radiated nanoemitters in a percolating cluster. The high level of the field localization allows rapid accumulation of nonlinear effects in this area that latter leads to a laser effect (random laser). It is of great interest to study the conditions when the localization happens. Since such phenomenon occurs in a highly disordered system it can be mentioned as one of variants of 3D optical Anderson localization. In considered percolating model similarly of widely used approaches in the theory of Anderson localization (see [37] and references therein) we will formulate the measure of localization by a dimensionless parameter g that is a ratio between two spatial scales: the mean free paths of photons in a sample \bar{l} and the wavelength of radiated emitters λ . The first parameter \bar{l} is proportional to the optimal distance l_{3N} that a photon can travel between N emitters inside a finite size 3D-sample (l_0 is a sample size) without visiting the same region twice, such that $l_{av} = l_{3N} / N$ and $\bar{l} = l_0 l_{av}$. The second parameter wavelength $\lambda_a = 2\pi c / \omega_a$ corresponds to a transition frequency ω_a of the radiated emitters incorporated in a percolating cluster. So that the parameter g reads

$$g = \bar{l} / \lambda_a = 2\pi \frac{\omega_a l_0}{c} \frac{l_{3N}}{N}. \quad (10.4)$$

Such definition of the parameter g in Eq. (10.4) leads to the following strategy of our numerical simulations: (i) First we define the 3D numerical grid as a $L \times L \times L$ nodes (where $L = 40, 50, 60, 75, 100$) and simulate the percolation for various values of the probability occupation p up to value $p = 0.34$ that slightly exceeds the critical percolation value $p_c = 0.318$; (ii) This allows to define the number and position of radiated emitters N belonging to the cluster; (iii) Then we calculate the dimensionless optimal path l_{3N} connecting such N emitters found from step (ii) with the use of Monte Carlo technique. The well-known principle of Fermat [38] asserts that the optimal optical path S between any two points r_0 and r_1 is defined by minimization of $S = \int_{r_0}^{r_1} n_r ds = \min$ where n_r is the refractive index of the media. In general, the definition of this optimal path is an advanced problem for structures with large number of randomly placed emitters. (We note that such optimization 3D problem for large N is quite involved computer task due to its NP complexity (see [39] and references therein)); (iv) The average length of the

free path is calculated as $l_{av} = l_{3N} / N$, and finally (v) we calculate the radiated field 3D structure from Eqs. (10.2)-(10.3) by FDTD technique (for more details see [22]).

Fig. 10.11 shows the calculated dimensionless average free path between optical emitters $l_{av} = l_{3N} / N$ (defined in step (iv)), where l_{3N} is the optimal optical path between N emitters for cases $L = 40, 50, 60, 75$. From Fig. 10.11 we observe that for small percolating probability $p < 0.3$ the value l_{av} is small, but also is small the number of emitters. In area of the percolation transition $p \approx 0.32$ the free path l_{3N} shows well expressed critical behavior for all used scales L . In this area the number of emitters N in the percolation cluster grows very rapidly with the increase p .

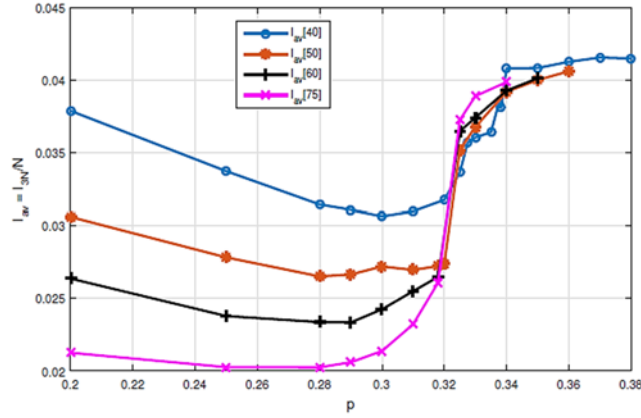


Fig. 10.11. (Color on line.) The dimensionless average free distance between field emitters $l_{av} = l_{3n} / N$ in percolating system as function of the occupation probability p , where l_{3N} is total optimal distance between N emitters (calculated by the Monte Carlo method) for 3D numerical grids L^3 with $L = 40, 50, 60, 75$.

Fig. 10.12 displays the values of parameter g from Eq. (10.4) for typical quantities: frequency of field $\omega_a = 2\pi \times 0.3 THz$, scale length $l_0 = 370 \mu m$, $\lambda = \lambda_{vac} = 62.8 \mu m$ and dimensionless free path l_{av} from Fig. 10.11. We observe that the threshold (the localization criterion) $g \sim 1$ can be achieved near the percolation threshold $p \sim 0.318$ already for $N > 50$. It is worth to note that above shown Figs. 10.7, 10.9, and 10.10 are obtained for parameters for which the localization condition $g \sim 1$ is fulfilled. This confirms the validity of above derived condition $g \sim 1$ for threshold of the optical field localization. Also one can see from Fig. 10.12 that for the overcritical case $p > p_c$ takes place $g > 1$, so for larger p the field localization can not be observed.

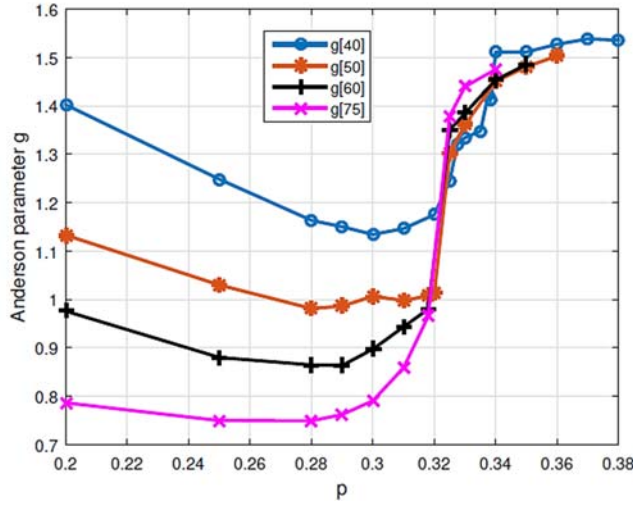


Fig. 10.12. (Color online.) The values parameter g from Eq. (10.4) for typical quantities: frequency of field $\omega_a = 2\pi \times 0.3 \text{ THz}$, scale length $l_0 = 370 \mu\text{m}$, $\lambda = \lambda_{vac} = 62.8 \mu\text{m}$ and dimensionless free path l_{av} from Fig. 10.11. We observe that the criterion $g \approx 1$ can be satisfied near the percolation threshold $p \approx 0.318$ already for $N > 50$. The increase of probability p leads to very long value of optimal path S , so becomes $g > 1$ and the localization does not occur.

10.5. Inverse Participation Ratio

Now we study if the field localization can be observed in area beyond the percolation phase transition. To evaluate quantitatively the degree of field localization, we calculate the inverse participation ratio (IPR) defined as

$$I = l_0^3 \frac{\int |E|^4 d^3r}{\int |E|^2 d^3r}. \quad (10.5)$$

From Eq. (10.5) one can see that I is close to 1 for smooth homogeneous fields, and $I < 1$ for inhomogeneous distributed fields. Fig. 10.13, 10.14 shows the values $I(p)$ and also the order parameters of percolation $P(p)$ as function of the occupation (percolating) probability p for different grids with $L = 75, 100$. We observe that $I(p) \sim 0.5$ up to percolation transition zone for $p < p_c$ and $I(p)$ changes sharply at $p \approx p_c$ where the percolating phase transition occurs. (Figs. 10.17, 10.18 display corresponded optical fields.) For larger $p > p_c$ the value IPR is $I(p) \sim 0.8$, see also Figs. 10.19, 10.20. That is in a good agreement with the above formulated condition of the field localization and explains why such effect occurs in area of the percolation phase transition p_c . Our FDTD numerical simulations confirm this result.

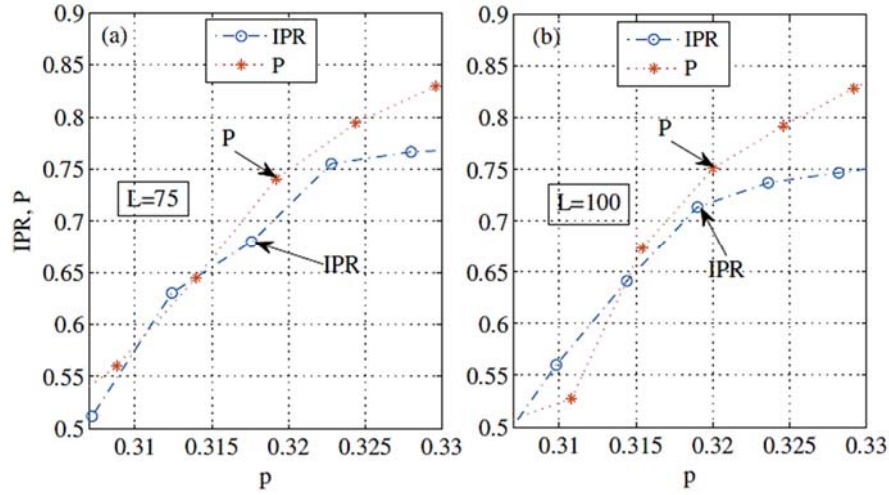


Fig. 10.13. (Color online.) The inverse participation ratio IPR $I(p)$ for radiated field (see Eq. (10.5)) and the order parameters $I(p)$ of the spanning percolation cluster as function of the occupation probability p in area of percolation phase transition $p \approx p_c = 0.318$ for 3D system $L \times L \times L$ with (a) $L = 75$; (b) $L = 100$.

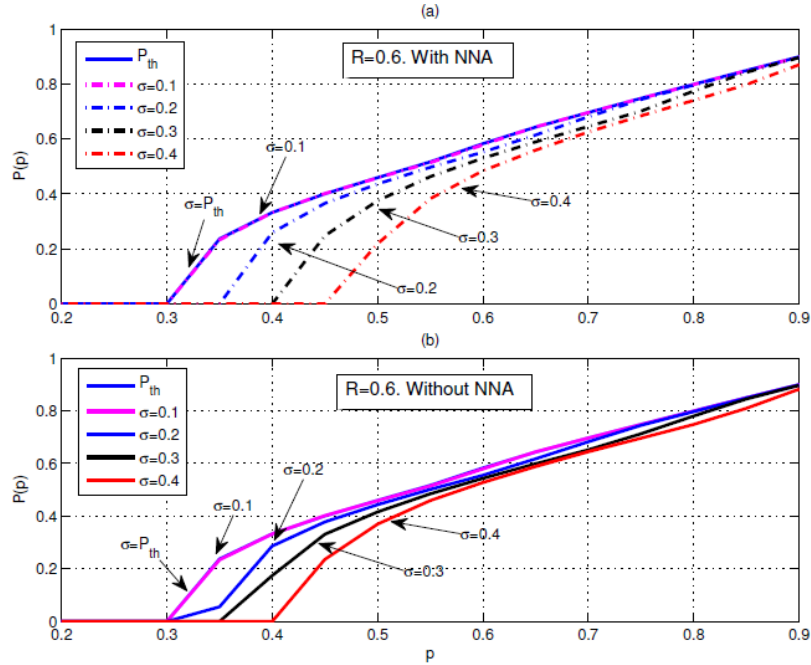


Fig. 10.14. (Color online.) The order parameter $P(p)$ of the spanning percolation cluster for pores with fixed $R = r = 0.6 > r_c$ and different random displacement with $\sigma = \sigma_s = 0.1, 0.2, 0.3$ and 0.4 . (a) With nearest neighbors approximation (NNA); (b) without NNA. For large value $\sigma_s \geq 0.3$ the difference between both cases becomes considerable.

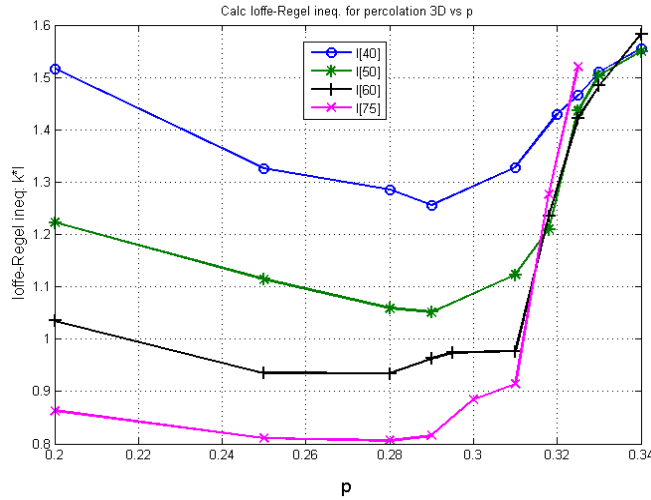


Fig. 10.15. (Color online.) The values of parameter g from Eq. (10.4). We consider the radius of the pores fixed $r=0.6$ and random displacement; $s=0.1$. We observe that the criterion $g \approx 1$ can be satisfied near the percolation threshold $p \approx 0.31$ already for $N > 60$. The increase of probability p leads to very long value of optimal path S , so becomes $g > 1$ and the localization does not occur.

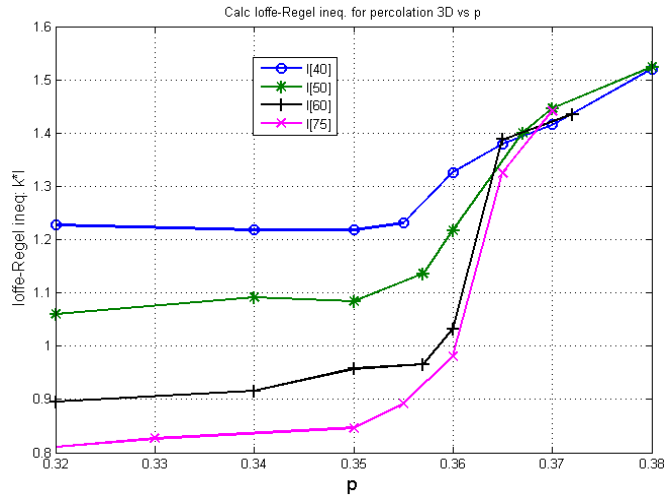


Fig. 10.16. (Color online.) The values of parameter g from Eq. (10.4). We consider the radius of the pores fixed $r=0.6$ and random displacement; $s=0.2$. We observe that the criterion $g \approx 1$ can be satisfied near the percolation threshold $p \approx 0.355$ already for $N > 60$. The increase of probability p leads to very long value of optimal path S , so becomes $g > 1$ and the localization does not occur.

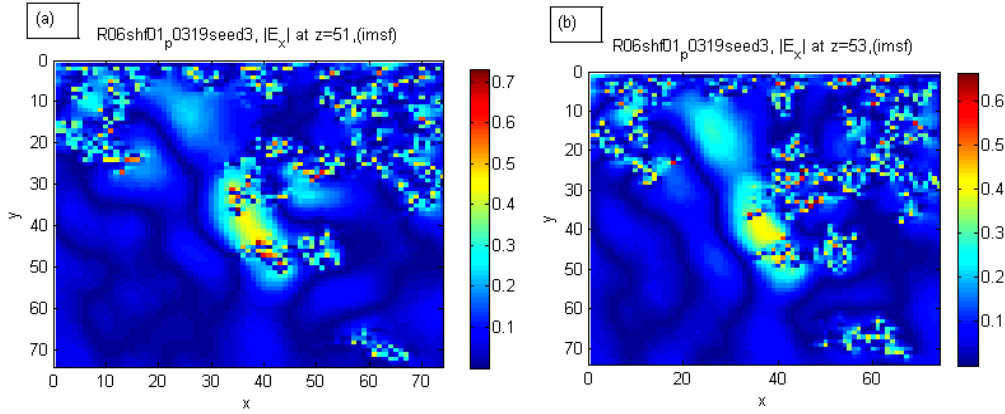


Fig. 10.17. (Color online.) The typical field distribution (in central intersection) $|E_x|$ in the percolating system with $p = 0.319$ nearly p_c for cub with $L = 75$ in planes (a) $z = 51$ and (b) $z = 53$. Small color squares display the radiated field in position of emitters inside the cluster. However the spot around $x = 40, y = 40$ exhibits the amplitude of localized mode beyond the cluster. Random displacement is applied with $s = 0.1$.

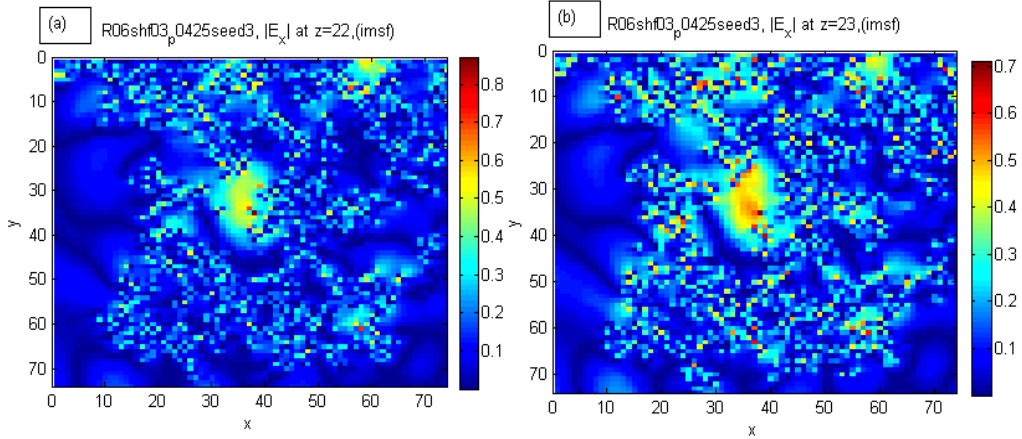


Fig. 10.18. (Color online.) The typical field distribution (in central intersection) $|E_x|$ in the percolating system with $p = 0.425$ nearly p_c for cub with $L = 75$ in planes (a) $z = 22$ and (b) $z = 23$. Small color squares display the radiated field in position of emitters inside the cluster. However the spot around $x = 40, y = 30$ exhibits the amplitude of localized mode beyond the cluster. The random displacement is applied with the standard deviation $s = 0.3$.

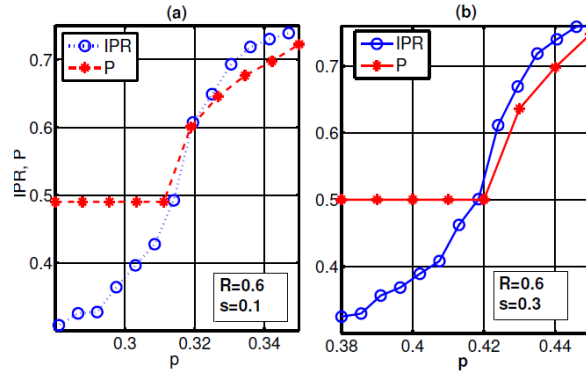


Fig. 10.19. (Color online.) The inverse participation ratio $I(p)$ for the radiated field and the order parameters $P(p)$ of the initial percolation cluster as a function of the occupation probability p . We consider the radius of the pores is fixed $r=0.6$ and random displacement has standard deviations: (a) $s=0.1$, (b) $s=0.3$. Percolation model with nearest neighbors approximation are applied, for 3D system $L \times L \times L$, with $L = 75$.

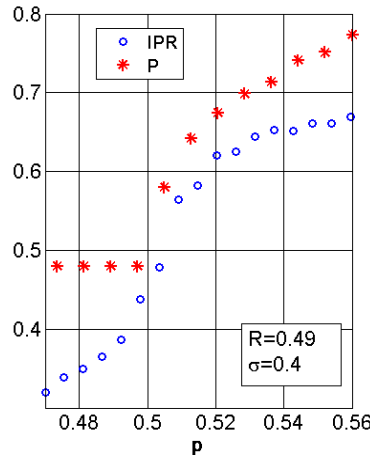


Fig. 10.20. (Color online.) The inverse participation ratio $I(p)$ for the radiated field and the order parameters $P(p)$ of the initial percolation cluster as a function of the occupation probability p . We consider random radius of the pores ($R=0.49$), with the standard deviation $\sigma=0.4$ and fixed pore's displacement. Percolation model with nearest neighbors approximation are applied, for 3D system $L \times L \times L$, with $L = 75$.

10.6. Localization for Large Time

In this section we consider the properties of localization at large time $t > t_c$, where t_c is the time when the random lasing begins. The results are shown in Figs. 10.21-10.38.

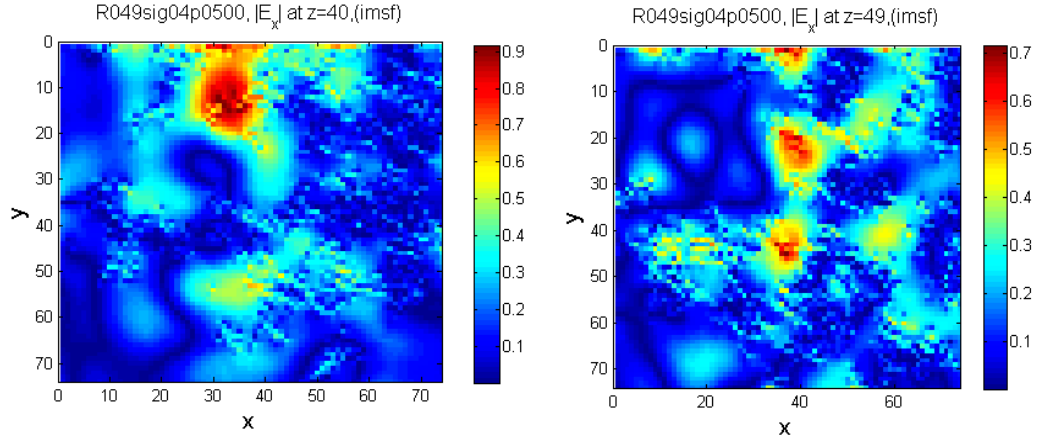


Fig. 10.21. (Color on line.) The field distribution $|E_x|$, in the percolating system with $p = 0.5$ for cube with $L = 75$ in planes (a) $z = 40$ and (b) $z = 49$. We observe various spots corresponding to localized modes with quite large amplitudes in a source-free area. For random pores the radius $R = 0.49$ is considered.

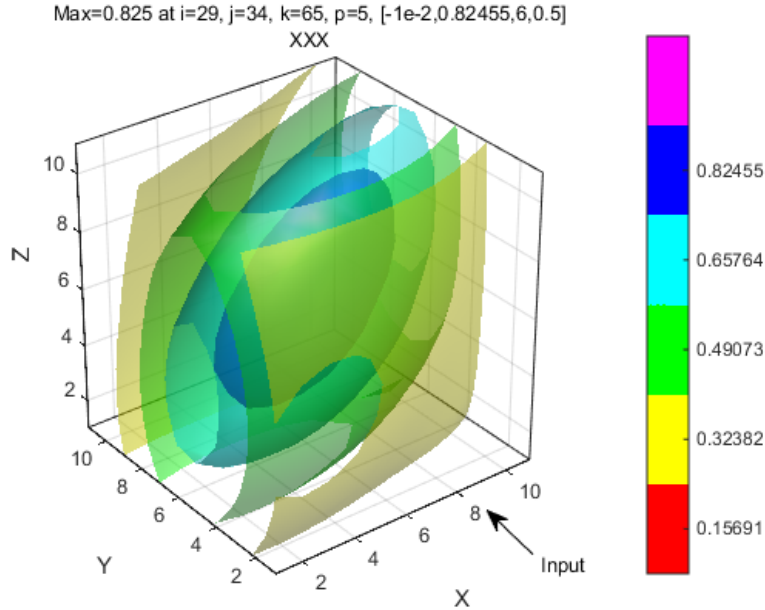


Fig. 10.22. The shape of the field amplitude $|E_x|$, corresponding to a localized mode in the percolating system, with $p = 0.65$ (in cube with $L = 100$) localized in coordinates $(29, 34, 65)$ at time $t = 17$. The random pores are considered with mean $R = 0.49$ and standard deviation $\sigma = 0.04$.

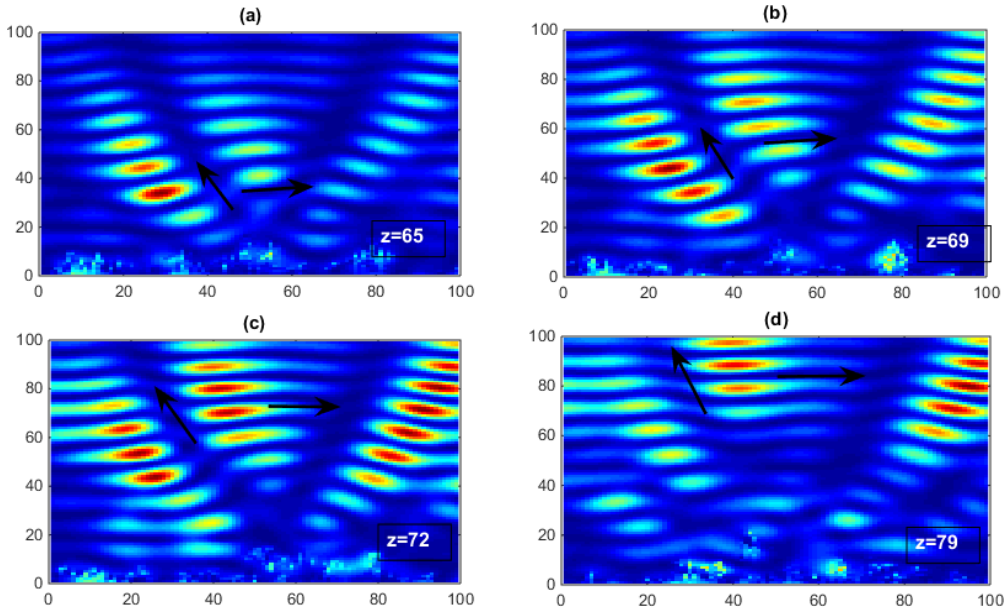


Fig. 10.23. The field distribution $|E_x|$ in the plane intercessions of the percolating system with $p = 0.42$ (cube with $L = 100$) in planes (a) $z = 65$, (b) $z = 69$, (c) $z = 72$ and (d) $z = 79$; at time $t = 17$. We observe various spots corresponding to localized modes, with quite amplitudes in a source-free area. Random pores have mean $R = 0.49$ and the standard deviation $s = 0.3$

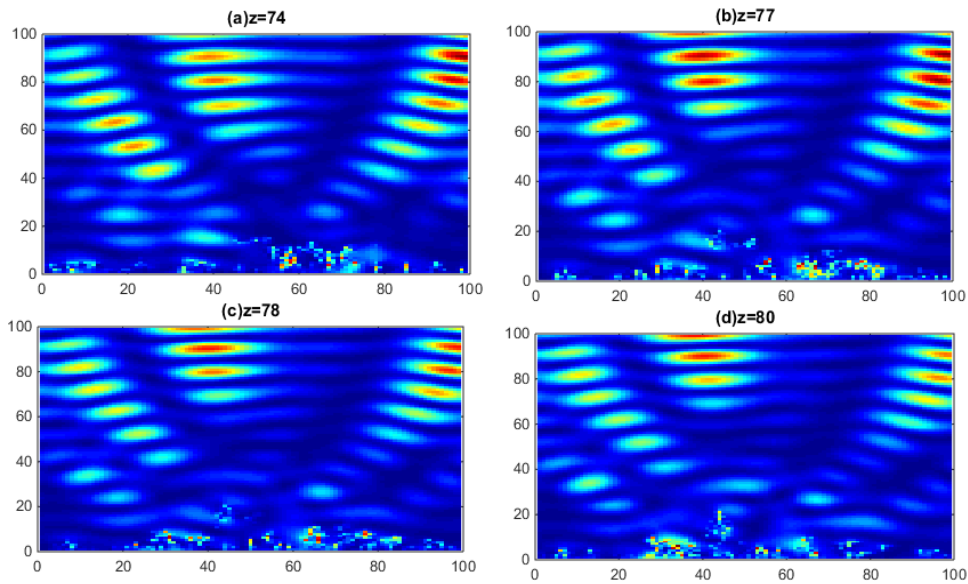


Fig. 10.24. The field distribution $|E_x|$ in the percolating system with $p = 0.42$ for a cube with $L = 100$ in planes (a) $z = 74$, (b) $z = 77$, (c) $z = 78$ and (d) $z = 80$; at time $t = 18$. Comparing with Fig. 10.23, the field undergoes a displacement toward the edges of the medium. Random pores have mean $R = 0.49$. See details in the text.

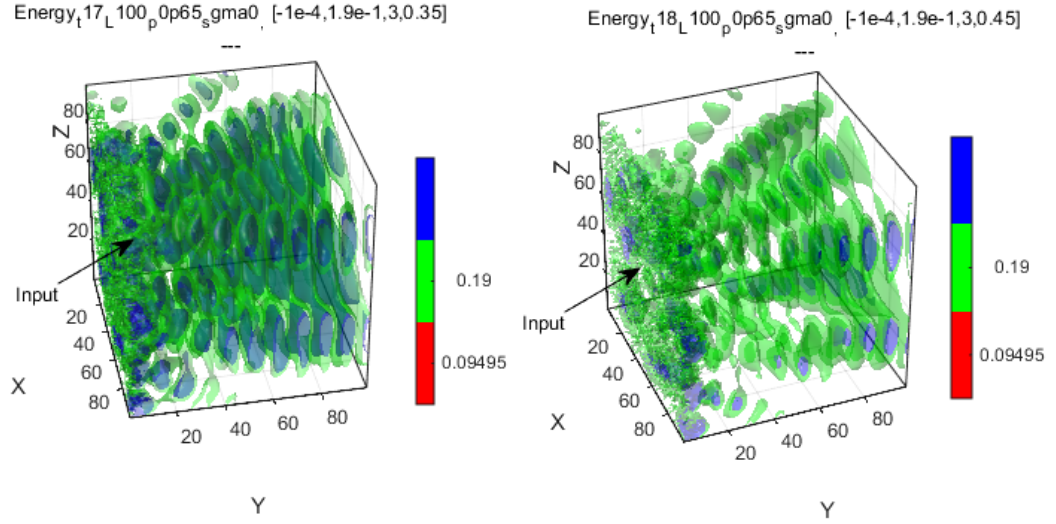


Fig. 10.25. The field amplitude $|E_x|$, generated by emitters incorporated in a percolating cluster with occupation probability $p = 65$ at time (a) $t = 17$ and (b) $t = 18$. We observe that at time $t = 18$, these modes move toward the boundary of the medium (output). Random pores is considered with mean $R = 0.49$ and the standard deviation $s = 0.3$.

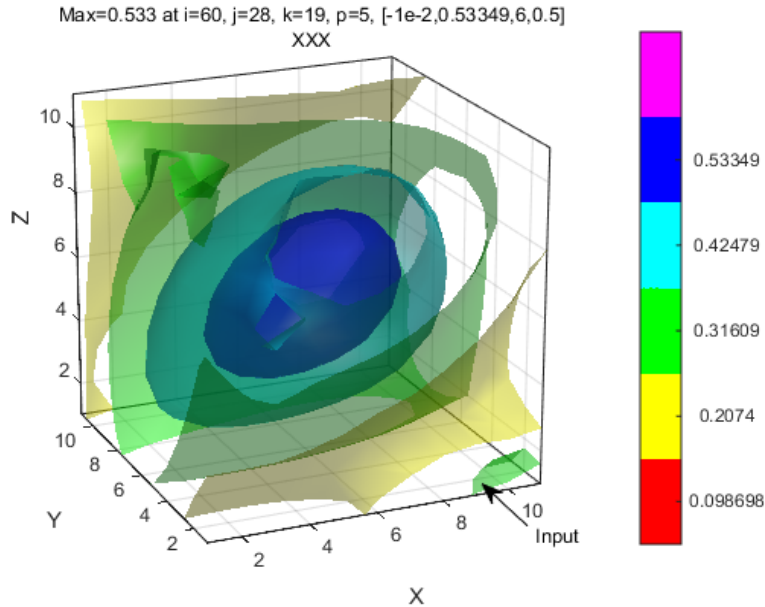


Fig. 10.26. The field shape $|E_x|$, showing a localized mode in radiated percolating system with $p = 0.7$ for a cube with $L100$ and localized in coordinates $(60, 28, 19)$ at time $t = 19$. Random pores have mean $R = 0.49$ and the standard deviation $s = 0.3$.

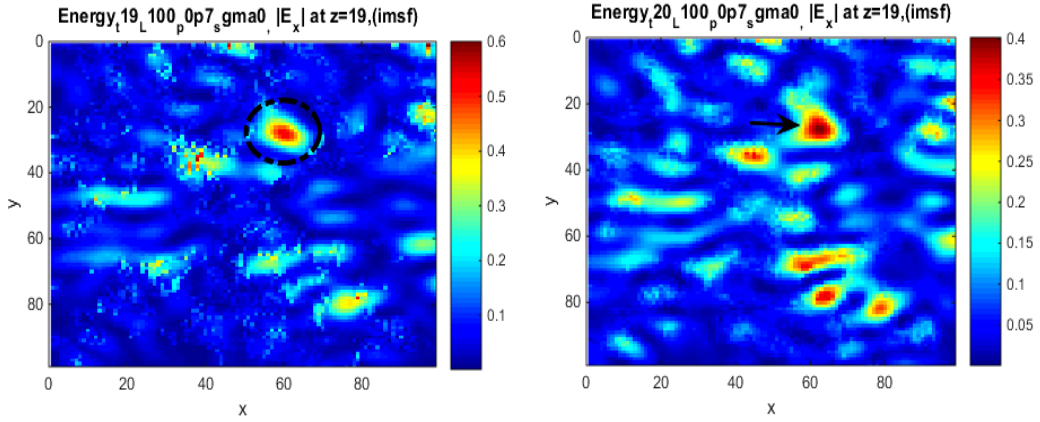


Fig. 10.27. The field distribution $|E_x|$ in the percolating system with $p = 0.7$ for a cube with $L = 100$ at time (a) $t = 19$, (b) $t = 20$. We observe various field spots corresponding to localized modes, with quite large amplitudes in a source-free area. Small displacement of the mode localized in the x, y plane $(60, 28)$ to the right side is observed. Random pores have mean $R = 0.49$ and $\sigma = 0.04$.

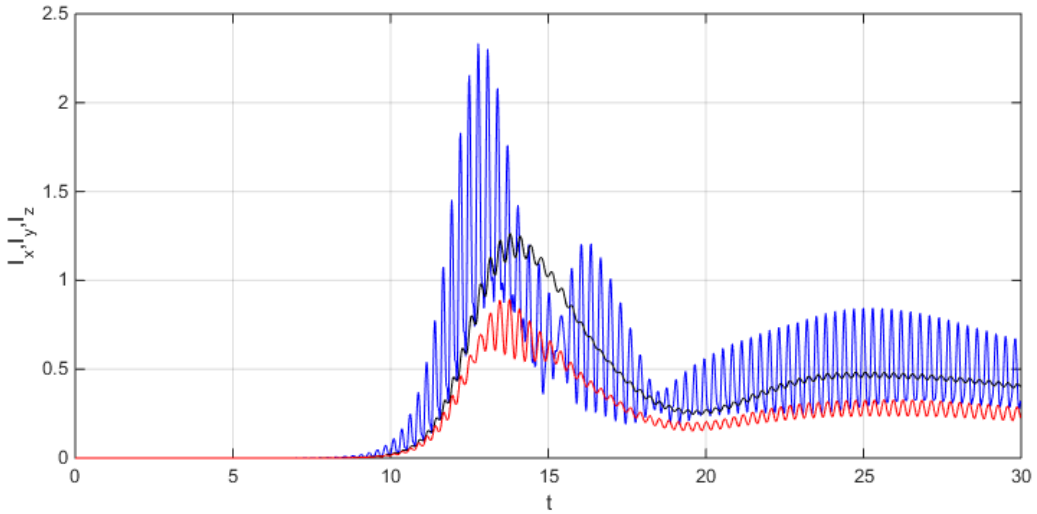


Fig. 10.28. The dynamics of laser generation for percolating probability $p = 0.47$. The picture shown the energy fluxes I_x, I_y and I_z , in the x, y, z directions of the system, with random pores having mean $R = 0.49$ and standard deviation. The numerical grid L^3 with $L = 100$ has been used.

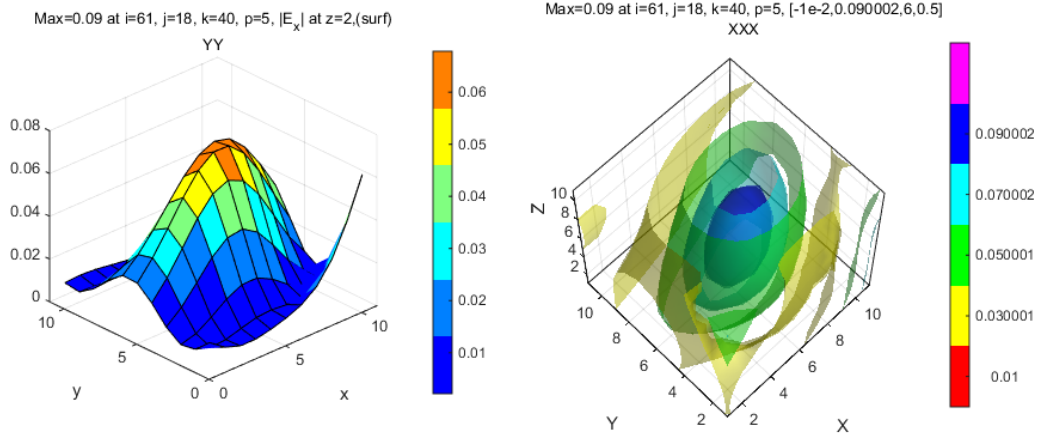


Fig. 10.29. (a) The amplitude of the field mode $|E_x|$ in the percolating system with $p = 0.47$ at time $t = 8$. (b) the ellipsoidal modes localized at the position (61, 18, 40) of the 3D medium. Random pores have mean $R = 0.49$ and the standard deviation $s = 0.3$

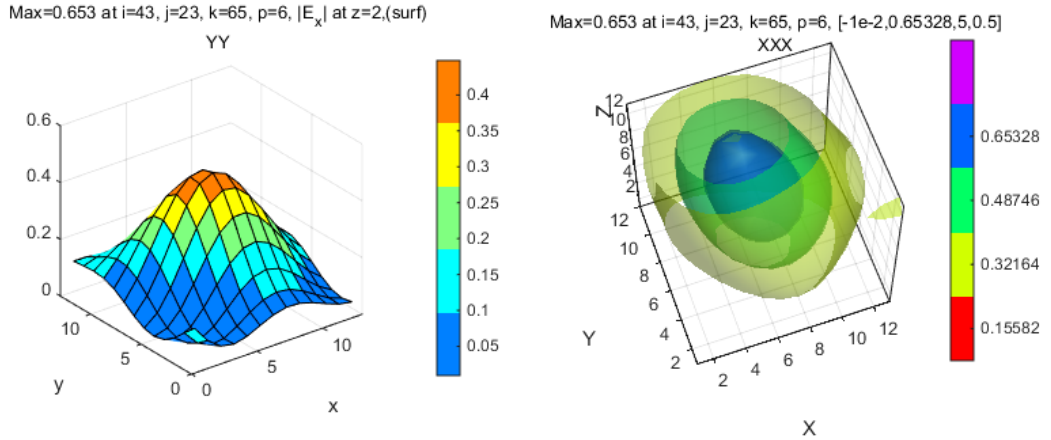


Fig. 10.30. (a) Amplitude of the localized mode $|E_x|$ in the percolating system with $p = 0.47$ at time $t = 15$. (b) One of the ellipsoidal modes localized at the coordinates (43, 23, 65) in the 3D system. We observe that in comparing with Fig. 10.29 when there is laser generation starts, the normalized amplitude of the field increases to $\text{Max} = 0.653$. Random pores have mean $R = 0.49$.

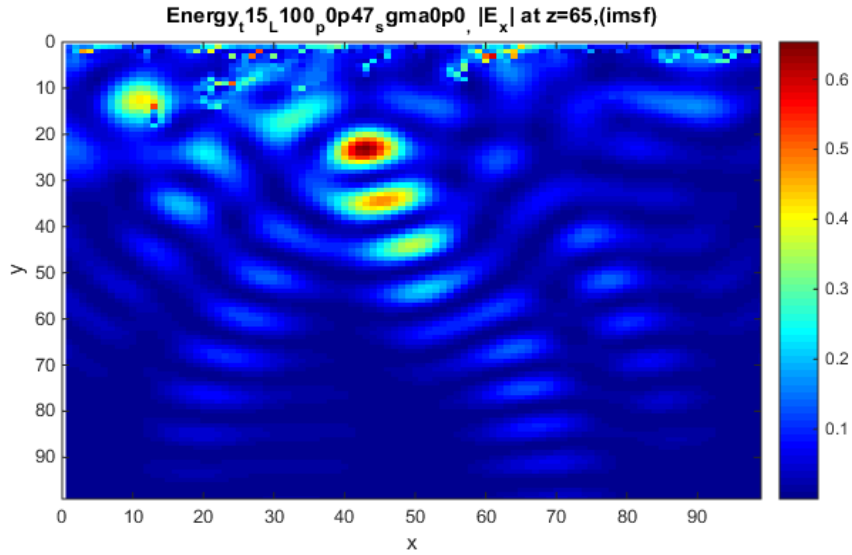


Fig. 10.31. The field distribution $|E_x|$ in the percolating system with $p = 0.47$ for a cube with $L = 100$ at time $t = 15$. We observe various field spots corresponding to localized modes, with quite large amplitudes in a source-free area. In particular we observe here the mode showing in Fig. 10.30, in the xy coordinates $(43, 23)$ at plane $z = 65$. Random pores have mean $R = 0.49$.

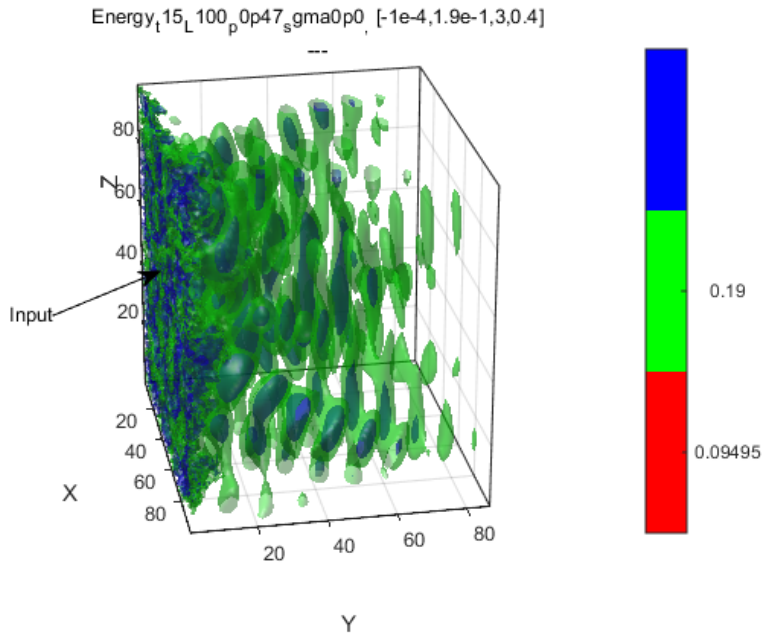


Fig. 10.32. The field amplitude $|E_x|$, generated by emitters incorporated in a percolating cluster with probability $p = 0.47$ at time $t = 15$. We observed some well-defined field bunches within the 3D percolation structure. Random pores have mean $R = 0.49$ and $s = 0.3$.

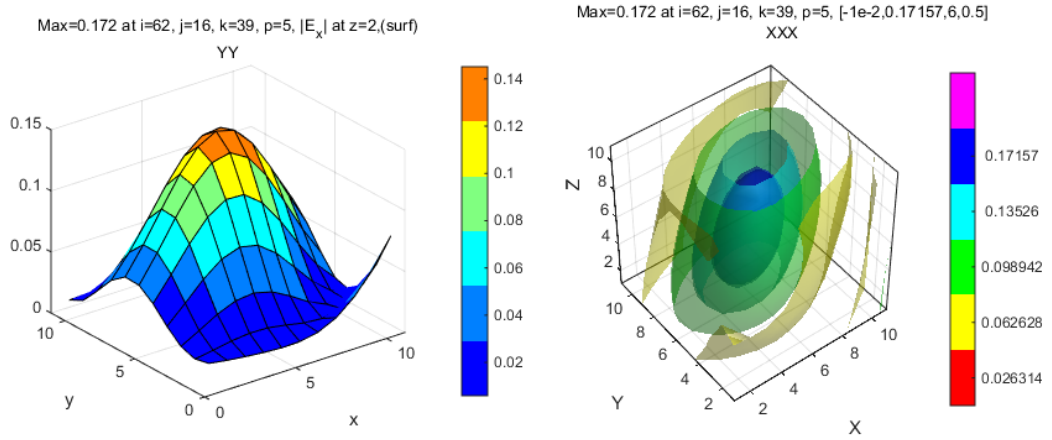


Fig. 10.33. (a) Amplitude of the mode $|E_x|$ in the percolating system with $p = 0.47$ at time $t = 26$. (b) One of the ellipsoidal modes localized at the coordinates (62, 16, 39) in the 3D system. Random pores have mean $R = 0.49$.

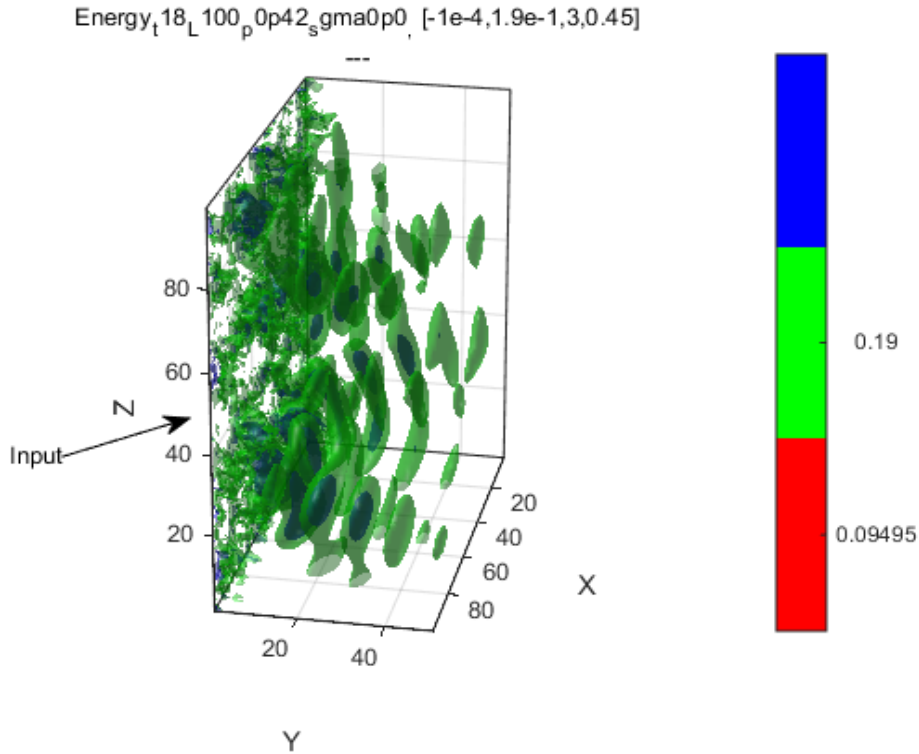


Fig. 10.34. The field amplitude $|E_x|$, generated by emitters incorporated in a percolating cluster with probability $p = 0.42 < p_c$ at time (a) $t = 18$. We observed some well-defined field bunches within the 3D percolation structure. Random pores have mean $R = 0.49$ and the deviation $s = 0.3$.

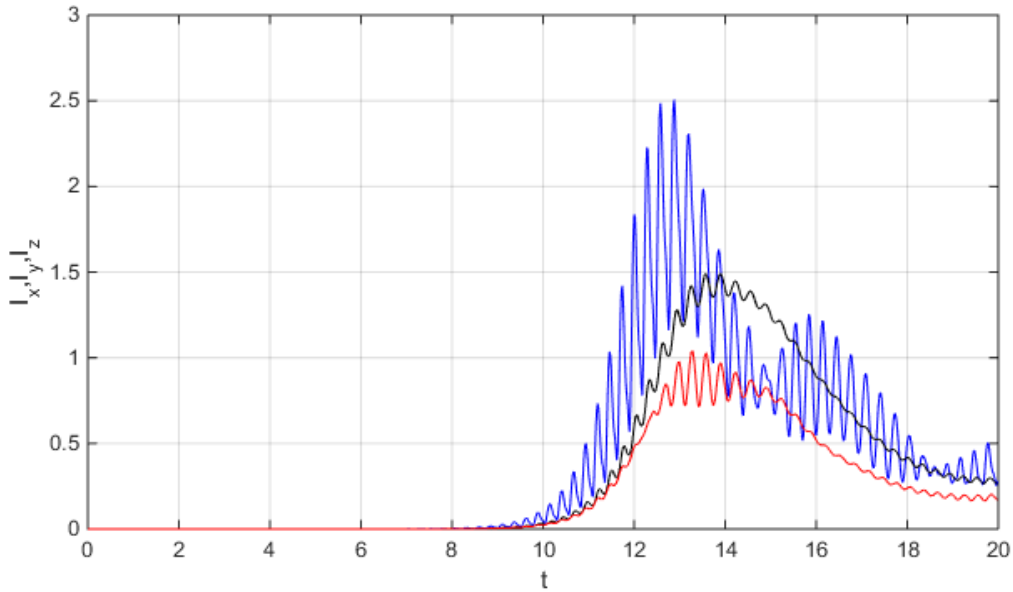


Fig. 10.35. The laser generation in percolating system with probability $p = 0.47$. The picture shown the field fluxes I_x , I_y and I_z , in x , y , z direction of the system, with random pores having $R = 0.49$ and the numerical grid with $L = 100$ has been used.

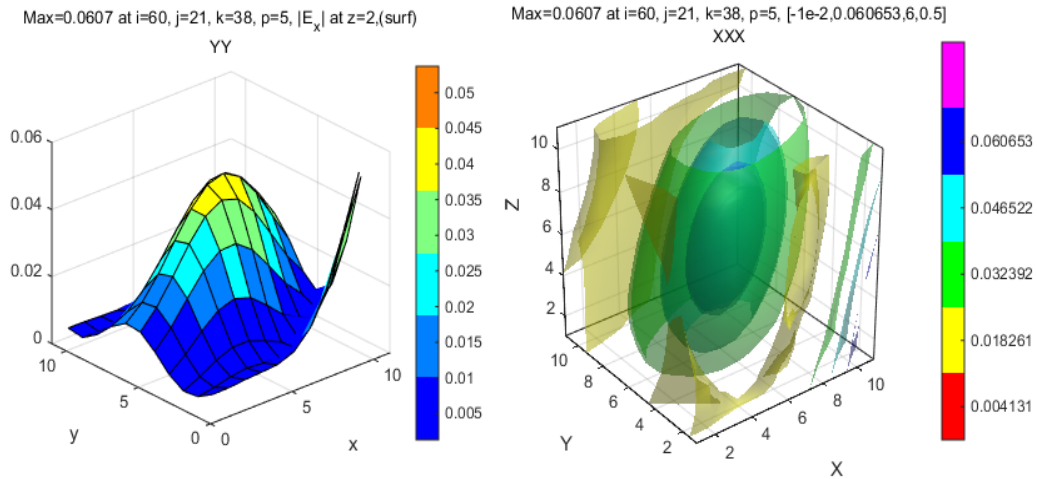


Fig. 10.36. (a) Amplitude of the mode $|E_x|$ in the percolating system with $p = 0.47$ at time $t = 4 < t_s$, where t_s is the time when the laser generation starts. (b) One of the ellipsoidal modes localized at the coordinates $(60, 21, 38)$ of the 3D system. Random pores have mean $R = 0.49$.

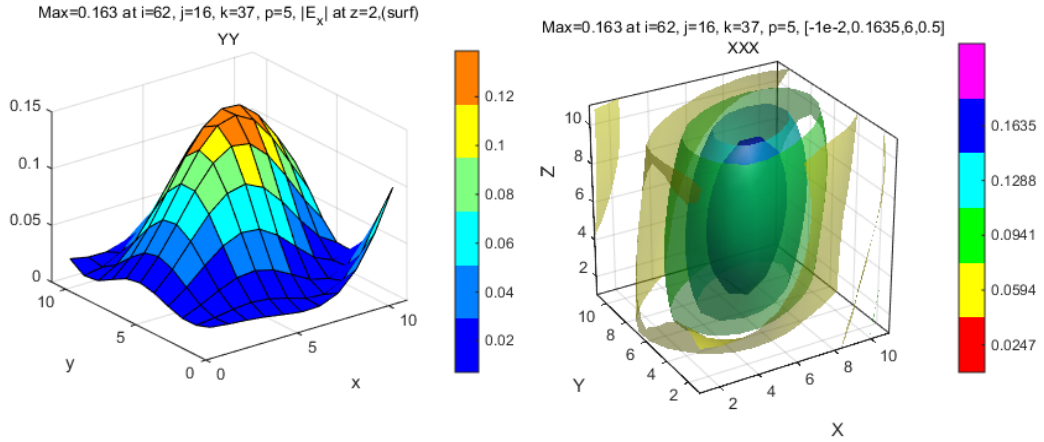


Fig. 10.37. (a) Amplitude of the mode $|E_x|$ in the percolating system with $p = 0.47$ at time $t = 19 > t_s$, where $t_s = 10$ is the laser generation time start. (b) One of the ellipsoidal modes localized at the coordinates (62, 16, 37) in the 3D system. In comparing with Fig. 10.36; where the laser generation starts at $t = 15$, we observe that the normalized amplitude of the field increases to $\text{Max} = 0.163$. Random pores have mean $R = 0.49$.

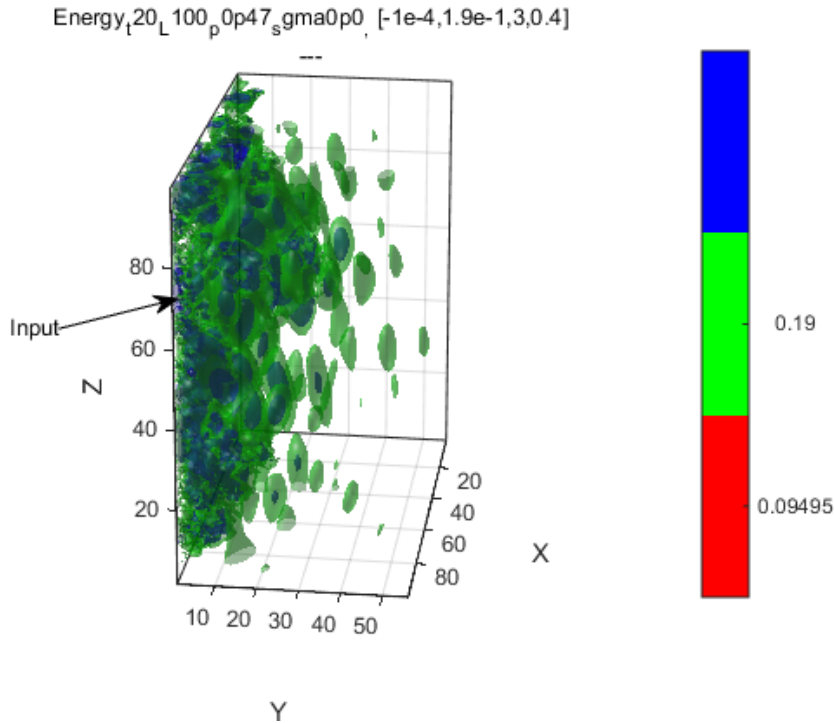


Fig. 10.38. The field amplitude $|E_x|$ generated at time $t = 20$ by the emitters incorporated in a percolating cluster with probability $p = 0.47$. We observed some well-defined field bullets within the 3D percolation structure. Random pores have mean $R = 0.49$.

10.7. Discussion

In this section we briefly discuss our results obtained by FDTD simulation of the generated optical fields whose dynamics are shown in Figs. 10.21-10.38. In 3D disordered percolating materials it is possible to incorporate the light nanoemitters by forcing the gas of emitters through the percolating channels (the optical percolation was registered in [20]). The external optical pumping will invert the 4-level integrated emitters to excited state where the laser generation becomes possible (random laser [22]). Due to fractal structure of the incipient percolating cluster (see Fig. 10.1) the emitters field has an intricate shape that allows forming the localized Anderson modes (see Figs. 10.7-10.10), and finally leads to increasing of the lasing lifetime. From the figures we observe that Anderson localization (AL) regions (having smooth shape) are normally beyond of point-like fields of the radiated emitters incorporated into percolating clusters. Our 3D FDTD simulations show that both Anderson localized field and the sharp field peaks (in positions of emitters) can be met inside of the cluster, see Figs. 10.13—10.38. The 3D optical Anderson localization criterion shows that such effect occurs closely to the critical value of the occupation probability of percolating, Fig. 10.11. The system considered here has some advantages. Disorder in a percolating material is created naturally and therefore does not require any special technology. Besides, such system of percolating light nanoemitters can be repeatedly re-pumped that allows increasing the lifetime of such compact device.

10.8. Conclusions

We investigated the optical Anderson localization associated with the properties of three-dimensional (3D) disordered percolation system, where the percolating clusters are filled by active media composed by light noncoherent emitters. We numerically studied 3D field structures where the wave localization occur and derived the expression for localization criterion that extends the optical Anderson localization condition to the percolating systems. The analysis of the Inverse participation ratio (IPR) shows that the localization arises closely to the threshold of 3D percolation phase transition and has critical behavior.

References

- [1]. F. Riboli, N. Caselli, S. Vignolini, F. Intonti, K. Vynck, P. Barthelemy, A. Gerardino, L. Balet, L. H. Li, A. Fiore, M. Gurioli, D. S. Wiersma, Engineering of light confinement in strongly scattering disordered media, *Nature Materials*, Vol. 13, Issue 7, 2014, pp. 720-725.
- [2]. E. Akkermans, G. Montambaux, *Mesoscopic Physics of Electrons and Photons*, 1st Ed., Cambridge University Press, 2007.
- [3]. K. Vynck, M. Burreli, F. Riboli, D. S. Wiersma, Photon management in two-dimensional disordered media, *Nature Materials*, Vol. 11, 2012, pp. 1017-1022.
- [4]. P. Sheng, *Introduction to Wave Scattering, Localization and Mesoscopic phenomena*, 2nd Ed., Springer, 2010.
- [5]. J. Wang, A. Z. Genack, Transport through modes in random media, *Nature*, Vol. 471, 2011, pp. 345-348.
- [6]. D. S. Wiersma, Disordered photonics, *Nature Photonics*, Vol. 7, 2013, pp. 188-196.

- [7]. D. Vollhardt, P. Wolfle, Scaling equations from a self-consistent theory of Anderson localization, *Phys. Rev. Lett.*, Vol. 48, 1982, pp. 699-702.
- [8]. P. Sebbah, D. Sornette, C. Vanneste, Anomalous diffusion in two-dimensional Anderson-localization dynamics, *Physical Review B*, Vol. 48, 1993, pp. 12506-12510.
- [9]. M. M. Sigalas, C. M. Soukoulis, C.-T. Chan, D. Turner, Localization of electromagnetic waves in two-dimensional disordered systems, *Physical Review B*, Vol. 53, 1996, pp. 8340-8348.
- [10]. T. Schwartz, G. Bartal, S. Fishman, M. Segev, Transport and Anderson localization in disordered two-dimensional photonic lattices, *Nature*, Vol. 446, 2007, pp. 52-55.
- [11]. F. Riboli, P. Barthelemy, S. Vignolini, F. Intonti, A. D. Rossi, S. Combrie, D. S. Wiersma, Anderson localization of near-visible light in two dimensions, *Optics Letters*, Vol. 36, 2011, pp. 127-129.
- [12]. P. Sebbah, C. Vanneste, Random laser in the localized regime, *Physical Review B*, Vol. 66, 2002, p. 144202.
- [13]. H. Noh, J. Yang, S. F. Liew, M. J. Rooks, G. S. Solomon, H. Cao, Control of lasing in biomimetic structures with short-range order, *Phys. Rev. Lett.*, Vol. 106, 2011, 183901.
- [14]. D. S. Wiersma, The physics and applications of random lasers, *Nat. Phys.*, Vol. 4, 2008, pp. 359-367.
- [15]. A. K. Sarychev, Anderson localization of surface plasmons and nonlinear optics of metal-dielectric composites, *Phys. Rev. B*, Vol. 60, 1999, pp. 16389-16408.
- [16]. F. Jendrzejewski, A. Bernard, K. Muller, P. Cheinet, et. al., Three-dimensional localization of ultracold atoms in an optical disordered potential, *Nature Physics*, Vol. 8, 2012, pp. 398-403.
- [17]. S. E. Skipetrov, Finite-size scaling analysis of localization transition for scalar waves in a three-dimensional ensemble of resonant point scatterers, *Phys. Rev. B.*, Vol. 94, Issue 6, 2016, 064202.
- [18]. S. Faez, A. Strybulevych, J. H. Page, A. Lagendijk, B. A. van Tiggelen, Observation of multifractality in Anderson localization of ultrasound, *Phys. Rev. Lett.*, Vol. 103, 2009, p. 155703.
- [19]. Lu, F.-F., Wang, C.-F., Anderson Localization in the Induced Disorder System, *Chinese Physics Letters.*, 33, 7, 2016, 074202.
- [20]. G. Burlak, M. Vlasova, P. A. Marquez Aguilar, L. Xixitla-Cheron, Optical percolation in ceramics assisted by porous clusters, *Opt. Commun.*, Vol. 282, 2009, pp. 2850-2856.
- [21]. G. Burlak, A. Diaz-de-Anda, Yu. Karlovich, A. B. Klimov, Critical behavior of nanoemitter radiation in a percolation material, *Phys. Lett. A*, Vol. 373, 2009, pp. 1492-1499.
- [22]. G. Burlak, Y. G. Rubo, Mirrorless lasing from light emitters in percolating clusters, *Phys. Rev. A*, Vol. 92, 2015, 013812.
- [23]. P. W. Anderson, Absence of diffusion in certain random lattices, *Phys. Rev.*, Vol. 109, 1958, pp. 1492-1505.
- [24]. A. E. Siegman, Lasers, *Mill Valley*, California, USA, 1986.
- [25]. X. Jiang, C. M. Soukoulis, Time dependent theory for random laser, *Phys. Rev. Lett.*, Vol. 85, 2000, 70.
- [26]. M. Sargent III, M. O. Scully, W. E. Lamb Jr., *Laser Physics*, Addison-Wesley, Reading, Mass., 1974.
- [27]. C. Conti, A. Fratalocchi, Dynamic light diffusion three-dimensional Anderson localization and lasing in inverted opals, *Nat. Phys.*, Vol. 4, 2008, pp. 794-798.
- [28]. M. A. Noginov, J. Novak, D. Grigsby, L. Deych, Applicability of the diffusion model to random laser with non-resonant feedback, *J. Opt. A: Pure Appl. Opt.*, Vol. 8, 2006, pp. S285-S295.
- [29]. A. Taflove, S. C. Hagness, Computational Electrodynamics: The Finite-Difference Time-Domain Method, 3rd Edition, *Artech House Publishers*, 2005.

- [30]. H. Cao, Y. G. Zhao, S. T. Ho, E. W. Seelig, Q. H. Wang, R. P. H. Chang, Random laser action in semiconductor powder, *Phys. Rev. Lett.*, Vol. 82, 1999, pp. 2278-2281.
- [31]. J. Sanghera, W. Kim, G. Villalobos, B. Shaw, C. Baker, J. Frantz, B. Sadowski, I. Aggarwal, Ceramic laser materials, *Materials*, Vol. 5, 2012, pp. 258-277.
- [32]. J. Wang, Z. Zhou, W. Zhang, T. M. Garoni, Y. Deng, Bond and site percolation in three dimensions, *Phys. Rev. E*, Vol. 87, 2013, 052107.
- [33]. D. Meschede, Optics, Light and Lasers, *Wiley-VCH*, New York, 2007.
- [34]. A. K. Sarychev, V. M. Shalaev, Electromagnetic field fluctuations and optical nonlinearities in metal-dielectric composites, *Phys. Rep.*, Vol. 335, 2000, pp. 275-371.
- [35]. W. J. M. Kort-Kamp, P. I. Caneda, F. S. S. Rosa, F. A. Pinheiro, Enhancing near-field heat transfer in composite media: Effects of the percolation transition, *Phys. Rev. B*, Vol. 90, 2014, 140202.
- [36]. P. Sheng, Introduction to Wave Scattering, Localization and Mesoscopic Phenomena, *Academic*, 1995.
- [37]. A. Z. Genack, J. Wang, Speckle statistics in the photon localization transition, Chapter 22, in 50 Years of Anderson Localization (E. Abrahams, Ed.), *World Scientific Publishing Co. Pte. Ltd.*, 2010.
- [38]. M. Born, E. Wolf, Principles of Optics, 7th Edition, *Pergamon*, 1980.
- [39]. W. H. Press, S. A. Teukovsky, W. T. Vetterling, B. P. Flannery, Numerical Recipes in C++, *Cambridge University Press*, Cambridge, 2002.
- [40]. G. Burlak, E. Martinez-Sanchez, The optical Anderson localization in three-dimensional percolation system, *Opt. Commun.*, Vol. 387, 2017, pp. 426-431.

Chapter 11

A Highly Directional Supercontinuum in the Visible upon Filamentation in Air

N. G. Ivanov, V. F. Losev, V. E. Prokop'ev and K. A. Sitnik

11.1. Introduction

Currently, there are many papers that present the observation and study of a supercontinuum (SC) in the visible spectrum range (white light laser) upon filamentation in air of the laser pulses with femtosecond duration [1–20]. The basis of SC formation is a non-linear optical space-time conversion of radiation of ultra-short high-intensity laser pulses in a dielectric medium with a cubic nonlinearity. The development of research in the field of SC generation in the visible spectrum range has shown the availability and great practical significance of this type of radiation for solving tasks in metrology, telecommunications, nanotechnology, optical coherence tomography, remote analysis of the atmosphere, and many others.

Filamentation and SC generation in the visible region in air was first observed by Braun et al. [1]. The term "white light filament" was introduced in the publications resulting from the European Teramobile project [2, 3]. The authors report the existence of a white light beam, observed at an altitude up to 6 km upon the propagation of a chirped femtosecond radiation pulse in the atmosphere. The initial beam had energy of 350 mJ, pulse duration of 70 fs, diameter of 50 mm. Later the same authors say that the radiation of the Teramobile laser extends over long distances in the form of a set of narrow (millimeter-diameter) light channels in the absence of ionization and, accordingly, energy loss. Similar structures were registered in [4, 5], authors explain their presence by the Bessel-like distribution of radiation intensity. In [7], radiation with a modulated spectrum was used to decrease a short wavelength cutoff. It is shown that when the distance between the spectral peaks is increased, SC has a more shortwave cutoff (up to 230 nm). The authors attributed the main mechanism of radiation conversion to four-wave parametric mixing. In [9] an overview of works devoted to the study of SC in the visible spectrum range is presented and it is shown that the main mechanism of the appearance of an SC is by self-phase modulation. In the above mentioned works, the main share of SC radiation is distributed in the form of a cone, and the radiation is distributed along the axis only in

some cases. In [10], it was shown that the filament was formed upon the long focal length ($F = 6\text{--}12\text{ m}$) focusing of radiation at $\lambda = 950\text{ nm}$ in air, and this filament was the source of directional on-axis white light (axial SC). Its appearance and direction depended on the beam power and the lens focus. At low intensities, the SC spectrum range was spread in the short-wavelength region of $450\text{--}500\text{ nm}$. With increasing intensity, the SC range expanded to $\lambda = 300\text{--}350\text{ nm}$ but became mostly conical. That is, the conditions of existence of axial SC are very limited.

The existence of a filament in most papers is explained by the balance between the self-focusing of laser radiation due to the Kerr effect and the defocusing of the radiation by the filament plasma. It is noted in [11] that the properties of the filament and the processes occurring in it largely depend on the numerical aperture of the focusing system $NA = \sin(R_0/F)$, where R_0 is the radius of the initial beam and F is the focal length. The authors introduce the concept of linear ($NA > 4 \times 10^{-3}$) and nonlinear ($NA < 3 \times 10^{-3}$) focusing modes.

In the first mode, the balance between geometric focusing and plasma defocusing is important for the existence of a filament. In the second mode the relationship between Kerr self-focusing and plasma defocusing becomes the main one. The spectral distribution of SC after the filament changes depending on the regime. With linear focusing, the SC broadening occurs only in the short-wave region of the spectrum, for nonlinear focusing, the SC broadens in both directions from the central pump wavelength. Such a division of the focusing regimes, as was shown in a number of works [1, 10, 7, 12], is conditional and not always the spectral distribution of the SC corresponds to one or another regime

Thus in [1], using a collimated beam, there was no shift of the spectrum to Stokes region, but white light was observed in SC. In [10], a supercontinuum in the visible region was observed at $NA \leq 1.5 \times 10^{-3}$, with no broadening of the emission spectrum to the Stokes region. In [12], at $NA = 1.3 \times 10^{-3}$, the SC spectrum was broadened only in the UV range up to $\lambda = 200\text{ nm}$, as well as in [7] with $NA = 2.5 \times 10^{-3}$. Thus, the available data are quite contradictory.

The influence of aberration on the process of filamentation of radiation in air was first studied in [13, 14] where stabilization of the filaments position was recorded with introduction of astigmatism into the wave front of the focused beam. Subsequently, studies of the influence of astigmatism on the filament and supercontinuum were devoted to [15-18]. In [15], it is shown that the intensity of SC in the visible spectrum range decreases with the tilt angle of the lens with a focal length of $F = 1\text{ m}$. At the same time, in [16] it was shown that the short-wave boundary of the SC does not change with an increase in the slope angle up to 10° . Observation of one or two bright light jets after the formation of a filament due to the shifting or tilting of the lens at a certain angle is presented in [17]. The filament is formed in air when the lens ($F = 10\text{ cm}$) focuses the radiation at $\lambda = 800\text{ nm}$ and a pulse duration of 50 fs . It is shown that the brightness of the light jets is also dependent on the polarization of the laser beam. The authors associate the emergence of SC radiation with a four-wave parametric process. However, the authors do not explain in any way the reasons for the appearance of 2-jet beams. In [18], it is shown

that by focusing the radiation with femtosecond pulse duration in air, the pump intensity and electron density in the filament increase with the tilt angle of the lens. As can be seen from the above articles, the conditions of appearance of the SC in the visible range by interaction of the femtosecond radiation with air differ significantly among the different publications.

The formation conditions of a highly directional supercontinuum in a visible spectrum range and its parameters are studied in this chapter.

11.2. Experimental Setup and Research Methods

In the experiments, we used a solid-state starter complex, "Start 480", developed and manufactured by the Russian company "Avesta-project", to form the radiation of femtosecond duration. The complex includes a Ti:Sa master oscillator with a continuous pump laser, optical stretcher, one regenerative and two multi-pass amplifiers with pulsed lasers pumping at a wavelength of 532 nm, and a compressor on two diffraction gratings. The output radiation parameters were as follows: a central wavelength of 940 nm, an FWHM spectral width of 26 ± 2 nm, the pulse duration of 70 ± 3 fs, energy of 50 mJ, and a beam diameter of 10 mm. The complex operated with a pulse repetition rate of 10 Hz, and the energy stability was 3 %. The output radiation was linearly polarized (horizontally), and the beam quality factor $M^2 = 2$. HR4000 spectrometer (Ocean Optics company) operating in the range of 193–1100 nm with a spectral resolution of 0.75 nm was used to measure the spectral parameters of radiation. The energy and power of the laser pump beam and the converted radiation were measured by Gentec and Ophir gauges. The duration of the radiation pulse was measured by ASF-20 femtosecond autocorrelators operating at wavelengths of 950 and 475 nm as well as by observing the interference fringes in the thin quartz plates. The spatial distribution of radiation was determined using a SP620U profilometer and the Beam Gage program, as well as by using beam patterns on photographic paper with subsequent processing in the Beam Gage program. In addition, the distribution of radiation intensity along the entire filament was recorded by removing the surface layer (photo paper, Al-mirror) method for a one or large number of pulses. Calculations of radiation propagation taking into account the third order nonlinearity of air and the aberrations were carried out in the Fresnel program. The nonlinear refractive index n_2 in simulations was assumed to be 3×10^{-19} cm²/W. Measurement of electron density in the filament plasma was carried out from the rate of change of the charge on the electrodes installed across the axis of radiation propagation. Photographing of filaments and images on the screen of visible rays was carried out by a Canon EOS 7D camera. The images obtained in the RAW format were subsequently processed in order to increase the contrast and reduce a noise.

11.3. Experimental Results and Discussion

Our experiments shown that the directional axial SC in air in case without aberration focusing is a most stable with a low numerical aperture of focusing system. This is done either through the use of a long-focus spherical mirror or by reducing the diameter of the

beam when using a reflecting telescope. With a numerical aperture $NA \geq 0.01$, only a conical SC whose directivity is extremely low is obtained. In our opinion, the observed difference is due to the fact that when using the short-length focusing, the geometric and nonlinear focuses are close to each other and a significant portion of the radiation passes through the filament in which the refraction on plasma electrons is large. In this case, after filament, the directional axial SC is virtually absent. The decrease of the numerical aperture reduces the plasma concentration in the filament and increases the field dominance length of Kerr nonlinearity over electron refraction. Under these conditions, an axial SC is possible. However, as experiments have shown, the range of conditions under which this possibility is realized is quite narrow. For example, when focusing radiation using a spherical mirror with $F = 12.2$ m, the red components of SC directed along the beam axis begin to appear at a radiation energy of 10 mJ. As the energy rises, the white components of SC arise. However, at the radiation energy of 15 mJ, the spectral components of axial white light acquire an angular dispersion that changes from pulse to pulse. The spectral composition of the SC is also extremely unstable.

Research on the SC when astigmatism is introduced in the focused beam was carried out in the optical scheme shown in Fig. 11.1.

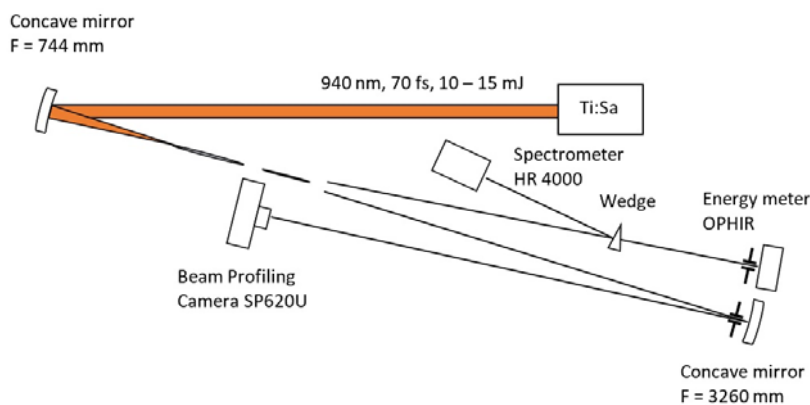


Fig. 11.1. Optical scheme of experiments.

At the minimum incidence angle of radiation on the focusing mirror, only a conical SC is observed after filament. With an increase in the incidence angle, the filament's emission intensity decreases and it splits into two parts located on the same axis. The first one is brighter and corresponds to the meridional focus region; the second part corresponds to the sagittal focus region. At a certain incidence angle, a ray of white light with a high directivity arises behind visible filament. With a further increase of the angle, the second beam appears, and then both beams disappear in reverse order. Fig. 11.2 shows a photograph of SC radiation spread behind filament at an incidence angle of the radiation on the mirror of 15° in the polarization plane of incident radiation. The figure shows that behind filament there are two directed light beams diverging relative to each other at a certain angle.

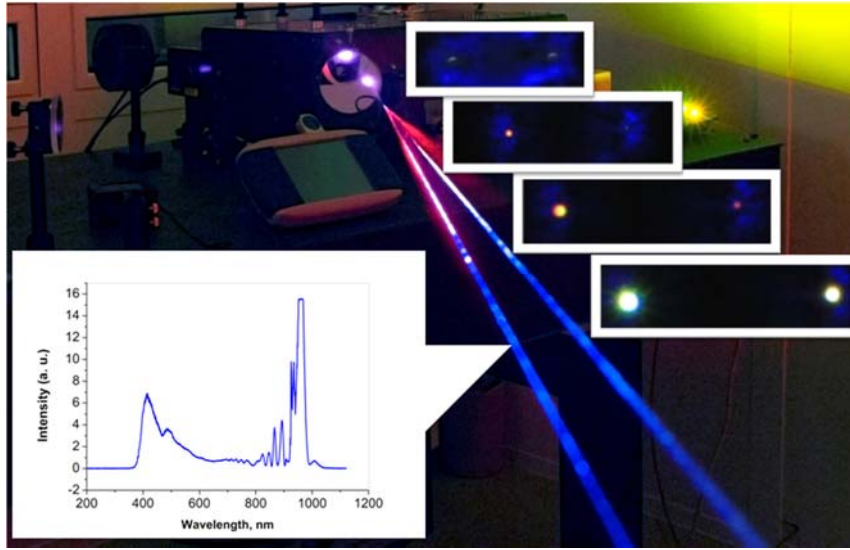


Fig. 11.2. View of two rays of SC behind the filament at an incidence angle of the laser beam on the mirror of 15° . Inset: on the right, the photographs of SC on the screen at different distances from the end of the filament (10, 20, 40, and 80 cm); on the left, the spectral composition of SC at a distance of 100 cm from the end of the filament.

As the distance from the filament increases, the colour of beams is changed from red at the beginning to blue at the end. The insets on the right show the photographs of radiation luminescence on a screen mounted at distances of 10, 20, 40, and 80 cm from the end of the filament. They also clearly show the dynamics of the change in the rays' colour. The inset on the left presents the spectral composition of the radiation at a distance of 100 cm from the end of the filament. Upon further beam transport, its spectral range did not practically change. The emergence of white light depended to a large extent on the incidence angle of radiation on the mirror and the degree of restriction of the original beam. Close by the threshold its appearance when limiting the beam on the left, the left beam appeared, and when limiting it on the right, the right beam appeared. If the threshold was exceeded when limiting the beam on the left, the right beam disappeared, and when the beam was limited on the right, the left beam disappeared. Upon deviation of the mirror axis in the vertical plane perpendicular to the plane of laser beam polarization, two bright directed beams also appeared, but the energy threshold of their appearance was increased 1.5-fold, roughly as in [8].

Fig. 11.3 shows the dynamics of the change of the spectral composition of directional SC for different distances of the wedge from the end of the visible filament (Fig. 11.1). It should be noted that a modulation in the laser radiation spectrum arises just before the filament (Fig. 11.3 (a)). A similar modulation was observed in [12], where it is explained by self-phase modulation. Just behind the filament the modulation increases substantially, and the maximum of the spectrum intensity envelope shifts to shorter wavelengths relative to the central wavelength of laser radiation (Fig. 11.3 (b, c)). Upon an increase in the distance from the filament in the SC emission, there is a growth of its short-wave part and

a shift of the maximum infrared (IR) component back to longer wavelengths (Fig. 11.3 (d)). Under certain conditions, depending primarily on the laser beam energy, this maximum becomes the Stokes shift relative to the laser radiation spectrum equal to ~ 20 nm (Fig. 11.3 (d)).

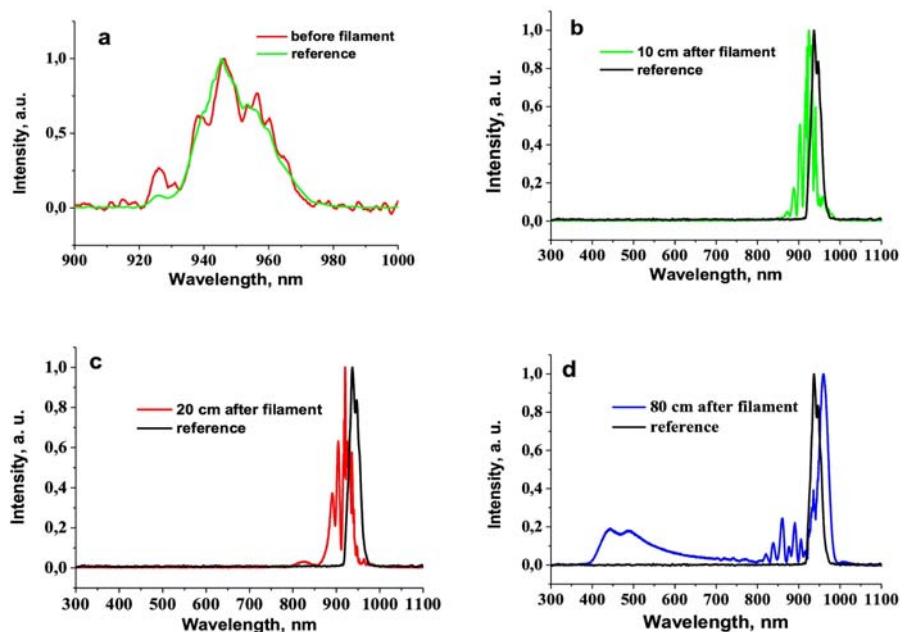


Fig. 11.3. Spectrum of reference and directional SC emission versus the distance from the visible filament.

Laser beam patterns both in and beyond the filamentation area were recorded to determine the role of astigmatism in the formation of directed SC; additionally, the calculations of aberrations for the experimental conditions were also carried out. Fig. 11.4 presents the experimental and computational results.

As can be seen from experiment (b) and calculation (d) under the condition of astigmatism in the meridional focus area (position 69 cm) the beam size along the x axis begins to increase with respect to the linear case (c). In the sagittal focus area behind the visible filament (position 75 cm) two areas with a higher intensity appear. Measurement of energy in these areas (position 75 cm) showed that they contain about 15 % of the laser beam energy. The presence of a ring structure (upper and lower half-rings) in the region of the focal waist (position 72) indicates the existence in the meridional focus area of the two high-intensity zones located above and below the filament. Their appearance is the result of two factors: astigmatism, which compresses the radiation in the horizontal planes, and the radiation refraction in the filamentation zone. Further, if a substantial fraction of radiation energy from these high intensity zones enters the rings due to the Kerr effect, two another high-intensity zones will occur in the horizontal plane at the location of their intersection. With further spreading in the sagittal focus area (75 cm), their intensity will

increase due to radiation compression in the vertical planes and refraction of radiation in the second part of filament. Thus, in contrast to the case without aberration, after filamentation, the existence of regions with increased intensity sufficient to exhibit a Kerr nonlinearity becomes possible.

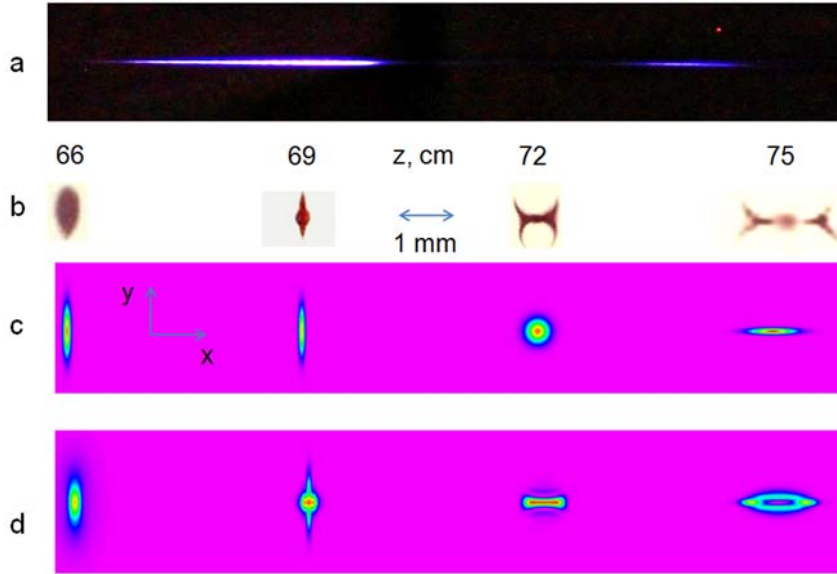


Fig. 11.4. Photograph of filament (a - the top view) and shape of the laser beam spots (b-d) for the different distances z from the mirror; (a, b) – experiment; (c, d) – calculation without (c) and with (d) taking into account the Kerr effect; (a-d) – the same scale along the z axis, angle of incidence on the mirror is 15° ; (b-d) - the same scale along the x and y axes.

Investigations of the intensity distribution in the filament by removing the surface layer technique were carried out to get a more information about the dynamics of zones occurrence with increased intensity. It was showed that the maximum radiation intensity is always in the central zone (on the filament axis). Fig. 11.5 shows the patterns of the laser beam on the Al-mirror for one pulse at different distances from the mirror.

Light area of a laser beam patterns is partly deleted Al-layer. In the central zone (dark area), the Al layer deposited on the glass substrate is completely removed. Just at the center of this region there is a filament with dimensions of $\sim 100 \mu\text{m}$. Their origin occurs before meridional focus area (68 cm). In our opinion, these are the zones from which two white beams are subsequently formed. It is seen from Fig. 11.5 that the distance between the centers of these regions is $\sim 350 \mu\text{m}$ before the meridional focus area. In the meridional focus area (69 cm) this distance decreases to $250 \mu\text{m}$, and then begins to increase and at the end of the visible filament is $850 \mu\text{m}$.

Fig. 11.6 shows the calculated dependence of the distance between these highly intensive areas as the laser radiation propagates. As can be seen from the figure, the distance varies nonlinearly, and consequently the course of two beams formed in these zones is not

rectilinear. This is due to their propagation in a medium with an inhomogeneous refractive index in the filamentation region more precisely near the filament). If we calculate the angle at which they propagate relative to each other along the (two extreme points (74 and 75 cm), then it exactly corresponds to the propagation angle (1.4°) of the two white rays in Fig. 11.2.

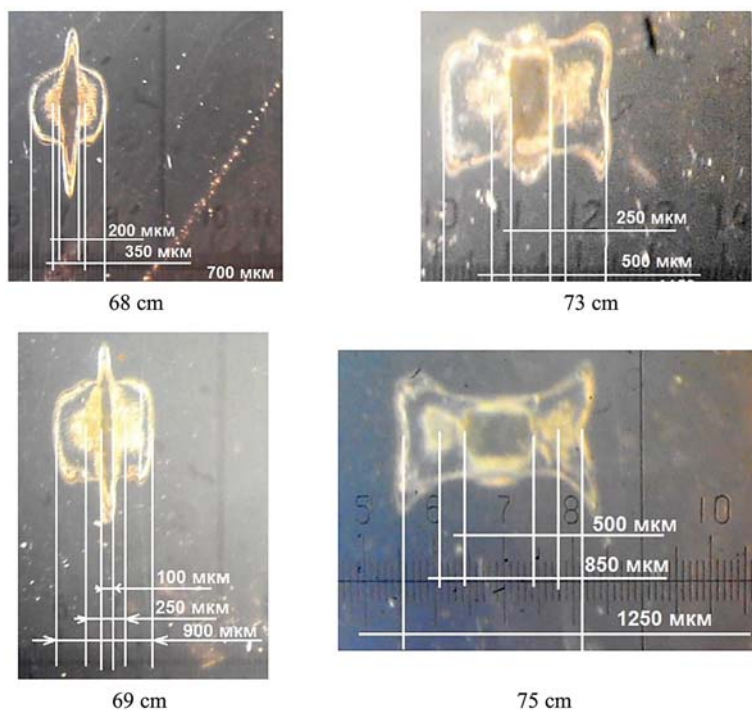


Fig. 11.5. Photographs of a laser beam patterns on an Al-mirror for one pulse at different distances from the mirror. Light area is partly deleted Al-layer. Scale of small division is 50 μm .

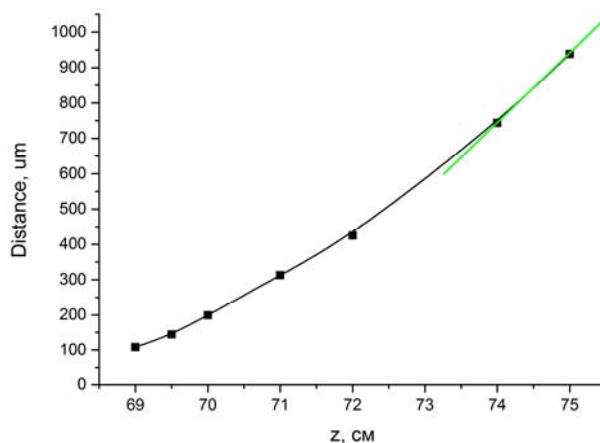


Fig. 11.6. Distance between two highly intensive areas versus distance to the mirror.

The nature of the appearance of two zones of higher intensity can be determined if we consider the dynamics of phase change (wavefront) of the laser beam. Fig. 11.7 shows the calculated phase distributions under conditions of astigmatism for different distances from the mirror. It can be seen that at position 66 cm a dip on the wave front along the center of the beam is observed, since the nonlinear addition to the refractive index is positive. Later at position 67 cm along the center of the beam, the wave front becomes convex, while on the periphery it is concave. Thus due to the Kerr effect two minima located outside the beam axis are realized in the phase distribution, which lead to the appearance of two off- axis areas with higher radiation intensity.

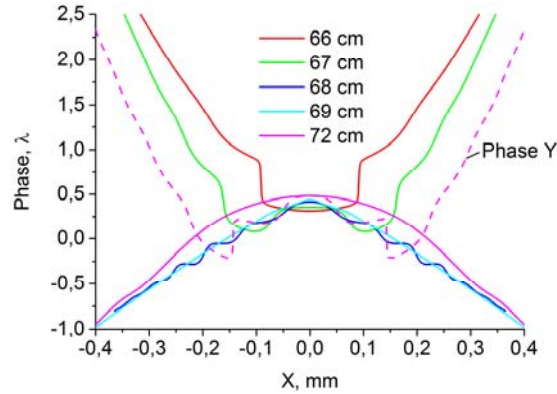


Fig. 11.7. Phase distribution over the cross section of the laser beam for different distances from mirror.

We consider that two independent beams were formed namely in these regions. As our investigations have shown, two spatial quasi-solitons appear in the zone behind the filament in conditions of astigmatism [16]. Fig. 11.8 presents experimental and calculated beam patterns and distributions of the energy density of the one beam in this zone at various distances from the filament end (75 cm from mirror).

The experimental distribution of beam energy density was built on the basis of beam patterns. In the calculations, the initial distribution of beam fluence was taken experimentally recorded fluence distribution at a distance of 15 cm from the filament. In this case the beam energy was 2.4 mJ. It should be noted that the calculation results were very critical to the energy of the laser beam, so an increase or decrease in energy by a factor of 2 led to a decrease of the quasi-soliton length.

According to our measurements, the beam does not change its diameter ($\sim 300 \mu\text{m}$) and peak intensity at a distance of $\sim 100 \text{ cm}$, although due to the linear diffraction at the end of the path it should have a size of 4 mm. Thus in this area there is a stable structure similar to a spatial soliton, where conversion of the laser radiation into the visible and ultraviolet regions of SC occurs until $\lambda = 350 \text{ nm}$. A similar stable structure has been observed previously in a number of papers, for example [1, 19, 20]. Despite the fact that the radiation intensity in calculation gradually decreases with distance increase from the

filament, nevertheless, a qualitative agreement of the theory with experiment is observed. At the same time, the calculation of the propagation of a Gaussian beam with an energy of 2.4 mJ and a diameter of 0.3 mm in e-2 shows that at a distance of 50 cm its fluence drops by a factor of 5. According to our measurements and estimates the existence of a quasisoliton is due to a balance between Kerr self-focusing and diffraction spreading.

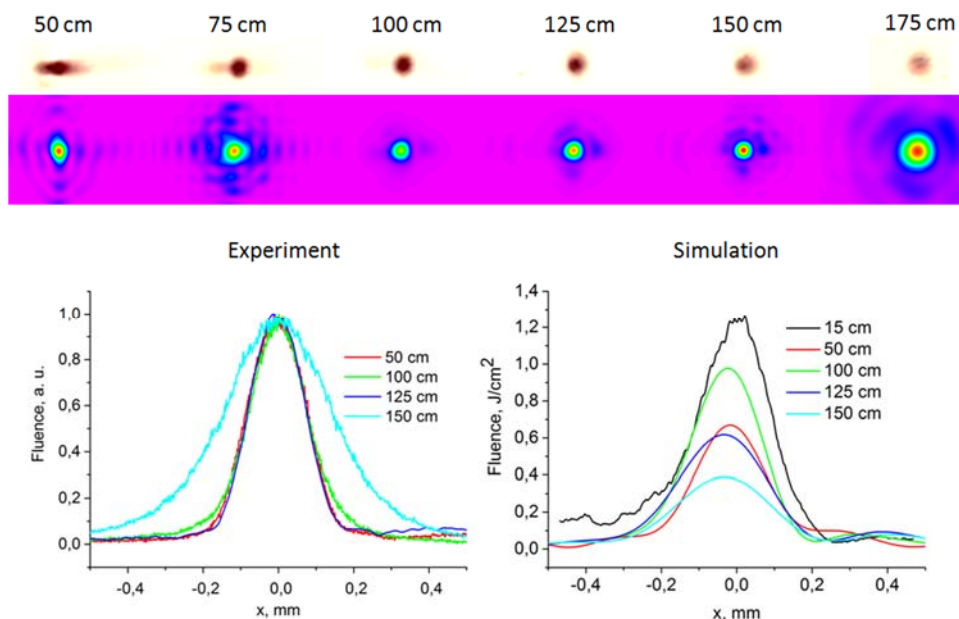


Fig. 11.8. Beam patterns and distribution of the laser beam fluence at different distances from the filament end (75 cm from mirror). Experimental (overhead row) and calculated (lower row) beam patterns have the same scale. One distribution of the laser beam fluence in simulation (black color – 15 cm) is experimental data.

Detailed studies of photographs in the field of filamentation from above and their processing showed that two thin filaments arise from the region of the meridian focus in the horizontal plane (Fig. 11.9). Their diameter is 20-30 μm and they diverge under the angle $\sim 1.3^\circ$. Their luminescence is much weaker than the axial filament. The plasma concentration in them is much lower and they exist, most likely, due to the balance of the Kerr self-focusing and diffraction spreading. These filaments also spread in the region of spatial quasi-solitons.

We also carried out measurements of electron density n_e in this area. According to the measurements and estimation, n_e is in the range of 10^{12} to 10^{13} cm^{-3} all the way from the filament end to 1.5 m away from it. It is known that change of the refractive index on the electron density is given by $\Delta n_e = -0.5(\omega_p/\omega)^2$, where ω_p is the plasma frequency and ω is the frequency of laser radiation. Under our conditions, this ratio does not exceed 4×10^{-9} , while the contribution of the Kerr nonlinearity is $\Delta n_{\text{kerr}} = \gamma I = 10^{-5}$, where $\gamma = 3 \times 10^{-19} \text{ cm}^2/\text{W}$ is the nonlinear refractive index coefficient and $I = 3 \times 10^{13} \text{ W/cm}^2$ is

the peak intensity of the laser beam. According to our estimates, this positive contribution to Δn can be compensated by the diffraction spreading of the beam with a diameter of 150 μm , which gives $\Delta n_{\text{diffr}} = (1.22\lambda)^2 / 8\pi n_0 r^2 = 10^{-5}$, where $\lambda = 940$ nm, r is the laser beam radius (FWHM), and n_0 is the refractive index coefficient of air. Thus in this case it is possible to have a balance between Kerr self-focusing and diffraction spreading.

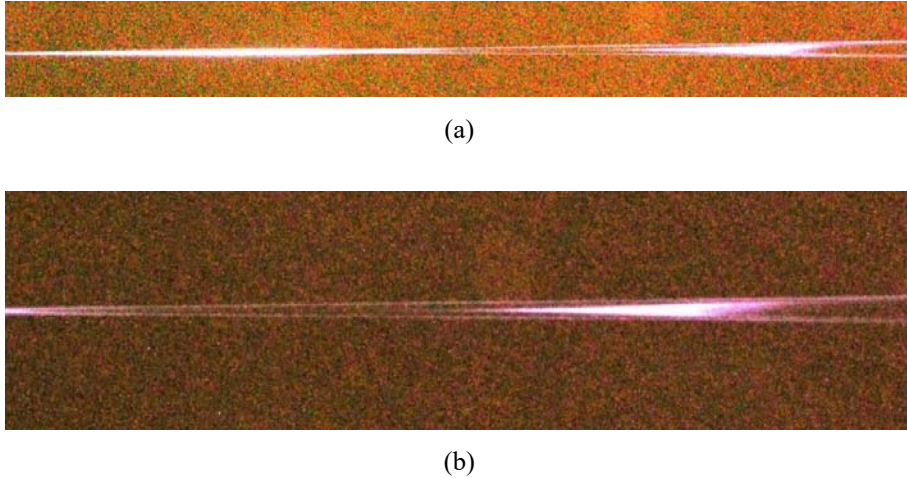


Fig. 11.9. Photo of the filamentation area (view from above) in optimal conditions to form two white rays (a), (b) is an enlarged part of this photo.

An important parameter for SC in the visible region is the conversion efficiency of the laser beam therein. Table 11.1 shows the dynamics of the change in energy for the laser beam and SC in one of the two rays. The measurements were made at a distance of 2 m from the filament. In order to cut off the infrared component with a large divergence before the power meter, the aperture was set with a diameter of 5 mm. Behind the diaphragm, the mirror of total reflection for the laser beam was set at an angle of 45°. According to Table 11.1, the conversion efficiency of radiation in the visible region of the SC is about 5 % relative to the energy of one beam measured behind the diaphragm and 0.5 % is relative to the initial energy (including two divergent beams). It should be noted that our measurements of the IR radiation energy in the range of 1–2.5 μm with the help of interference filters showed that it was not more than 5–6 μJ .

Table 11.1. Energy, pump, efficiency.

Energy before filament, mJ	Energy behind filament, mJ	Energy behind diaphragm, mJ	Energy of SC, mJ
11	3	0.66	0.033

Information about the pulse duration of the converted light component is extremely important from the point of view of understanding the formation mechanism of directional radiation of SC and from a practical point of view. Our attempts to measure the duration

of SC radiation in the range of 470 nm using an ASF-20 autocorrelator were unfortunately unsuccessful, probably due to insufficient radiation intensity. However, for the IR radiation of SC measured in the far field, the duration was set. Fig. 11.10 shows the autocorrelation function of laser radiation with $\lambda = 940$ nm (top picture) and SC radiation in the range of 900–1000 nm (bottom picture), measured at a distance of 10 m behind the filament. As can be seen from Fig. 11.10, the duration of the SC pulse is less than half that of the initial laser beam. It can be expected that the other SC spectral components in visible will have the similar pulse duration. At least our measurements of white light pulse duration according to its interference in thin plates have shown that its duration is less than 1 ps.

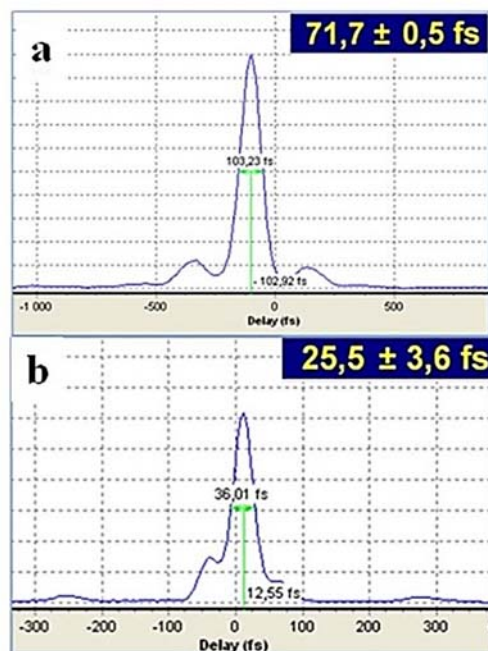


Fig. 11.10. Autocorrelation function of the laser beam with $\lambda = 940$ nm (a) and SC radiation in the range of 900–1000 nm (b).

Another important parameter of the obtained radiation, from the point of view of its application, is the beam divergence of the SC. Measurement of divergence was carried out by using a concave mirror with $F = 326$ cm installed at a distance of 7 m from the filament. Fig. 11.11 shows the intensity distribution of the SC in the visible range measured in the far field. The circle shows the dimensions of the focal spot in the diffraction-limited Gaussian beam with a diameter of 5 mm (the size of the near-field).

The circle (a) corresponds to the diameter of the diffraction-limited Gaussian beam with a central wavelength $\lambda = 600$ nm. If the size of the circle in (b) is more or less than a certain value, the size in (a) is undefined. Numerical processing of blue light ($\lambda = 415$ nm) data (b) gives a quality factor of the beam $M^2 = 1.15$. Measurement of radiation

polarization of the directed SC showed that, like the laser beam, it is linearly polarized with a polarization ratio of 95 %.

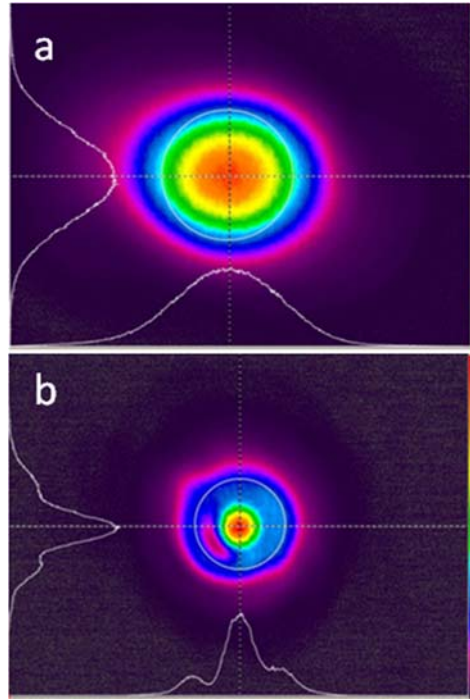


Fig. 11.11. Intensity distribution of SC radiation measured in the far field. a – total radiation, b – part of the radiation with a central wavelength of 415 nm and half-width of 60 nm. The circle is the diffraction-limited dimension for a Gaussian beam with a diameter of 5 mm.

11.4. Conclusion

Thus, experimental and theoretical studies of conditions for the formation of a highly directional SC in the visible spectral region were carried out. The laser beam having a wavelength of $\lambda = 940$ nm, pulse duration of 70 fs and energy in the range $8 \div 15$ mJ was used in these experiments. It was shown that directional axial SC in air in case without aberration focusing is possible only with a low numerical aperture of focusing system ($NA \leq 0.01$). In this case the plasma concentration in the filament reduces and the field dominance length of Kerr nonlinearity over electron refraction increases. Under these conditions, an axial SC is possible. When using the short-length focusing ($NA \geq 0.01$), the geometric and nonlinear focuses are close to each other and a significant portion of the radiation passes through the filament in which the refraction on plasma electrons is large. In this case, after filament, the directional axial SC is virtually absent.

In case aberrational focusing of radiation the aberrations were determined by varying the angle of incidence of the laser beam on a spherical mirror. The most stable pattern of highly directional SC in the visible region of the spectrum is observed at the angle of

incidence of radiation on the focusing mirror 15° . In this case two highly directional beams of light diverging relative to each other behind the filament are formed.

The study of the conditions and the causes of nucleation of two highly directional beams showed that the initial regions with higher intensity from which these beams develop appear before the meridional focus area in the initial filamentation region on the right and on the left of the filament which has a size of $\sim 100 \mu\text{m}$. When they propagate along the filament, the path is not rectilinear, the distance between the beams increases nonlinearly up to the end of the filament. After this, they are extended rectilinearly in the angle of 1.3° - 1.4° . The nature of the appearance of two zones of higher intensity is due to the distortion of the wavefront of the laser beam in conditions of the astigmatism and the Kerr effect. In result two minima located outside the beam axis are realized in the phase distribution, which lead to the appearance of two off-axis areas with higher radiation intensity and as a consequence of this the formation of two laser beams. It should be noted that all the experimental results agree well with the results of the calculations, in which only the Kerr effect was taken into account and plasma was not taken into account.

The main transformation of the emission spectrum into the short-wavelength region occurs in a stable spatial structure of length $\sim 100 \text{ cm}$ and diameter $\sim 0.3 \text{ mm}$. This region is characterized by a low concentration of plasma (10^{12} - 10^{13} cm^3) and one exists due to a balance between Kerr self-focusing and diffraction spreading of radiation. The infrared radiation component of the SC has duration 2.8 times shorter than the length of the laser beam. It is probable that the pulse duration of the visible supercontinuum has the same duration or shorter. The divergence of this radiation is close to the diffraction limit and it is linearly polarized with a polarization ratio of 95 %. Analysis of the literature and our experimental data showed that the self-phase modulation and cascading four-wave parametric processes in the area before filament, in the visible filament, and after it one play an important role in the formation of such radiation.

References

- [1]. A. Braun, G. Korn, X. Liu, D. Du, J. Squier, G. Mourou, Self-channeling of high-peak-power femtosecond laser pulses in air, *Opt. Lett.*, Vol. 20, No. 1, 1995, pp. 73-75.
- [2]. H. Wille, M. Rodriguez, J. Kasparian, D. Mondelain, J. Yu, A. Mysyrowicz, R. Sauerbrey, J. P. Wolf, L. Wöste, Teramobile: A mobile femtosecond-terawatt laser and detection system, *Eur. Phys. J. AP*, Vol. 20, 2002, pp. 183-190.
- [3]. J. Kasparian, M. Rodriguez, G. Mejean, E. Salmon, H. Wille, R. Bourayou, S. Frey, Y. B. Andre, A. Mysyrowicz, R. Sauerbrey, J. P. Wolf, L. Wöste, White-light filaments for atmospheric analysis, *Science*, Vol. 301, Issue 5629, 2003, pp. 61-64.
- [4]. J.-F. Daigle, O. N. Kosareva, N. Panov, T.-J. Wang, S. Hosseini, S. Yuan, G. Roy, S. L. Chin, Formation and evolution of intense, post-filamentation, ionization-free low divergence beams, *Optics Communications*, Vol. 284, 2011, pp. 3601-3606.
- [5]. Yu. E. Geints, A. A. Ionin, D. V. Mokrousova, L. V. Seleznev, D. V. Sinitsyn, E. S. Sunchugasheva, A. A. Zemlyanov, Post-filamentation high-intensive light channels formation upon ultrashort laser pulses self-focusing in air, *Proceedings of SPIE*, Vol. 10254, 2017, 102541F-1.

- [6]. P. Bejot, L. Bonacina, J. Extermann, M. Moret, J. P. Wolf, R. Ackermann, E. Lascoux, N. R. Salamé, E. Salmon, J. Kasparian, L. Bergé, S. Champeaux, C. Guet, 32 TW atmospheric white-light laser, *Appl. Phys. Lett.*, Vol. 90, 2007, 151106.
- [7]. F. Theberge, W. Liu, Q. Luo, S. L. Chin, Ultrabroadband continuum generated in air (down to 230 nm) using ultra short and intense laser pulses, *Appl. Phys. B*, Vol. 80, Issue 2, 2005, pp. 221-225.
- [8]. Yu. Geints, A. A. Zemlyanov, A. M. Kabanov, G. G. Matvienko, Nonlinear Femtosecond Atmospheric Optics, *Publishing House of the IAO SB RAS*, Tomsk, 2010, p. 212.
- [9]. V. P. Kandidov, S. A. Shlenov, O. G. Kosareva, Filamentation of high-power femtosecond laser radiation, *Quantum Electron.*, Vol. 39, Issue 3, 2009, pp. 205-228.
- [10]. N. G. Ivanov, V. F. Losev, V. E. Prokop'ev, Study of the population inversion mechanisms and superradiance on ions transitions of molecular nitrogen in the filament, *Proceedings of SPIE*, Vol. 9810, 2015, 98100L.
- [11]. K. Lim, M. Durand, M. Baudelet, M. Richardson, Transition from linear- to nonlinear-focusing regime in filamentation, *Scientific Reports*, Vol. 4. No. 7217, 2014, pp. 1-8.
- [12]. N. Aközbek, S. A. Trushin, A. Baltuška, et al., Extending the supercontinuum spectrum down to 200 nm with few-cycle pulses, *New Journal of Physics*, Vol. 8, 2006, p. 177.
- [13]. G. Fibich, S. Eisenmann, B. Ilan, A. Zigler, Control of multiple filamentation in air, *Optics Letters*, Vol. 29, No. 15, 2004, pp. 1772-1774.
- [14]. G. Méchain, A. Couairon, M. Franco, B. Prade, A. Mysyrowicz, Organizing multiple femtosecond filaments in air, *Physical Review Letters*, Vol. 93, No. 3, 2004, pp. 035003 (1-4).
- [15]. Y. Kamali, Q. Sun, J. F. Daigle, A. Azarm, J. Bernhardt, S. L. Chin, Lens tilting effect on filamentation and filament-induced fluorescence, *Opt. Commun.*, Vol. 282, 2009, pp. 950-954.
- [16]. S. Sreeja, T. Shuvan Prashant, Ch. Leela, V. Rakesh Kumar, S. P. Tewari, S. Venugopal Rao, P. Prem Kiran, Effect of lens tilt on SCE and filamentation characteristics of femtosecond pulses in air, *Proceedings of SPIE*, Vol. 8434, 2012, 84340T.
- [17]. Z. Xu, X. Zhu, Y. Yu, N. Zhang, J. Zhao, Super-luminescent jet light generated by femtosecond laser pulses, *Sci. Rep.*, Vol. 4, Issue 3892, 2014, pp. 1-7.
- [18]. A. A. Dergachev, A. A. Ionin, V. P. Kandidov, D. V. Vokrousova, L. V. Seleznev, D. V. Sinitsyn, E. S. Sunchugasheva, S. A. Shlenov, A. P. Shustikova, Plasma channels during filamentation of a femtosecond laser pulse with wavefront astigmatism in air, *Quantum Electron.*, Vol. 44, No. 12, 2014, pp. 1085-1090.
- [19]. G. Merchain, C. D. Amico, Y. B. Andrer, S. Tzortzakis, M. Franco, B. Prade, A. Mysyrowicz, A. Couairon, E. Salmon, E. R. Sauerbrey, Range of plasma filaments created in air by a multi-terawatt femtosecond laser, *Optics Commun.*, Vol. 247, 2005, pp. 171-180.
- [20]. J. F. Daigle, O. N. Kosareva, N. Panov, T.-J. Wang, S. Hosseini, S. Yuan, G. Roy, S. L. Chin, Formation and evolution of intense, post-filamentation, ionization-free low divergence beams, *Opt. Commun.*, Vol. 284, 2011, pp. 3601-3606.

Chapter 12

Study of Third-Harmonic Generation at Interfaces Taking into Account the Contribution of Self-Focusing Effect

**Emerson Cristiano Barbano, Sergio Carlos Zilio
and Lino Misoguti**

12.1. Introduction

Self-focusing (SF) and third-harmonic generation (THG) are two fundamental third-order nonlinear effects that have great importance in nonlinear optics. Besides the same nonlinear order, they are very unlike in terms of behavior and origin, for instance, while THG has origin from pure instantaneous electronic nonlinearity, SF may have several contributions, including pure electronic and other non-instantaneous nonlinearities such as orientational effect. In this way, most of the time these two effects are studied separately in order to simplify the analysis. Here, we present a study of THG at interfaces under the influence of the SF effect.

THG at interfaces [1] has long been of interest for third-order nonlinear materials characterization [2-4], ultrashort pulse measurements [5, 6], and nonlinear microscopy [7, 8]. Any of these applications requires the precise knowledge of the fundamental processes involved for a correct interpretation of the results. It is well known that for a thick sample (thickness much larger than the Rayleigh range of the laser beam), the THG vanishes in bulk media because the Gouy phase shift [9] leads to the destructive interference between third harmonics (TH) generated before and after the focus [3, 10, 11]. The THG occurs only at interfaces because the focal symmetry is broken by the presence of two materials with different refractive indices and/or third-order susceptibilities. As a result, TH at interfaces can be used to measure the discontinuity in the fast electronic part of the cubic susceptibility, $\Delta\chi^{(3)}(-3\omega, \omega, \omega, \omega)$. However, there are cases where such a nonlinear process is not acting alone during the light-matter

interaction, leading to difficulties in the reliable interpretation of the expected signal behavior.

We have studied the THG of femtosecond (fs) pulses in glass slab [12] (where the sample was scanned along the beam propagation direction) and we observed the TH generated at the entrance and exit interfaces are unequal. There are previous works where the asymmetry of THG peaks was observed and a few explanations were proposed. For example, we may quote the constructive interference due to the Fresnel reflections [12], the overcompensation of group velocity dispersion [13] at output interfaces, and even some theoretical studies where the non-negligible third-order susceptibility of air may cause asymmetry between front and back interface peaks [14]. Therefore, it seems worthwhile to investigate more carefully the contributions of linear and nonlinear refractive indices at the interfaces of different materials and the main goal of the research presented here is to study this asymmetry of THG peaks.

We report on THG measurements carried out with a setup similar to the Z-scan technique [15, 16] at the tightly focused laser beam condition. We performed THG measurements at air-glass interfaces of several optical glasses, as well as at glass-liquid interfaces, where we have used a cuvette filled with selected organic solvents. In the first case, by carefully controlling the sample position and laser power, we could follow the evolution of the THG intensity at both interfaces (entrance and exit) of different glasses. We have found that the TH generated at the entrance interface follows the expected cubic power dependence, which is characteristic of a third-order process, but the THG at the exit presents a slope higher than three, indicating that another nonlinear effect is also present. We propose that SF strongly contributes to the THG asymmetry because the positive nonlinear refraction decreases the effective beam waist as the pulse propagates, resulting in higher laser irradiance at the exit interface. This unbalances the TH generated at the entrance and exit interfaces. In the case of a cuvette filled with different solvents, we extended our study to a different configuration. Such a system presents four interfaces separating three materials of different refractive indices and third-order nonlinear susceptibilities. Using different cuvettes and solvents we have a variety of interfaces with different linear and nonlinear optical properties. Although the THG is due to the electronic part of the nonlinear susceptibility, it can suffer from the influence of the SF effect, which can have both instantaneous electronic and slow nuclear contributions. This mixture of two distinct third-order nonlinear processes was never considered for such interfaces. Here, using the cuvette walls as a reference and taking into account the SF effect, we shall show how to determine the nonlinear refractive indices, n_2 , and third-order nonlinear susceptibilities of the solvents, $\chi^{(3)}$, by the THG signals.

We also proposed a study of the SF effect on the THG by changing the pulse duration of the pump beam. Due to the non-instantaneous nature of the orientational process, the SF from the solvent may be influenced by the pulse duration. In this case, the THG, which is an instantaneous electronic phenomenon, can be indirectly affected by pulse duration by means of the SF effect from non-instantaneous nonlinearity.

12.2. Basic Theory of Nonlinear Polarization and Nonlinear Refractive Index

There are several theoretical details about THG (especially for tightly focused beam) and SF effects which we are not going to present here, but the reader can find those in Refs. [10, 17, 18]. In this text, we just mentioned one simple usual model to introduce the nonlinearity related to harmonic generation and nonlinear refractive index. For sake of simplicity, let us consider the simple case where the electric field is monochromatic and can be represented as plane wave:

$$E(t) = E_0 \cos(\omega t), \quad (12.1)$$

where E_0 is the amplitude of the wave, ω is the wave's angular frequency and t is a given point in time. Considering the nonlinear interaction between light and matter up to the third order, we obtain the following expression for the induced polarization P :

$$P = \varepsilon_0 \chi^{(1)} E_0 \cos(\omega t) + \varepsilon_0 \chi^{(2)} E_0^2 \cos^2(\omega t) + \varepsilon_0 \chi^{(3)} E_0^3 \cos^3(\omega t), \quad (12.2)$$

where ε_0 is the permittivity of free space, $\chi^{(1)}$ is the linear electric susceptibility, and $\chi^{(2)}$ and $\chi^{(3)}$ are known as second- and third-order nonlinear susceptibilities. It is possible to show by trigonometric identities that:

$$\begin{aligned} P = & \varepsilon_0 \chi^{(1)} E_0 \cos(\omega t) + \frac{1}{2} \varepsilon_0 \chi^{(2)} E_0^2 [1 + \cos(2\omega t)] \\ & + \varepsilon_0 \chi^{(3)} E_0^3 \left[\frac{3}{4} \cos(\omega t) + \frac{1}{4} \cos(3\omega t) \right]. \end{aligned} \quad (12.3)$$

One can see from Eq. (12.3) the rise of a host of optical phenomena. The first term is the linear response of the material by the presence of light (the material's response to the applied optical field is linear with the strength of the applied optical field). The second term presents the second-order polarization $P^{(2)}$ where we have two contributions, a constant one and another one at the frequency 2ω . The first contribution (at zero frequency) is a process known as optical rectification, which is a static electric field created across the nonlinear medium, and the second one (at 2ω) is related to the second-harmonic generation (SHG) process. It is important to point out that we are considering a simple case in which the applied electric field is monochromatic. For the general case in which the field is made up of several frequency components, the second-order polarization is related to even more processes such as sum-frequency generation (SFG) and difference-frequency generation (DFG) [10]. Here, the monochromatic approach is interesting since the expression for the third-order contribution to the nonlinear polarization (last term in Eq. (12.3)) is very complicated for the general case. The third-order contribution in Eq. (12.3) describes a nonlinear response to the polarization at the frequency of the input electric field and another one at frequency 3ω that leads to the THG process.

Equation (12.3) shows the second-order effects depend on $\chi^{(2)}$ while the third-order effects depend on $\chi^{(3)}$. For centrosymmetric materials, the response of the induced polarization to

an external electric field must respect the relation $P(-E) = -P(E)$, that is, when the electric field is reversed, the polarization remains its magnitude but is opposite to the field [19]. Consequently, for centrosymmetric materials, all nonlinear susceptibilities of even powers of E , $\chi^{(2)}$, $\chi^{(4)}$, etc., have to be equal to zero and the nonlinear effects related to $\chi^{(3)}$ are the most important. Therefore, in this case, the induced polarization can be written as:

$$P = \varepsilon_0 \chi^{(1)} E_0 \cos(\omega t) + \varepsilon_0 \chi^{(3)} E_0^3 \left[\frac{3}{4} \cos(\omega t) + \frac{1}{4} \cos(3\omega t) \right]. \quad (12.4)$$

From Eq. (12.4) we see the part of the polarization which influences the propagation of the beam at frequency ω is

$$P(\omega) = \varepsilon_0 \chi^{(1)} E_0 \cos(\omega t) + \frac{3}{4} \varepsilon_0 \chi^{(3)} E_0^3 \cos(\omega t), \quad (12.5a)$$

$$P(\omega) = \varepsilon_0 \left[\chi^{(1)} + \frac{3}{4} \chi^{(3)} E_0^2 \right] E(t). \quad (12.5b)$$

Taking into account the wave equation for a beam that propagates at z direction and using $c^2 = 1/\mu_0 \varepsilon_0$ (where μ_0 is the permeability of free space), we obtain the following ratio between the speed of light in vacuum, c , and in the medium, v :

$$\left(\frac{c}{v} \right)^2 = n^2 = 1 + \chi^{(1)} + \frac{3}{4} \chi^{(3)} E_0^2, \quad (12.6)$$

where n is the refractive index. Therefore, considering the linear refractive index, $n_0 = (1 + \chi^{(1)})^{1/2}$, we have

$$n = n_0 \left[1 + \frac{3\chi^{(3)} E_0^2}{4(n_0^2)} \right]^{1/2} \approx n_0 \left[1 + \frac{3\chi^{(3)} E_0^2}{8(n_0^2)} \right], \quad (12.7)$$

where the approximation comes from the Taylor series expansion considering that the second term is much smaller than 1.

Using the irradiance as:

$$I = \frac{1}{2} c n_0 \varepsilon_0 E_0^2. \quad (12.8)$$

The effective refractive index as a function of I is:

$$n = n_0 + \frac{3}{4} \frac{\chi^{(3)}}{c n_0^2 \varepsilon_0} I. \quad (12.9)$$

An alternative way (and more commonly used) of defining the irradiance-dependent refractive index is:

$$n = n_0 + n_2 I, \quad (12.10)$$

where n_2 is the material nonlinear refractive index. Then, in the SI system of units:

$$n_2 = \frac{3}{4} \frac{\chi^{(3)}}{cn_0^2 \epsilon_0}. \quad (12.11)$$

One can see from Eq. (12.10) that the refractive index depends on the irradiance of the light. Usually, n_2 values are relatively small, typical values are of the order of $10^{-20} \text{ m}^2/\text{W}$ for common optical glasses.

SF effect is a consequence of the nonlinear refractive index. Briefly, it can be thought as an induced lens due to the nonlinear interaction between light and matter. Considering a medium whose refractive index is given by Eq. (12.10), with positive n_2 , and a laser beam with Gaussian transverse intensity profile, $I = I(r)$, propagating into this medium, we have:

$$n(r) = n_0 + n_2 I(r). \quad (12.12)$$

In this case, the laser beam induces a refractive index variation in such a manner that the beam is caused to come to a focus [10, 20]. The effective refractive index is larger at the center of the beam than at its borders, the beam is slowed at the center in comparison to the edge and, the material acts as a positive lens. For negative n_2 , the material presents a reduced refractive index on the beam axis and it acts as if it were a negative lens, self-defocusing.

There are different physical mechanisms that can produce an irradiance-dependent refractive index such as electronic polarization, electrostriction and thermal effects [17]. For short-pulse laser beam focused in solids and liquids, one of the most important mechanisms is the electronic polarization, also known as optical Kerr lens effect, and occurs due to electronic cloud distortion. But SF has been observed for different types of materials and with many types of laser sources [18]. For instance, SF of mode-locked laser pulses in liquid argon [21] and molecular gases [22], pulsed dye lasers in gases due to resonant interactions [23] and, also, SF of CW (continuous-wave) dye laser in sodium vapor has been reported [24].

The SF effect is a self-action phenomenon because the beam modifies its own propagation by means of the nonlinear response of the medium. As the beam propagates in the medium, the wavefront of the beam gets distorted and this distortion is similar to that imposed by a positive lens. SF is often responsible for optical damage of transparent materials. Since the propagation is perpendicular to the wavefront, the laser beam appears to focus by itself and the beam waist radius is decreased compared with that of a weak laser pulse [25]. More information about SF effect can be found, for example, in the books of Boyd [10] and Koechner [17].

12.3. THG under Influence of SF Effect

In the previous section we discussed the THG and the SF effect separately. Now, we propose a simple model for analyzing the THG under the influence of the SF. In section 12.2, to derive the nonlinear refractive index, we considered, in the wave equation for the light-matter interaction, the part of the induced polarization at frequency ω in Eq. (12.4),

on the other hand, considering only the part at 3ω , one can see that this term of the nonlinear polarization acts as a source term for the THG process. In this case, solving the wave equation for the induced polarization at 3ω and an electric field whose amplitude varies spatially according to a Gaussian distribution, it is possible to show that in the tightly focused beam condition [26]:

$$I_{sample}^{(3\omega)} \propto |\Delta\chi^{(3)}|^2 (I^{(\omega)})^3, \quad (12.13)$$

where $I_{sample}^{(3\omega)}$ and $I^{(\omega)}$ are the irradiances of the TH and fundamental beam, respectively. $\Delta\chi^{(3)}$ is the difference between the third-order nonlinear susceptibility of the two materials at the interface. For the case of a single glass slab, Eq. (12.13) can be rewritten as:

$$I_{sample}^{(3\omega)} \propto (\chi_{sample}^{(3)})^2 (I^{(\omega)})^3, \quad (12.14)$$

since the nonlinear susceptibility of air is negligible. As mentioned earlier, the THG at the entrance and exit interfaces are different and we propose the SF effect explains this asymmetry. In this case, the SF changes the beam waist, w , and the irradiance becomes a function of w ,

$$I_{sample}^{(3\omega)} \propto (\chi_{sample}^{(3)})^2 [I^{(\omega)}(w)]^3. \quad (12.15)$$

Considering the beam waist radius changes with the laser power explains why the slope of the THG signal does not follow the cubic power law. In this simple model, we consider that the beam waist for a given laser power, P , depends on the sample's nonlinearity because of the SF process. In the case of a positive nonlinearity, we assume that the beam radius at the exit surface is reduced to:

$$w_{output} = w(P) = w_0(1 - BP), \quad (12.16)$$

where w_0 is the radius without SF, B is an empirical constant proportional to the sample's nonlinearity, n_2 , and the product BP represents the amount of beam decrease. Since the laser irradiance is inversely proportional to the square of the beam radius:

$$I \propto P/w^2 = P/[w(P)]^2, \quad (12.17)$$

we propose the following empirical equation:

$$I_{sample}^{(3\omega)} = A' \left(\frac{P}{[w_0(1-BP)]^2} \right)^3, \quad (12.18)$$

where A' is a constant that depends on the particular sample. For our purpose here we are going to define $A = A'/w_0^6$ and, in this way,

$$I_{sample}^{(3\omega)} = A \left[\frac{P}{(1-BP)^2} \right]^3. \quad (12.19)$$

The quantity BP changes the slope of the logarithmic THG intensity versus the input power curve, while A adjusts its offset.

The SF of light occurs when a laser beam with a Gaussian transverse intensity profile propagates through a medium whose refractive index is given by Eq. (12.12). For a positive nonlinearity, as is usual for the instantaneous electronic polarization of optical glasses, the refractive index is larger at the center of the laser beam than at its periphery, resulting in a positive lens whose focal length depends on the laser irradiance and sample nonlinearity [27]. In the Z-scan method, the interfaces translate through the laser focal plane. If the beam is tightly focused at the exit interface, Fig. 12.1 (a), the sample is in the pre-focus region, and the nonlinear interaction occurs just in this region. On the other hand, when the beam is focused at the entrance interface, Fig. 12.1 (b), only the post-focus region is important. It is worth mentioning that, in both cases, THG and SF effects occur in a typical Rayleigh range, z_R , where the laser irradiance is strong enough to induce such nonlinear effects.

As seen in Fig. 12.1 (a), stronger THG is expected due to the reduction of the beam waist at the exit interface. Furthermore, the THG at the entrance interface is not enhanced because the SF changes the beam waist just after the focus where the irradiance is smaller due to beam expansion.

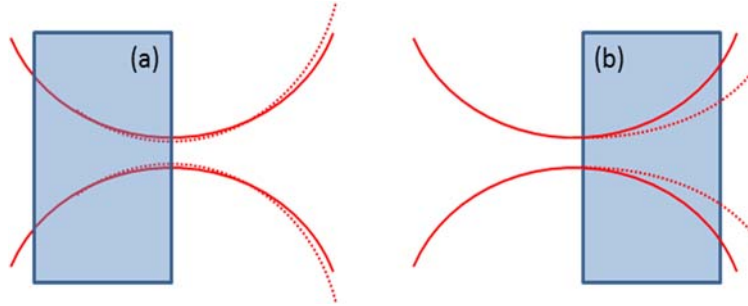


Fig. 12.1. Schematic SF action represented by the dashed line for beam focused at the exit (a), and entrance (b) interfaces. Source: [28].

As we have already mentioned, the THG, in the tightly focused laser beam condition (thick sample), is observed when the interface between two different materials, which we can label 1 and 2, is within the laser confocal parameter, and the THG intensity is proportional to the difference between the third-order nonlinear susceptibility of the two materials, Eq. (12.13). In a more general approach we have to consider the phase-matching, J , integrals, since the sample dispersion may significantly affect the harmonic generation efficiency as well. If medium 2 is air, which has a very weak nonlinearity, we just need the nonlinear susceptibility of medium 1, $\chi_1^{(3)}$ (glass, for example) to obtain:

$$I_{medium\ 1/air}^{(3\omega)} \propto \left| J(b, \Delta k_1) \chi_1^{(3)} \right|^2 (I^{(\omega)})^3. \quad (12.20)$$

On the other hand, for two media with non-negligible nonlinearities, which are the cases of the internal interfaces of a cuvette with solvent:

$$I_{cuvette/solvent}^{(3\omega)} \propto \left| J(b, \Delta k_1) \chi_1^{(3)} - J(b, \Delta k_2) \chi_2^{(3)} \right|^2 (I^{(\omega)})^3, \quad (12.21)$$

where $J(b, \Delta k)$ is the phase-matching integral defined as,

$$J(b, \Delta k) = \int_0^\infty \frac{e^{i\Delta k z}}{(1+2iz/b)^2} dz, \quad (12.22)$$

b is the confocal parameter ($2z_R$), and Δk is the wave vector mismatch,

$$\Delta k = 3k_1 - k_3 = \frac{6\pi}{\lambda} (n_1 - n_3), \quad (12.23)$$

where n_1 and n_3 are the refractive indices at the fundamental and the TH wavelengths, respectively.

It is possible to determine the $\chi^{(3)}$ ratio between two media (1 = cuvette wall and 2 = solvent) using Eqs. (12.20) and (12.21), according to:

$$\frac{\chi_2^{(3)}}{\chi_1^{(3)}} = \frac{J(b, \Delta k_1)}{J(b, \Delta k_2)} \left[1 \pm \left(\frac{I_{cuvette/solvent}^{(3\omega)}}{I_{cuvette/air}^{(3\omega)}} \right)^{1/2} \right]. \quad (12.24)$$

In this case, one needs to know the values of J and use the correct signal, which is usually positive when the nonlinearity of the solvent is higher than that of the cuvette [29]. J can be numerically calculated with Eq. (12.22). For a practical evaluation of the nonlinear susceptibility ratio, it is possible to graphically obtain an empirical equation that fits J as a function of $|b\Delta k|$ [29]. In this case we propose the following Sellmeier's type of dispersion equation:

$$J(x)/b = \left(C_1 + \frac{C_2(x+C_3)^2}{(x+C_3)^2 - C_4} + C_5(x+C_3)^2 \right)^{1/2}, \quad (12.25)$$

where $x = |b\Delta k|$, and C_1, C_2, C_3, C_4 and C_5 are the constants to be determined by the fitting procedure. This expression can be justified by its similarity to the refractive index dispersion.

12.4. THG at Optical Glasses Interfaces

In experiments where the sample is scanned along the beam propagation direction, the TH generated at the entrance and exit interfaces are different [12]. The TH generated at the exit interface is systematically stronger than one generated at the entrance. It seems worthwhile to study more carefully this phenomenon since THG is an important third-order process used in a variety of applications. In the present section we report on THG measurements, using femtosecond laser pulses, at the interfaces of several optical glasses

carried out in a similar way to that of the Z-scan technique [15, 16] at the tight focusing limit. By controlling the laser power, we could follow the evolution of the THG intensity at both interfaces for different materials. We found the TH generated at the entrance interface follows the expected cubic power law dependence while the TH generated at the exit presents a slope higher than three. We next show how this behavior can be explained considering the SF effect.

12.4.1. Experiments

In order to verify that SF effect affects the THG at interfaces, we carried out THG Z-scan measurements using a commercial optical parametric amplifier (OPA) (TOPAS, from Light Conversion), synchronously pumped by a chirped-pulse-amplified system (CPA 2001-Clark MXR) with a 775 nm central wavelength and 150 fs pulse duration at 1 kHz repetition rate. The OPA was set to generate pulses at the near IR range ($\lambda = 1300$ nm) with about 120 fs of pulse duration (TH centered at 433 nm). In these pump and harmonic wavelengths, the samples used here (silica, BK7, SK11, and F2) present no significant linear absorption. As the Z-scan method requires, the sample translates through the focus of the laser beam located at $z = 0$. Negative values of z correspond to sample positions between the focusing lens and its focal plane such that the exit (entrance) interface appears at the negative (positive) z direction. The average laser power was carefully adjusted with a calcite polarizer in the range from 0.3 to 0.8 mW ($I_0 \sim 2.4$ to 6.4×10^{16} W/m²) and was then focused with a 50 mm focal length lens ($w_0 \approx 8$ μ m) such that the Rayleigh range ($z_R = 151$ μ m) was shorter than the sample thickness, allowing us to resolve the two THG signal peaks at the two interfaces. To avoid damage to the detector, we used dichroic mirrors (high reflectance to TH beam) to separate the strong fundamental beam from the TH beam. A lens placed after the sample was used to direct most of the TH light into a large-area silicon PIN detector coupled to a lock-in amplifier, Fig. 12.2.

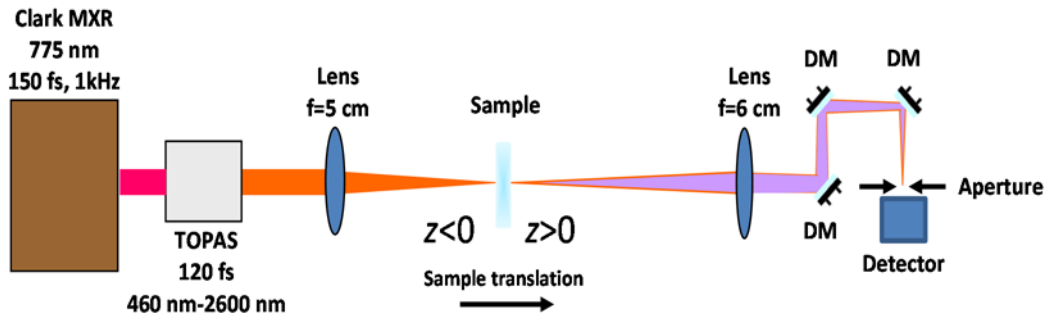


Fig. 12.2. Experimental setup for measuring the TH signal. DM, dichroic mirrors.
Source: Adapted from [30].

12.4.2. Results and Discussion

Here, we have measured three optical glasses, BK7, SK11 and F2. In Table 12.1 we present their thickness and linear refractive index at 433 nm and 1300 nm. We have used silica as a reference because it has a well-known nonlinearity, $n_2 = 2.71 \times 10^{-20} \text{ m}^2/\text{W}$ [31].

Table 12.1. Optical glasses used with corresponding refractive index and thickness.

Sample	n_0 (433 nm)	n_0 (1300 nm)	thickness (mm)
Silica	1.46693	1.44692	1.225
BK7	1.52697	1.50370	0.925
SK11	1.57564	1.54965	1.155
F2	1.64271	1.59813	1.165

Fig. 12.3 shows the THG curves as a function of z -position. As it can be seen, the THG peaks generated at the exit are systematically more intense than those generated at the entrance interface for all samples measured. It is worth mentioning that although the Fig. 12.3 shows THG curves for a fixed power (0.44 mW), this behavior of stronger peak at the exit interface (negative z) and a smaller one at the entrance (positive z) is observed for all laser powers used.

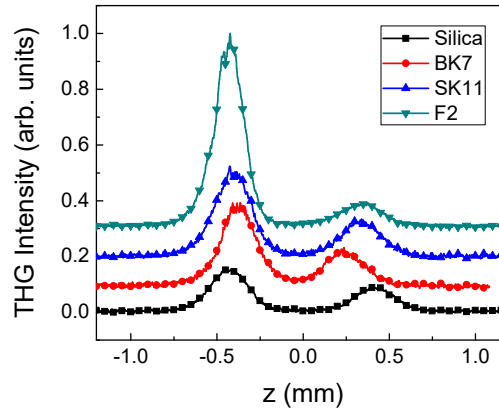


Fig. 12.3. THG Z-scan signal for silica, BK7, SK11, and F2 obtained at an average laser power of ~ 0.44 mW. An offset of 0.1 between consecutive curves was introduced for better visualization. Source: [28].

It also possible to analyze the THG peak widths. Fig. 12.4 shows the normalized TH for the entrance and exit interface peaks presented in Fig. 12.3. One can see clearly in Fig. 12.4 that the harmonics generated at entrance and exit interfaces exhibit different widths. The TH generated at the exit interfaces are systematically narrower than those generated at the entrance interfaces. For silica (which presents the lowest nonlinearity) the width variation is more difficult to notice, however, with a higher power, 0.8 mW ($I_0 \approx 6.4 \times 10^{16} \text{ W/m}^2$), this difference becomes clear, Fig. 12.5.

This reduction in the width indicates a change in the Rayleigh range (or Rayleigh parameter) of the Gaussian beam because it corresponds to the region where the laser beam is strong enough to generate TH. As predicted by the Gaussian beam theory, the reduction in the Rayleigh range may be related to a beam waist radius reduction [17].

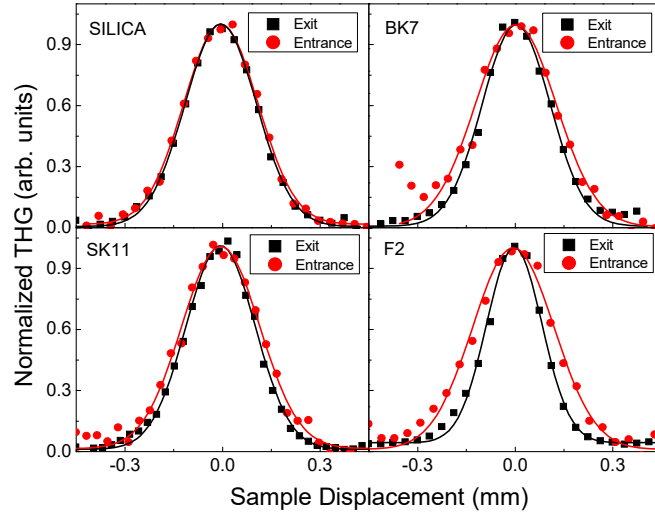


Fig. 12.4. Normalized TH peaks generated at the entrance (red) and exit (black) interfaces of optical glasses obtained with 0.44 mW pump power. The points are the experimental results, and the lines are fittings with Gaussian functions. Source: [26].

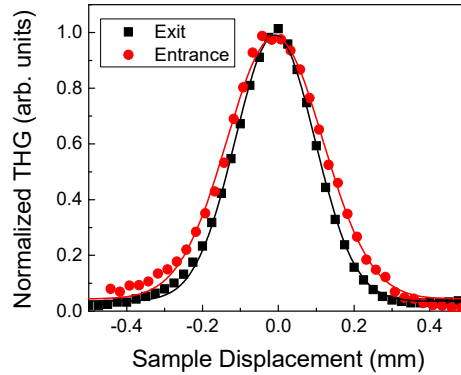


Fig. 12.5. Normalized TH peaks generated at the entrance (red) and exit (black) interfaces for silica obtained with 0.8 mW pump power. The points are the experimental results, and the lines are fittings with Gaussian functions. Source: [28].

Measurements as a function of the laser power reveal similar TH curve shapes, and considering only the peak values at the interfaces, we could establish the power law dependence through the log-log plots, Fig. 12.6. The slopes of the THG for entrance interfaces (red circles) are exactly 3 while the slopes for exit interfaces (black squares) are

slightly higher than 3. The slopes in Fig. 12.6 were obtained by considering a simple linear fit in the log-log scale. As expected, F2 presents the uppermost slope because it has the highest nonlinear refractive index among the glasses studied here. The solid lines correspond to the fitting using Eq. (12.19), which takes into account the beam waist decrease. The fitting of the experimental data with Eq. (12.19) allows finding the sample nonlinearity using the value of silica as a reference. By comparing the relative B values of each fitting, one can estimate n_2 , once the ratio $B_{\text{glass}}/B_{\text{silica}}$ should be equal to $n_{2\text{glass}}/n_{2\text{silica}}$. Table 12.2 shows the results achieved by this procedure (THG slope).

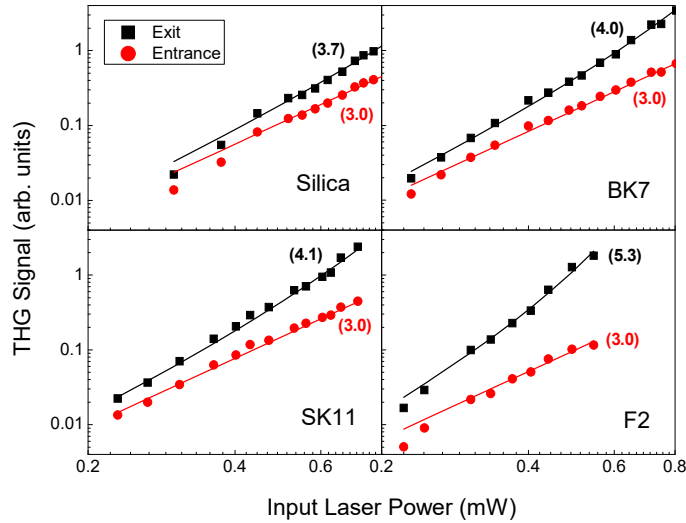


Fig. 12.6. THG peak values at the entrance (red) and exit (black) interfaces as a function of the incident laser power for silica, BK7, SK11, and F2. The symbols are the experimental data, while the solid lines are results obtained with the model proposed. The numbers in parenthesis indicate the slopes (THG dependence with laser power). Source: [28].

There is a trend of broadening (entrance) and shortening (exit) of the THG peaks as a function of the laser power, Fig. 12.7, which can also be explained by the SF effect as the TH is generated near the focal plane. When the nonlinear sample is placed after the focus position (entrance interface), the SF effect tends to recollimate the beam, and z_R increases. On the other hand, for a sample positioned before the focal plane (exit interface), z_R decreases. The broadening of z_R at the entrance interface does not affect the fundamental laser irradiance, since most of the TH is generated near the interface, where the SF did not take place yet because of the lack of beam propagation. This explains why the TH generated at the entrance interface follows the cubic power law. In the case of the exit interfaces, most of the THG occurs in the vicinity of the output face where the beam waist compression already took place due to SF and propagation along the Rayleigh distance. Due to this reduction of the beam waist radius, the laser irradiance increases and leads to a THG law with an exponent greater than 3. Such confocal parameter behavior can be understood considering the z_R dependence with w ,

$$2z_R(P) = \frac{2\pi n}{\lambda} [w(P)]^2, \quad (12.26)$$

or, using Eq. (12.16), we have:

$$2z_R(P) = 2z_{R0}(1 \pm BP)^2, \quad (12.27)$$

where z_{R0} corresponds to the Rayleigh range in the absence of SF. The positive (negative) signal is relative to the entrance (exit) interface. In Fig. 12.7, the symbols are the experimental data, while the solid lines are obtained from Eq. (12.27). With the B values obtained by these fittings we also calculated the $n_{2,glass}/n_{2,silica}$ ratio. Table 12.2 shows the results achieved by this procedure ($2z_R$ approach).

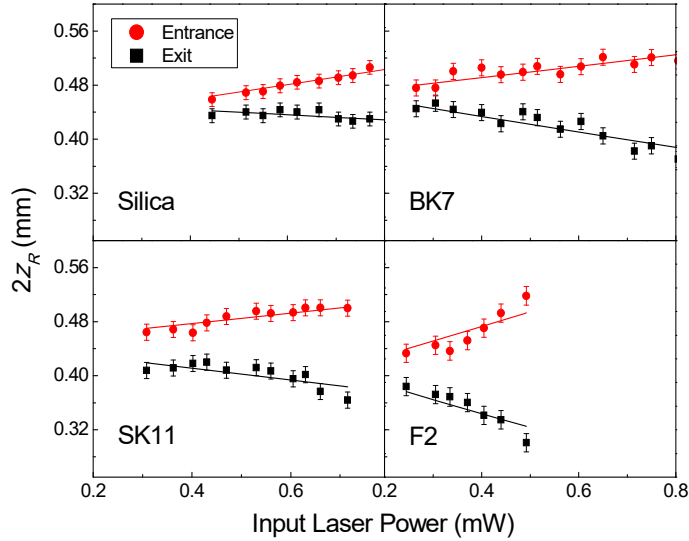


Fig. 12.7. $2z_R$ from the TH collected as a function of the pump laser power for silica, BK7, SK11, and F2 glasses. The solid lines were obtained from Eq. (12.27). Source: [26].

Table 12.2. Experimental n_2 values obtained from the ratio B_{glass} / B_{silica} .

Sample	n_2 -THG slope- $\times (2.71 \times 10^{-20} \text{ m}^2/\text{W})$	n_2 - $2z_R$ approach- $\times (2.71 \times 10^{-20} \text{ m}^2/\text{W})$
Silica	1.0	1.0
BK7	1.7	1.3
SK11	1.8	1.3
F2	3.8	3.1

In summary, in this section, we studied THG at interfaces of different optical glasses using femtosecond laser pulses. The TH intensities asymmetry can be explained by the SF effect, which increases the laser irradiance at the exit interface due to the reduction of the

beam waist. The beam waist radius should change as a function of sample nonlinearity and input laser power. It is important to mention that there are SF models already available in the literature [10, 17], but none of those correctly predict the results reported here. This is probably due to the simplifying assumptions they make such as flat-top beams or unchangeable beam waist radius at the focal plane. We proposed a simple empirical equation, Eq. (12.16), considering there is a change in the beam waist due to SF where the beam waist decreases as the laser power and/or material's nonlinearity increases. Using our model we could follow the TH signal behavior as a function of the laser power and estimate the material's nonlinear refractive index.

12.5. THG at the Interfaces of a Cuvette Filled with Organic Solvents

In the previous section, we discussed the effect of SF occurring in Kerr media with pure electronic nonlinear refractive index n_2 (where $n_2 \propto \chi^{(3)}(-\omega; \omega, -\omega, \omega)$) on the THG intensity profile at interfaces. We showed that SF effect clearly affects the THG at entrance and exit interfaces of a single slab, and that the power dependence, of the signal at the exit interface, can be greater than 3 due to the reduction of the beam waist radius. Now, using a cuvette containing liquid, we extend the study to a configuration with four different interfaces separating different materials which include liquids with the presence of both electronic and non-instantaneous nonlinearities. In fact, the linear and nonlinear optical properties of the surrounding medium significantly affect the THG signal depending on the propagation direction. Using silica and Schott B270 Superwhite crown glass cuvettes, filled with three organic solvents (acetone, chloroform, and dimethyl sulfoxide (DMSO)), we have a large variety of interfaces with materials of different linear and nonlinear properties. The nonlinearity of organic solvents may present significant contributions from both pure fast electronic nonlinearities and slow nuclear contributions, such as orientational effects.

The TH peaks for each interface in a particular cuvette-solvent pair have their own magnitude and power dependence. Similar to the single slab interfaces, the cuvette interfaces are also very asymmetric in terms of beam propagation direction. In other words, the THG signal observed when the laser beam propagates from the cuvette to the solvent may be completely different from that generated in the opposite direction. Such dependence was never carefully studied before and it is very relevant for materials characterization and the understanding of fundamental properties. There are several works studying the THG response at interfaces between liquid and solid media where the unknown third-order nonlinear optical susceptibilities of liquids are determined [3, 4, 32-34]. All these works require the precise value of phase-matching integrals (J integrals) for the case of tightly focused beams. Here we also decided to consider the contribution of J since we have more complex material's interfaces. This is not a simple task for interfaces between two media (1 and 2) with unequal dispersion ($\Delta k_1 \neq \Delta k_2$) and nonlinear susceptibilities ($\chi_1^{(3)} \neq \chi_2^{(3)}$) [3]. It is possible to do a fast $\chi^{(3)}$ ratio determination if the J integrals are previously determined with a numerical calculation [4, 29]. In this section we are going to present THG measurements at liquid-glass interfaces and interpret the results. We shall determine the nonlinear optical coefficients of the liquids using a method

that makes use of the phase-matching integral [4] and, also, show how we can obtain these coefficients using our approach that considers the SF contribution. In our method the nonlinear refractive indices, n_2 , and third-order nonlinear susceptibilities of the solvents, $\chi^{(3)}$, are determined simultaneously by the THG signals using the cuvette walls as a reference. The results are compared and discussed.

12.5.1. Experiments

The experimental setup is the same one shown in Fig. 12.2, where the glass slab is replaced by the cuvette with solvent. The OPA was also set to generate pulses at 1300 nm, leading the TH to be centered at 433 nm. This fundamental wavelength was carefully chosen because it has negligible linear absorption, as does its TH, in all solvents studied. Note that organic solvents usually present several absorptions peaks in the IR range which severely restricts the useful spectral range of THG.

As done in the previous experiment, the sample was scanned through the focus of the laser beam at $z = 0$. The laser is focused into the sample with a 50 mm focal length lens ($b \approx 0.3$ mm, $w \approx 8$ μ m) and a 60 mm focal length lens is placed after the sample to collect the TH light into the detector. Since we are using a relatively high laser irradiance (1.4 to 2.0×10^{16} W/m²), the TH generated at the interfaces was strong enough to be detected by a large area silicon PIN detector coupled into a lock-in amplifier. Using the same setup with a germanium PIN detector, we carried out open-aperture Z-scan measurements [15] to verify the presence of any absorptive nonlinearity. With the irradiances used here, no significant two-photon absorption was observed in any of the solvents.

As shown in Fig. 12.8 (I), there are four different interfaces for laser focus positioning when a silica cuvette is filled with solvent: (a) air-SiO₂ (entrance interface), (b) SiO₂-solvent, (c) solvent-SiO₂, and (d) SiO₂-air (exit interface). A similar configuration occurs for the B270 cuvette. Owing to these interfaces, there are four positions giving rise to THG, as depicted in Fig. 12.8 (II).

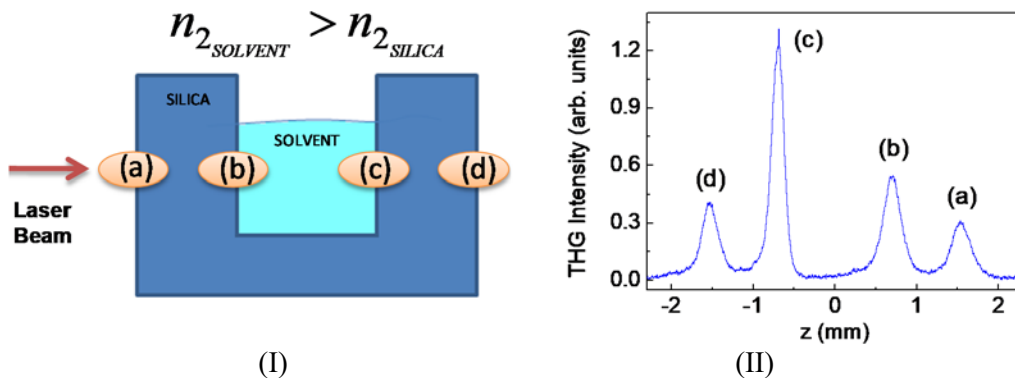


Fig. 12.8. (I) Definition of the cuvette interfaces and, (II) typical TH intensity curve for silica/DMSO system. Source: [30].

12.5.2. Results and Discussion

As already mentioned, we used two cuvettes (silica and B270) and three different organic solvent: acetone, chloroform, and DMSO. Table 12.3 shows their linear refractive indices, n_0 , for the fundamental and TH wavelengths. Both commercial cuvettes have a 2 mm internal optical path with two windows with 1.2 mm thickness each. The total cuvette thickness is 4.4 mm.

Table 12.3. Dispersion properties of the materials used.¹

	n_0 (0.43 μm)	n_0 (1.3 μm)	$ \Delta k $ (μm^{-1})	$ b\Delta k $
Silica	1.467	1.447	0.290	87
B270	1.534	1.510	0.348	104
Acetone	1.368	1.352	0.232	70
Chloroform	1.458	1.434	0.348	104
DMSO	1.492	1.464	0.406	122

In order to achieve a fast $\chi^{(3)}$ (n_2) ratio calculation, we first found several values of J by numerically solving Eq. (12.22) (squares in Fig. 12.9), and then we found the best parameters that fit Eq. (12.25) to these points. We obtained the values $C_1 = -0.2396$, $C_2 = 0.23924$, $C_3 = 4.959$, $C_4 = 19.40569$, and $C_5 = 2.93 \times 10^{-8}$.

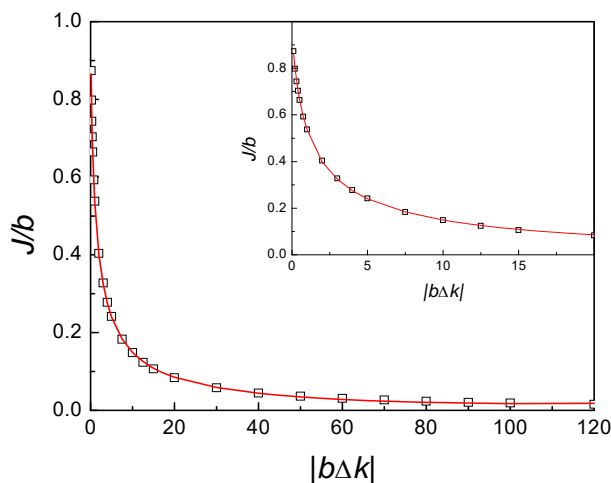


Fig. 12.9. Open squares show values of J/b as a function of $|b\Delta k|$ calculated with Eq. (12.22). The solid curve is the best fit obtained by using Eq. (12.25) and the inset is just an expanded view. Source: [30].

¹The refractive indices were obtained using the Sellmeier coefficients of Silica [35], B270 [36], acetone [37], chloroform [38], and DMSO [39]. The confocal parameter is $b = 0.3$ mm.

Fig. 12.10 shows typical TH profiles for the cuvettes containing the three solvents measured, where four TH peaks with different amplitudes can be seen. By making measurements with different laser powers, each curve presents its own THG peak values at each interface. We first used the procedure described in Barille's paper [4], based on the ratio between the THG peak values from two distinct interfaces: air/glass and glass/solvent, without considering SF, to find the nonlinear susceptibilities ratios of our solvents with respect to silica and B270. As it can be seen in Fig. 12.10, peaks (b) and (c) (or (a) and (d)) have different magnitudes due to the presence of SF. In this way, there are four possible values for the nonlinear susceptibility ratio for each laser power. These are obtained by using the proper $I^{(3\omega)}$ ratio in Eq. (12.24), i.e., $I^{(3\omega)}$ generated in surface (b) divided by those of (a) and (d), and from surface (c) with (a) and (d). Using these four values we determined an average value and the deviation of each of them from it. Fig. 12.11 shows that the nonlinear susceptibilities ratios are nearly constant with the laser power, but the deviations increase as a function of the laser power and nonlinearity strength. These deviations reflect the asymmetry existing in the TH profiles whose origin is the SF effect. DMSO has the highest discrepancy among all solvents because it has the highest nonlinearity. Table 12.4 shows the averaged (AVG) nonlinear susceptibilities ratios obtained with all different laser powers (Fig. 12.11) where the respective uncertainties came from the difference between the maximum (highest laser power) and the minimum values (lowest laser power). From the experimental data, we can also find that the ratio $\chi^{(3)}_{\text{B270}}/\chi^{(3)}_{\text{silica}}$ is about 1.5 ± 0.2 . It is important to observe that in this method, the error rapidly increases with laser power and nonlinearity due to the influence of SF. Note that in the model given by Barille et al. [4], no SF is taken into account.

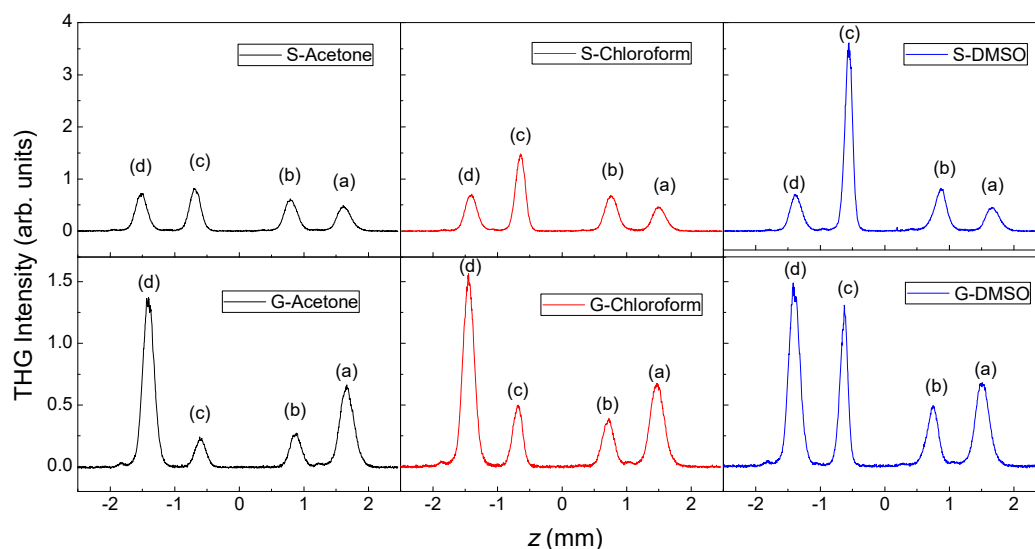


Fig. 12.10. THG signals for three different solvents as a function of the sample position for the four interfaces. We used two different cuvettes: silica (S) and B270 glass (G). The curves were obtained using pump powers of 0.430 mW and 0.415 mW for silica and B270 cuvettes, respectively. Other pump powers present very similar curves. Source: [30].

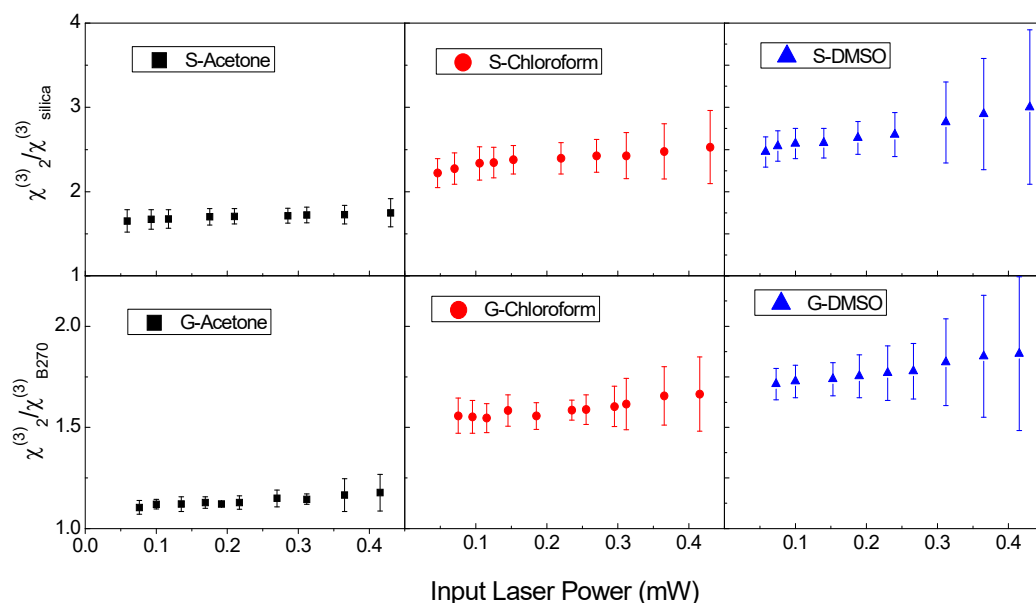


Fig. 12.11. Nonlinear susceptibility ratios for all solvents obtained with TH peak ratios using Eq. (12.24) for different laser powers. We used two different cuvettes: silica (S) and B270 glass (G). The error bar for each laser power comes from dispersion among the four possible ratios. Source: [30].

Table 12.4. Nonlinear susceptibility ratios of the solvents obtained by averaging the experimental data of Fig. 12.11¹.

Sample	$\chi_2^{(3)}/\chi_{silica}^{(3)}$	J/b	$\chi_2^{(3)}/\chi_{silica}^{(3)}$ AVG	$\chi_2^{(3)}/\chi_{B270}^{(3)}$ AVG
Silica	1	0.0209	-	-
B270		0.0195	-	-
Acetone	1.5 [40]	0.0252	1.7 ± 0.1	1.1 ± 0.1
Chloroform	2.3 [40]	0.0195	2.4 ± 0.3	1.6 ± 0.1
DMSO	2.4 [41]	0.0200	2.7 ± 0.6	1.8 ± 0.2

Next, we determined the nonlinearity ratio using the procedure where SF is considered. The power-dependent TH signals for each interface of the two cuvettes are plotted in Figs. 12.12 and 12.13. As seen, the slopes of the TH peaks can be significantly different from 3, the value expected for a third-order nonlinear process. This can be explained by taking into account the SF effect. We propose a self-consistent procedure to explain the results by using Eq. (12.18) and one parameter, the nonlinear refraction of the cuvette, as a reference. In order to simplify our study, we start by separately analyzing the interfaces that present the same $\Delta\chi^{(3)}$, that is, interfaces (a) and (d), and interfaces (b) and (c).

¹Here we can consider the silica nonlinearity: $n_2 = 2.7 \times 10^{-20} \text{ m}^2/\text{W}$, $\chi^{(3)} = 2.0 \times 10^{-22} \text{ m}^2/\text{V}^2$ [31].

According to Eq. (12.13), they should generate the same TH if no SF was present. As seen in Fig. 12.10, the harmonic generated at interface (d) is more intense than one generated at interface (a), and also, it presents a slope higher than 3 (Figs. 12.12 and 12.13) which agrees with previous results obtained with single glass slabs.

In interface (c), the laser beam already propagated from the solvent to silica (or B270) and, as the nonlinearity of all solvents used here is greater than that of the cuvettes walls, the high value of the slope (> 3) can be explained with the SF effect in the output interface of the solvent. In our model, the SF in the output interfaces, which reduces the beam waist, explains why the slope is greater than 3, which is also responsible for the high magnitude of these peaks in comparison to peaks (b).

Fig. 12.10 shows that the TH generated at the interface (c) is always greater than that generated at (b), except for acetone in the B270 cuvette. In this case, the THG signals are small because their nonlinearities are of the same order of magnitude. This implies that interfaces (b) and (c) experience roughly the same beam waste reduction and, even though the acetone nonlinearity is slightly greater than that of B270, it is not enough for the peak (c) to overcome the peak (b). Furthermore, acetone presents the largest linear absorption at the pump frequency which reduces the irradiance at the interface (c). This absorption can be noticed in peak (d), which is the output of the B270 cuvette, where the THG signal is somewhat smaller with acetone in comparison with other solvents.

Nonlinear susceptibilities and refractions can be found by fitting the slope of the power dependence of each peak, as shown in Figs. 12.12 and 12.13. In order to do it, we fit the experimental data by means of Eq. (12.19) and obtain the coefficients A ($A = A'/w_0^6$) and B for each solvent. The constant A is related to the symmetry breaking of the nonlinear susceptibility at the interface, $\Delta\chi^{(3)}$, B is proportional to the nonlinear refraction, n_2 , responsible for changing the beam waist radius owing to the SF action. We assume that B must be the same at the interfaces (b) and (d), $B_b = B_d$, since the SF occurs in the same material of the cuvette (silica or B270), and must be zero at the interface (a), $B_a = 0$, since the THG slope is unaffected by SF, i.e., the slope is constant and equal to 3. Also, we use the condition that the two internal interfaces must present the same value of A , $A_b = A_c$, as well as the two external interfaces, $A_a = A_d$. Using these considerations it is possible to carry out the adjustment at each interface by placing only one parameter as a variable, which provides more reliable values for each coefficient. We start with the interface (a) where we already know the coefficient B , ($B_a = 0$), and then we obtain the parameter $A_a = A_d$ by the fitting.

A_d must be used at the interface (d) to obtain $B_d = B_b$. The procedure follows by using B_b to find $A_b = A_c$ at the interface (b). Finally, knowing A_c we obtain B_c at interface (c). Tables 12.5 and 12.6 summarize all values found for silica and B270 cuvettes. Both A and B are related to n_2 (or $\chi^{(3)}$) and it is possible to estimate the nonlinearity of each solvent by using directly the ratio $B_c/B_b = n_{2,\text{solvent}}/n_{2,\text{silica or B270}}$, or by using Eq. (12.24) just replacing I 's by respective A 's, since A carries information about the TH intensity. As can be seen, it is not possible to get the values of nonlinearity ratios by simply using the A 's ratios. It is necessary to use an indirect method based on Eq. (12.24). In this case, it is also necessary

to use the J values for each solvent and cuvette material. Using a linear fit in a log-log scale we can note the slope is equal to 3 only at the first interface, where there is no SF contribution in the TH peak value. For the other interfaces, the power law dependence of the THG is greater than 3 due to the influence of the Kerr lens effect.

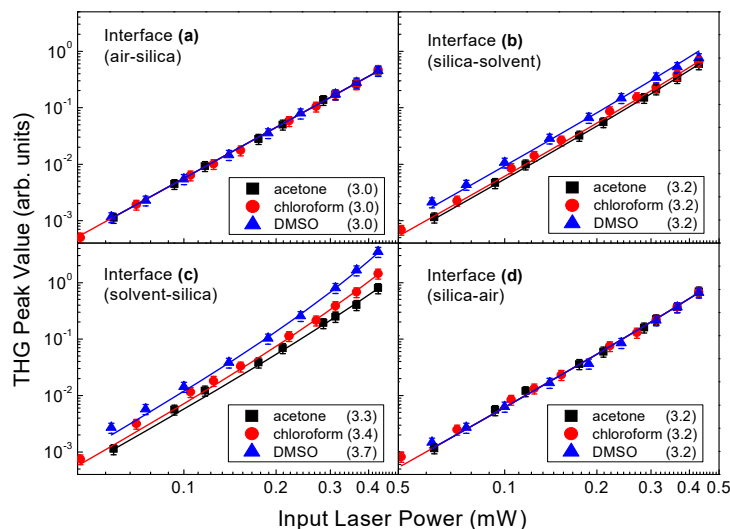


Fig. 12.12. Peak values of THG curves at the silica cuvette interfaces as a function of input laser power for all solvents and all interfaces. The symbols are the experimental data and the solid lines are the fits using the empirical model. The numbers in parentheses are the approximate slopes. Source: [30].

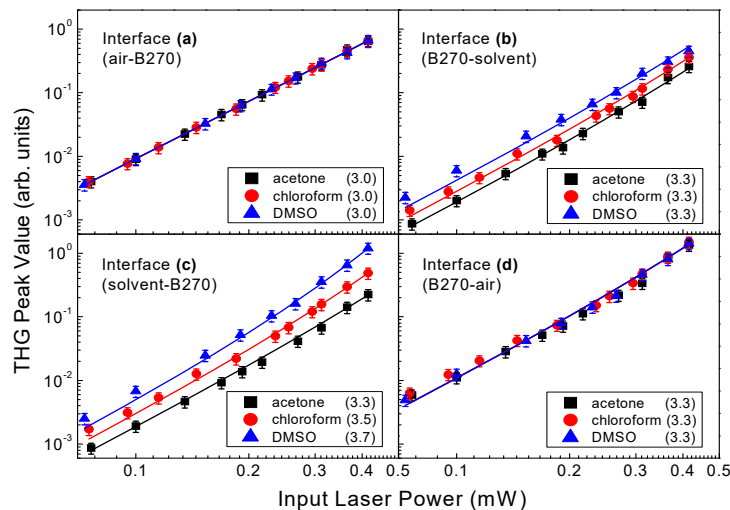


Fig. 12.13. Peak values of THG curves at the B270 cuvette interfaces as a function of input laser power for all solvents and all interfaces. The symbols are the experimental data and the solid lines are the fits using the empirical model. The numbers in parentheses are the approximate slopes. Source: [30].

Table 12.5. Coefficients A and B obtained for the silica cuvette for interfaces (b) and (c)¹.

Sample	(b)	(c)
Acetone	$A = 4.95 \pm 0.06$ $B = 0.15 \pm 0.01$	$A = 4.95 \pm 0.06$ $B = 0.25 \pm 0.01$
Chloroform	$A = 5.6 \pm 0.1$ $B = 0.15 \pm 0.01$	$A = 5.6 \pm 0.1$ $B = 0.41 \pm 0.02$
DMSO	$A = 8.5 \pm 0.1$ $B = 0.15 \pm 0.01$	$A = 8.5 \pm 0.1$ $B = 0.55 \pm 0.02$

Table 12.6. Coefficients A and B obtained for the B270 cuvette for interfaces (b) and (c)².

Sample	(b)	(c)
Acetone	$A = 1.62 \pm 0.02$ $B = 0.28 \pm 0.02$	$A = 1.62 \pm 0.02$ $B = 0.24 \pm 0.01$
Chloroform	$A = 2.38 \pm 0.04$ $B = 0.28 \pm 0.02$	$A = 2.38 \pm 0.04$ $B = 0.39 \pm 0.02$
DMSO	$A = 3.55 \pm 0.05$ $B = 0.28 \pm 0.02$	$A = 3.55 \pm 0.05$ $B = 0.55 \pm 0.02$

As expected, interfaces (a) and (d) for both silica and B270 cuvette have about the same results independent of the solvent.

Table 12.7 summarizes the results obtained for nonlinear refractive index ratios using coefficients B and A , in silica and B270 cuvettes. Here, we observe that the coefficients A and B lead to roughly similar n_2 ratios. The discrepancies between the two values increase systematically with the sample's nonlinearity as in the Barille's method [4]. However, in our model, Eq. (12.18), the coefficients A and B have reverse contribution to the THG intensity. An overestimation of A 's corresponds to underestimation of B 's and, in this way, the most correct values can be obtained by averaging the two values. In fact, if we take the average ratio values of the n_2 's in Table 12.7, for each solvent, we can get about the same values shown in Table 12.4. Since all measurements were carried out using the same laser configuration, we can also find the ratio of the $n_2(\text{B270})/n_2(\text{silica})$ using the respective coefficients A and B . Using A 's (using Eq. 12.20) and B 's we got about 1.37 and 1.87, respectively. The average value is about 1.6 ± 0.2 which is very similar to one obtained from intensity ratio measurements, 1.5 ± 0.2 , and they are in the range of that expected for such borosilicate type of glass [42]. We believe that the main reason for the difference between the ratio's values obtained by these two coefficients is the simplicity of our model. Despite this drawback, our model explains very well the contribution of the SF on the change of the slope for the THG versus P curves in organic solvents.

¹For interfaces (a) and (d) for all solvents, $A = 5.6 \pm 0.1$, $B = 0$ and $A = 5.6 \pm 0.1$, $B = 0.15 \pm 0.01$, respectively.

²For interfaces (a) and (d) for all solvents, $A = 9.2 \pm 0.1$, $B = 0$ and $A = 9.2 \pm 0.1$, $B = 0.28 \pm 0.02$, respectively.

Table 12.7. Ratios $n_{2,\text{solvent}}/n_{2,\text{cuvette}}$ obtained with the coefficients B and A ¹.

Sample	Literat. (silica)	B (silica)	A (silica)	AVG (silica)	B (B270)	A (B270)	AVG (B270)
Acetone	1.7 [40]	1.7 ± 0.1	1.6 ± 0.1	1.7 ± 0.1	0.9 ± 0.1	1.1 ± 0.1	1.0 ± 0.1
Chloroform	2.3 [40]	2.7 ± 0.2	2.1 ± 0.2	2.4 ± 0.3	1.4 ± 0.1	1.5 ± 0.1	1.5 ± 0.1
DMSO	2.4 [41]	3.7 ± 0.3	2.3 ± 0.2	3.0 ± 0.7	2.0 ± 0.2	1.6 ± 0.1	1.8 ± 0.8

It is also possible to notice the contribution of the SF in narrowing the peaks (c)'s width. In each measurement, the peak (c) is the narrowest peak in the curve. Although this effect is expected at the exit interface in the presence of the SF, here, it is not clear about entrance or exit interfaces since both media no longer have negligible nonlinearities. In fact, according to our model, there is competition between narrowing (exit interface of the solvent) and broadening (entrance interface of the cuvette wall) in these interfaces. In this way, we have not considered this study here because no clear behavior was observed in terms of narrowing or broadening of the two, (b) and (c), internal interfaces. In general, we notice that both peaks, (b) and (c), were narrowed (beam waist reductions) and, consequently, an increase in the peaks' slopes were observed.

12.6. THG at Organic Solvents Interfaces as a Function of Pulse Duration

As mentioned, although the TH is a pure electronic process, the SF may have electronic and other nonlinear mechanism contribution. The instantaneous electronic nonlinearity is a constant as a function of the pulse width, which is not the case for the orientational one. Orientational n_2 is only constant for pulses much longer than the orientation response time [43]. Hence, it is possible to change the effective nonlinearity of a solvent controlling the pulse duration (chirp) if the orientational response time is in the order of the pulse duration. Here, we propose a study of the SF effect on the THG just changing the pulse duration of the excitation beam.

In this experiment we have used the same laser system Ti:sapphire (CPA 2001 – Clark MXR) which delivers pulses with 150 fs duration at 1 kHz repetition rate. This time, we did not use the OPA since we wanted to control the pulse duration. The laser emission is centered at $\lambda = 775$ nm with THG at the ultraviolet, 258 nm. The laser beam is focused into the sample using a 50 mm focal length lens ($z_R \approx 0.17$ mm, $w_0 \approx 7$ μ m, $I_0 \approx 1 \times 10^{16}$ W/m²). After the sample we use a 60 mm focal length lens to collect the TH light into the detector (UV-VIS USB 2000+, Ocean Optics). This collection lens must be transparent to 258 nm and, in this way, we have used a silica lens, instead of the BK7 lens typically used. The temporal pulse duration (~ 150 fs to ~ 2 ps) can be directly controlled by the laser pulse compressor already inside the laser system. The pulse duration was measured by nonlinear refraction measurements using Z-scan technique, in the thin sample condition and silica as a reference ($n_2 = 2.7 \times 10^{-20}$ m²/W). The Z-scan ΔT_{pv} for a

¹AVG corresponds to averaged values obtained with the A and B coefficients.

refractive signature is proportional to the nonlinearity of the sample, the sample's thickness and the laser irradiance. Using a well-known sample, silica, with known thickness and n_2 , it is possible to calibrate the irradiance necessary to obtain the ΔT_{pv} measured. Finally, with the laser power and beam waist radius, w_0 , one can find the pulse duration. As sample we have used a silica cuvette filled with the same organic solvents: acetone, chloroform and DMSO.

12.6.1. Result and Discussion

In the cuvette we have four interfaces: air-silica (entrance interface), silica-solvent, solvent-silica and silica-air (exit interface), that were labeled as (a), (b), (c) and (d), respectively, as we did in the previous section, Fig. 12.8. The cuvette was scanned along the z axis and the TH signal was collected into the spectrometer. Due to the shorter TH wavelength, peaks (a) and (b) are strongly attenuated by the solvent absorbance and, in this way, we can only use the peaks (c) and (d) for analysis.

Fig. 12.14 (I) shows THG measurements for different pulse durations. In order to preserve the same TH signal at the interface (d), that we can use as a reference due to the electronic origin of its nonlinear response, the irradiance was set to be approximately constant, for each solvent, for the different pulse durations.

As it can be seen in Fig. 12.14 (I), the TH signal for acetone and chloroform at the interface (c) is enhanced for longer pulses and there is no significant variation for DMSO. These results indicate that, for long pulses, the orientational nonlinear refractive index increases its contribution on the SF effect in acetone and chloroform. By our results, chloroform has somewhat stronger orientational nonlinearity in comparison with acetone, which is in agreement with Rau et al. [40]. In the case of DMSO, it seems that this solvent has no significant orientational nonlinearity for the pulse widths we have used here. That means the DMSO may have a low orientational contribution and most of its nonlinearity has electronic origin (the pure electronic nonlinearity of the solvent should have the same pulse width dependence of the silica and, consequently, no enhancement should be observed) [44]. Also, the TH generated at the entrance interface, peak (a), disappears for acetone due to the strong linear absorption at 258 nm. For chloroform and DMSO, it is possible to see the peak (a) because they have lower linear absorption at the TH wavelength, Fig. 12.14 (II). Peaks (b)s are not seen in the Fig. 12.14 (I), for none of the solvents, probably due to the presence of both linear and nonlinear absorption. Our results, are also in accordance with the expected by the SF model.

12.7. Conclusions

In this chapter, we have investigated the THG at the interfaces on different materials and understood the influence of the SF on its intensity profile. In a first experiment, we have studied slabs of different optical glasses using femtosecond laser pulses. The results have shown that the SF is the main reason for the TH intensity asymmetric profile and dependence of pump power at the interfaces.

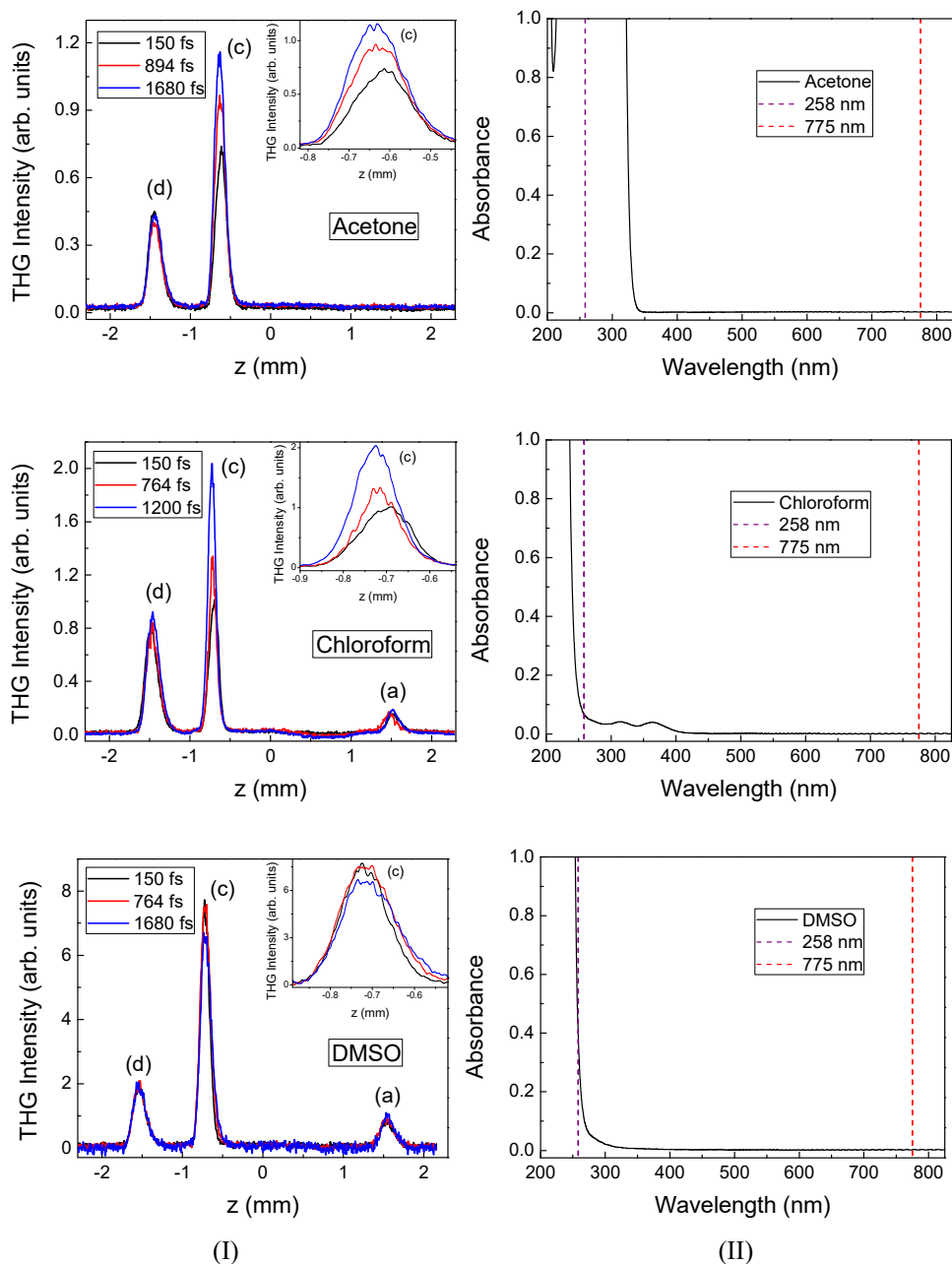


Fig. 12.14. (I) Integrated THG signal. The interfaces were labeled as (a) air-silica (entrance interface); (b) silica-solvent (there is no signal at this interface); (c) solvent-silica. and (d) silica-air (exit interface). The TH profiles are presented for acetone, chloroform and DMSO for different pulse durations. (II) Absorbance spectrum for each solvent. The insets in (I) show the peaks (c) in more detail. Source: Adapted from [26].

Although the THG is a pure third-order nonlinear process, the TH power dependence at the interfaces can be greater than 3 due to the influence of another nonlinear process, the SF, which changes the beam waist at the interfaces. Taking into account the SF contribution, it was possible to explain the TH behavior and obtain the glasses nonlinearity.

In a second experiment, we have measured the THG of solvents inside cuvettes. To be clearer, THG at interfaces of solvent and cuvette walls using femtosecond laser pulses. The solvents may have, in addition to the electronic nonlinearity, also an orientational contribution, which is not present in optical glasses. Even with these differences, our model works and, again, the SF effect, which now has contribution of pure fast electronic and orientational nonlinearity, is important to explain the THG signal.

Results related to the THG intensity profile as a function of pulse duration were reported and a clear variation on the TH peak values at the solvent-silica interface was observed. This indicates an enhancement of the SF due to the orientational nonlinearity contribution which increases with pulse duration. This result is in accordance with our expectation since SF from orientational nonlinearity is not instantaneous and does depend on the pulse width.

In general, the studies presented in this text correspond to advancements in both basic and applied strands of research in THG and/or nonlinear optics. We were able to understand the contribution of the SF on the THG at interfaces, which is very important for fundamental and applied point of view such as for materials characterization.

Acknowledgements

The authors gratefully acknowledge funding support from the São Paulo Research Foundation (FAPESP) under Grant 2011/22787-7.

References

- [1]. T. Y. F. Tsang, Optical third-harmonic generation at interfaces, *Physical Review A*, Vol. 52, Issue 5, 1995, pp. 4116-4125.
- [2]. J. M. Schins, T. Schrama, J. Squier, G. J. Brakenhoff, M. Muller, Determination of material properties by use of third-harmonic generation microscopy, *Journal of the Optical Society of America B, Optical Physics*, Vol. 19, Issue 7, 2002, pp. 1627-1634.
- [3]. V. Shcheslavskiy, G. Petrov, V. V. Yakovlev, Nonlinear optical susceptibility measurements of solutions using third-harmonic generation on the interface, *Applied Physics Letters*, Vol. 82, Issue 22, 2003, pp. 3982-3984.
- [4]. R. Barille, L. Canioni, L. Sarger, G. Rivoire, Nonlinearity measurements of thin films by third-harmonic-generation microscopy, *Physical Review E*, Vol. 66, Issue 6, 2002, pp. 067602 (1-4).
- [5]. J. A. Squier, D. N. Fittinghoff, C. P. J. Barty, K. R. Wilson, M. Muller, G. J. Brakenhoff, Characterization of femtosecond pulses focused with high numerical aperture optics using

- interferometric surface-third-harmonic generation, *Optics Communications*, Vol. 147, Issue 1-3, 1998, pp. 153-156.
- [6]. T. Tsang, M. A. Krumbugel, K. W. DeLong, D. N. Fittinghoff, R. Trebino, Frequency-resolved optical-gating measurements of ultrashort pulses using surface third-harmonic generation, *Optics Letters*, Vol. 21, Issue 17, 1996, pp. 1381-1383.
- [7]. Y. Barad, H. Eisenberg, M. Horowitz, Y. Silberberg, Nonlinear scanning laser microscopy by third harmonic generation, *Applied Physics Letters*, Vol. 70, Issue 8, 1997, pp. 922-924.
- [8]. R. Carriles, D. N. Schafer, K. E. Sheetz, J. J. Field, R. Cisek, V. Barzda, A. W. Sylvester, J. A. Squier, Invited review article: Imaging techniques for harmonic and multiphoton absorption fluorescence microscopy, *Review of Scientific Instruments*, Vol. 80, Issue 8, 2009, pp. 081101 (1-23).
- [9]. L. G. Gouy, Sur une propriété nouvelle des ondes lumineuses, *Comptes Rendus des Séances de l'Académie des Sciences*, Vol. 110, 1890, pp. 1251-1253.
- [10]. R. W. Boyd, Nonlinear Optics, 3rd Edition, *Academic Press*, San Diego, 2008.
- [11]. G. C. Bjorklund, Effects of focusing on third-order nonlinear processes in isotropic media, *IEEE Journal of Quantum Electronics*, Vol. QE11, Issue 6, 1975, pp. 287-296.
- [12]. E. C. Barbano, J. P. Siqueira, C. R. Mendonca, L. Misoguti, S. C. Zilio, Broadband third-harmonic generation on interfaces using femtosecond pulses, in *Proceedings of the Conference on Nonlinear Frequency Generation and Conversion: Materials, Devices and Applications X*, San Francisco, CA, USA, 24-27 January 2011, pp. 79170T (1-4).
- [13]. D. Stoker, M. F. Becker, J. W. Keto, Optical third-harmonic generation using ultrashort laser pulses, *Physical Review A*, Vol. 71, Issue 6, 2005, pp. 061802 (1-4).
- [14]. J. W. Wilson, R. A. Bartels, Coherence-modulated third harmonic generation for vibrational spectroscopy: a theoretical treatment, *Journal of the Optical Society of America B, Optical Physics*, Vol. 29, Issue 8, 2012, pp. 1875-1883.
- [15]. M. Sheikbahae, A. A. Said, T. H. Wei, D. J. Hagan, E. W. Vanstryland, Sensitive measurement of optical nonlinearities using a single beam, *IEEE Journal of Quantum Electronics*, Vol. 26, Issue 4, 1990, pp. 760-769.
- [16]. M. Sheikbahae, A. A. Said, E. W. Vanstryland, High-sensitivity, single-beam n2 measurements, *Optics Letters*, Vol. 14, Issue 17, 1989, pp. 955-957.
- [17]. W. Koechner, Solid-State Laser Engineering, 6th Edition, *Springer-Verlag*, New York, 2006.
- [18]. J. F. Reintjes, Nonlinear Optical Parametric Processes in Liquids and Gases, *Academic Press*, Orlando, 1984.
- [19]. M. G. Papadopoulos, A. J. Sadlej, J. Leszczynski, Non-Linear Optical Properties of Matter: From Molecules to Condensed Phases, *Springer Science & Business Media*, Dordrecht, 2007.
- [20]. P. L. Kelley, Self-focusing of optical beams, *Physical Review Letters*, Vol. 15, Issue 26, 1965, pp. 1005-1008.
- [21]. J. P. McTague, C. H. Lin, T. K. Gustafson, R. Y. Chiao, Observation of filaments in liquid argon, *Physics Letters A*, Vol. 32, Issue 2, 1970, pp. 82-83.
- [22]. M. E. Mack, R. L. Carman, J. Reintjes, N. Bloembergen, Transient stimulated rotational and vibrational Raman scattering in gases, *Applied Physics Letters*, Vol. 16, Issue 5, 1970, pp. 209-211.
- [23]. D. Grischkowsky, Self-focusing of light by potassium vapor, *Physical Review Letters*, Vol. 24, Issue 16, 1970, pp. 866-869.
- [24]. J. E. Bjorkhol, A. Ashkin, CW self-focusing and self-trapping of light in sodium vapor, *Physical Review Letters*, Vol. 32, Issue 4, 1974, pp. 129-132.
- [25]. Y. R. Shen, The Principles of Nonlinear Optics, *John Wiley & Sons*, New York, 1984.
- [26]. E. C. Barbano, Third-harmonic generation at interfaces with femtosecond pulses: self-focusing contribution and nonlinear microscopy, PhD Thesis (Basic Physics), Instituto de Física de São Carlos, *Universidade de São Paulo*, São Carlos, SP, Brazil, 2016, p. 101.

- [27]. V. Magni, G. Cerullo, S. Desilvestri, ABCD matrix analysis of propagation of Gaussian beams through Kerr media, *Optics Communications*, Vol. 96, Issue 4-6, 1993, pp. 348-355.
- [28]. E. C. Barbano, S. C. Zilio, L. Misoguti, Influence of self-focusing of ultrashort laser pulses on optical third-harmonic generation at interfaces, *Optics Letters*, Vol. 38, Issue 23, 2013, pp. 5165-5168.
- [29]. A. N. Naumov, D. A. Sidorov-Biryukov, A. B. Fedotov, A. M. Zheltikov, Third-harmonic generation in focused beams as a method of 3D microscopy of a laser-produced plasma, *Optics and Spectroscopy*, Vol. 90, Issue 5, 2001, pp. 778-783.
- [30]. E. C. Barbano, K. Harrington, S. C. Zilio, L. Misoguti, Third-harmonic generation at the interfaces of a cuvette filled with selected organic solvents, *Applied Optics*, Vol. 55, Issue 3, 2016, pp. 595-602.
- [31]. D. Milam, Review and assessment of measured values of the nonlinear refractive-index coefficient of fused silica, *Applied Optics*, Vol. 37, Issue 3, 1998, pp. 546-550.
- [32]. R. S. Pillai, G. J. Brakenhoff, M. Muller, Anomalous behavior in the third harmonic generation z response through dispersion induced shape changes and matching $\chi(3)$, *Applied Physics Letters*, Vol. 89, Issue 11, 2006, pp. 111123 (1-3).
- [33]. R. S. Pillai, G. J. Brakenhoff, M. Muller, Analysis of the influence of spherical aberration from focusing through a dielectric slab in quantitative nonlinear optical susceptibility measurements using third-harmonic generation, *Optics Express*, Vol. 14, Issue 1, 2006, pp. 260-269.
- [34]. V. Shcheslavskiy, G. I. Petrov, S. Saltiel, V. V. Yakovlev, Quantitative characterization of aqueous solutions probed by the third-harmonic generation microscopy, *Journal of Structural Biology*, Vol. 147, Issue 1, 2004, pp. 42-49.
- [35]. I. H. Malitson, Interspecimen comparison of refractive index of fused silica, *Journal of the Optical Society of America*, Vol. 55, Issue 10, 1965, pp. 1205-1208.
- [36]. B270 Datasheet – 1993, <http://refractiveindex.info/?shelf=glass&book=SCHOTT-multipurpose&page=B270>
- [37]. J. Rheims, J. Koser, T. Wriedt, Refractive-index measurements in the near-IR using an Abbe refractometer, *Measurement Science & Technology*, Vol. 8, Issue 6, 1997, pp. 601-605.
- [38]. S. Kedenburg, M. Vieweg, T. Gissibl, H. Giessen, Linear refractive index and absorption measurements of nonlinear optical liquids in the visible and near-infrared spectral region, *Optical Materials Express*, Vol. 2, Issue 11, 2012, pp. 1588-1611.
- [39]. I. Z. Kozma, P. Krok, E. Riedle, Direct measurement of the group-velocity mismatch and derivation of the refractive-index dispersion for a variety of solvents in the ultraviolet, *Journal of the Optical Society of America B, Optical Physics*, Vol. 22, Issue 7, 2005, pp. 1479-1485.
- [40]. I. Rau, F. Kajzar, J. Luc, B. Sahraoui, G. Boudebs, Comparison of Z-scan and THG derived nonlinear index of refraction in selected organic solvents, *Journal of the Optical Society of America B, Optical Physics*, Vol. 25, Issue 10, 2008, pp. 1738-1747.
- [41]. G. Shi, C. He, Y. Li, R. Zou, X. Zhang, Y. Wang, K. Yang, Y. Song, C. H. Wang, Excited-state nonlinearity measurements of ZnPcBr₄/DMSO, *Journal of the Optical Society of America B, Optical Physics*, Vol. 26, Issue 4, 2009, pp. 754-761.
- [42]. Y. H. Chen, S. Varma, I. Alexeev, H. M. Milchberg, Measurement of transient nonlinear refractive index in gases using xenon supercontinuum single-shot spectral interferometry, *Optics Express*, Vol. 15, Issue 12, 2007, pp. 7458-7467.
- [43]. M. Reichert, H. H. Hu, M. R. Ferdinandus, M. Seidel, P. Zhao, T. R. Ensley, D. Peceli, J. M. Reed, D. A. Fishman, S. Webster, D. J. Hagan, E. W. van Stryland, Temporal, spectral, and polarization dependence of the nonlinear optical response of carbon disulfide, *Optica*, Vol. 1, Issue 6, 2014, pp. 436-445.
- [44]. M. L. Miguez, T. B. G de Souza, E. C. Barbano, S. C. Zilio, L. Misoguti, Measurement of third-order nonlinearities in selected solvents as a function of the pulse width, *Optics Express*, Vol. 25, Issue 4, 2017, pp. 3553-3565.

Chapter 13

Transformations and Evolution of Phase Singularities in Diffracted Optical Vortices

A. Bekshaev, O. Angelsky and S. G. Hanson

13.1. Introduction

Diffraction is one of the most traditional and deeply investigated phenomena of classical optics [1, 2], and it is difficult to believe that its further study can bring any peculiar news on the physical features of optical fields. Of course, there are many quantitative details and special cases of diffraction that still need refinement and further elucidation but the general principles of diffracted field formation and evolution seem to be firmly established and widely known. However, this is not the case with structured light fields that have become a hot topic of modern optics during the past decades [3], especially with light beams carrying optical vortices (OV) [4–6]. The edge diffraction of circular OV beams [7–23] shows many impressive non-trivial details associated with their special physical attributes: helical wavefront shape and transverse energy circulation. In case of the OV beam diffraction, common and well-studied diffraction effects (fringes, transverse diffusion of the light energy, etc. [1, 2]) are supplemented with the OV-specific diffraction transformations which seem bizarre and surprising at first glance. The first of such effects is the asymmetric penetration of light energy into the shadow region [9, 13–18] impressively testifying for the transverse energy circulation in the incident beam, which comes to light due to the beam symmetry violation caused by the diffraction obstacle [16] (see Fig. 13.2 below for an example). Beside this, much attention was recently paid to the distribution and migration of the OVs within the diffracted beam [7, 8, 11, 12, 14–15, 19–21]; especially, it was shown that even if the incident OV is stopped by the obstacle, it is restored after a certain propagation distance (OV regeneration [7, 8, 10, 19]). The interest to the OVs in the diffracted beams is supported by the peculiar character of the OV cores as amplitude zeros and phase singularities, whereby they are physically highlighted and can be precisely detected and localized [24–28], which is employed, e.g., in sensitive metrology [29–32].

It is well established, both theoretically and in experiment, that after diffraction of an incident circular OV beam, at which its singular point is not ‘screened’ by the obstacle, the OV core shifts from its initial axial position, and an m -charged OV is decomposed into a set of $|m|$ secondary single-charged ones thus forming the ‘singular skeleton’ [6] of the diffracted beam. During propagation of the diffracted beam, the OV cores move along intricate spiral-like trajectories [19, 23] carrying distinct ‘fingerprints’ of the incident beam and its disposition with respect to the diffraction obstacle. A similar transformation of the singular skeleton can be observed in a fixed cross section of the diffracted beam when the screen edge performs a monotonous translation in the transverse direction towards or away from the beam axis [20–22].

However, the singular skeleton evolution is not limited by the ‘smooth’ migration of the secondary OVs within the diffracted beam ‘body’. Generally, this process is accompanied by various topological reactions [4, 6]: the OV disappearance and regeneration [7, 8, 10], emergence of new OVs, their annihilation, etc. Normally, such events occur at the beam periphery and are related with the diffraction fringes, etc. [15, 19] but some sorts of topological reactions are intimately connected with the ‘regular’ OV migration and constitute its part [22]. Importantly, the progress of these reactions is highly sensitive to the incident beam properties and the diffraction conditions (e.g., the screen edge position or the propagation distance behind the screen plane). Therefore, in addition to the general physical interest, these topological events offer potentially valuable and prospective means for precise measurements and diagnostics of the OV beam’s characteristics.

In this chapter, we make an effort for a systematic study of the singular skeleton formation, evolution and transformations associated with the diffraction of circular OV beams. The presentation is based mainly on the recent works [20, 21, 23, 33]. In contrast to some previous studies [7, 8, 15, 19], the consideration is mostly restricted to the case of ‘weak diffraction perturbation’ (WDP) when the screen edge is located far enough from the incident beam axis. We do not specify exactly which perturbation can be called ‘weak’. Practically this implies that the beam visually preserves the initial circular shape immediately after the screen, which for typical OV beams takes place if the screen is separated from the axis by two or more beam radii measured at e^{-1} intensity level (see, for example, Figs. 13.17 (a) and 13.18 (a) for the illustration), but realizing that some important conclusions of the WDP-based reasoning can be applicable well beyond any formal limits of its validity. Surprisingly enough, it is the WDP situations that appear the most favourable for explicit manifestation of the OV-related phenomena in the diffracted beam singular skeleton behaviour.

We start with description of the experimental approach (Section 13.2) and of the general numerical means for the OV diffraction simulation (Section 13.3). The consideration is based on the Kummer [34] and Laguerre-Gaussian (LG) [4–6] models for the incident OV beam: the former adequately describes OV beams obtained from an initial Gaussian beam with the help of suitable vortex-generating elements widely used in experimental practice; the latter is a standard OV beam model convenient for calculations. Comparative analysis of the results obtained with the two different models will highlight the diffraction effects related to the vortex character of the incident beam ‘per se’ and separate them from

‘occasional’ consequences of the incident beam radial profile, etc. The physical nature of the diffraction-induced beam transformations is explained with the help of a simple analytical model of the diffracted field formation based on interference of the incident beam with the edge wave [2] formed due to the incident field scattered by the screen edge. The analytical model is refined by means of the asymptotic analysis of the Fresnel–Kirchhoff diffraction integral [21, 33] (Section 13.4).

Further, the developed theoretical instruments are applied to the singular skeleton analysis in the diffracted beam in both basic situations: when the screen edge is fixed but the observation plane moves along the propagation direction (z -dependent evolution, cf. Fig. 13.1) and when the observation plane is fixed and the diffracted beam structure changes due to the screen edge translation (a -dependent evolution). In Section 13.5, the simplified situation of diffraction of a low-order LG beam discloses the nature of the spiral-like OV trajectories (‘vortex filaments’) and its intrinsic relations with the vortex beam structure.

In the first approximation, these trajectories look smooth but further examination reveals the existence of topological discontinuities that are studied both for the LG and Kummer beams’ diffraction in Sections 13.6 and 13.7. We describe the typical manifestations of such discontinuities (‘jumps’) as a series of topological reactions associated with the birth of the OV dipole at a remote point of the beam cross section followed by collision of one of the dipole constituents with the initial OV and their subsequent annihilation. This enabled us to introduce the numerical criterion for the OV trajectory ‘jumps’ whose validity is demonstrated in a number of examples of the singular skeleton transformation both for the a -dependent and z -dependent diffracted beam evolution. The observed discontinuities are also interpreted based on the transverse projections of the everywhere smooth and continuous 3D vortex filaments in the diffracted field. Possible applications of the results and the prospective lines of further research are discussed in the conclusion.

13.2. Experimental Setup

We start with outlining the typical experimental situation in which the OV diffraction is explored [7–15, 19, 20] (see Fig. 13.1 (a)). A circular OV beam whose axis coincides with the z -axis is directed onto the diffraction obstacle S – an opaque screen with sharp vertical edge, mounted with possibility of precise adjustment of the off-axis distance a in the horizontal x -direction. The diffracted beam formed behind the screen is observed with the help of a CCD camera positioned orthogonally to the incident OV beam axis at different adjustable distances z from the obstacle. The nature of the incident OV beam is less important; many previous works [7–13] deal with the standard Laguerre-Gaussian beams of various topological charges that are characterized by the Gaussian envelope parameters: the waist radius at e^{-1} intensity level b_0 , the corresponding Rayleigh range [2]

$$z_{Rc} = kb_0^2, \quad (13.1)$$

and the longitudinal distance from the waist to the screen z_c (see Fig. 13.1 (b)). In particular, impressive experiments were performed by Arlt with the LG beams of zero

radial index $z_c = 0$ [9], which clearly show internal energy circulation in the OV beams. In Fig. 13.2 the original results of [9] are re-arranged according to the geometry of Fig. 13.1 (a); here, as well as in the whole chapter, all views of the beam cross section are presented as seen against the beam propagation (from the positive end of the optical axis z).

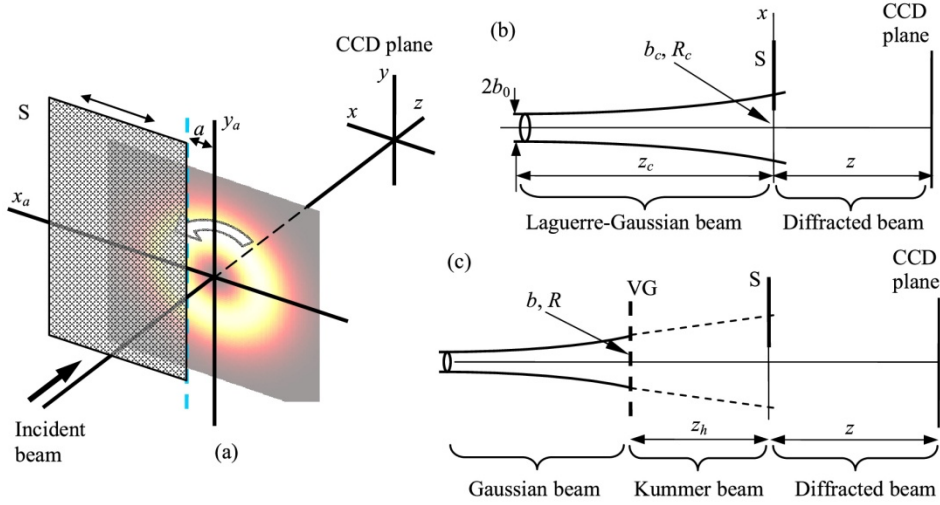


Fig. 13.1. (a) General view of the OV beam diffraction with the beam screening and the involved coordinate frames; schemes of (b) diffraction of the incident LG beam and (c) formation and diffraction of the incident Kummer beam. VG is the OV-generating element, S is the diffraction obstacle (opaque screen with the edge parallel to axis y , its position along axis x is adjustable), the diffraction pattern is registered in the observation plane by means of the CCD camera. Further explanations appear in the text.

Now our aim is to study the individual singularities of the diffracted beam rather than its general amplitude and phase profile. The adopted experimental scheme presented by Fig. 13.1 (c) [20] includes the He-Ne laser (wavelength $\lambda = 633$ nm, wavenumber $k = 2\pi/\lambda \approx 10^5$ cm $^{-1}$) that generates a Gaussian beam with the waist plane at the laser output window where its amplitude is distributed according to the equation

$$u_0(r) \propto \exp\left(-\frac{r^2}{2b_0^2}\right), \quad (13.2)$$

with r being the radial distance from the beam axis, the waist beam radius equals $b_0 = 0.165$ mm (corresponding Rayleigh range is $z_{Rc} = 27$ cm). After passing a distance $z_0 = 27$ cm, the beam impinges on the center of the vortex-generating element VG (in [20], the holographic grating with groove bifurcation – i.e., the ‘fork’ hologram [4, 5]); in the grating plane, the Gaussian beam radius b and its wavefront curvature radius R are

$$b = 0.232 \text{ mm}; \quad R = 54 \text{ cm} \quad (13.3)$$

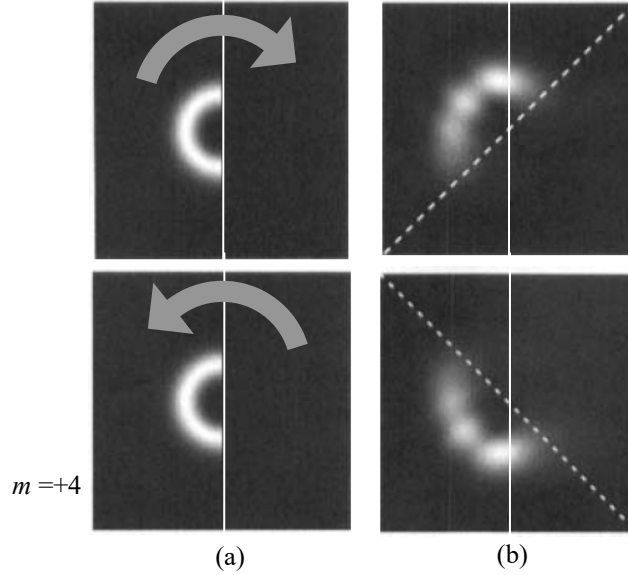


Fig. 13.2. Intensity distribution of the diffracted LG beams with the topological charge (azimuthal index) $m = \pm 4$ (the beam waist plane coincides with the screen plane S, screen edge position $a = 0$): (a) immediately behind the screen and (b) at the distance $z = z_R$ (13.1). The gray arrows show the energy circulation handedness, dotted line indicates the rotation angle 45° ; the vertical line is the screen edge projection.

(the value of b defines the transverse characteristic scale for the system geometry). The holographic grating produces a ‘fan’ of diffracted beams, $\pm n$ -th diffraction order carrying an OV beam with topological charge $\pm n$. The selected grating-generated OV beam was directed to the screen. The total distance from the grating VG to the screen S was $z_h = 11$ cm (see Fig. 13.1 (c)). In all cases, the OV beam approaching the screen plane was apparently circularly symmetric and of a negative topological charge m ($m = -1, -2, -3$); the transverse energy circulation in the screen plane is shown in Fig. 13.1 (a) by the curve arrow. The intensity profile of the diffracted beam was registered by the CCD camera with the sensitive area 4.8×3.6 mm², or 768×576 pixels where the pixel size is 6.25 μ m.

The exact localization of the OV cores in the diffracted beam cross section is not a trivial task; usual interference methods for the OV detection (by finding the fork-like defects in the fringe pattern formed due to interference with a reference wave without wavefront dislocation [4]) are hardly applicable because of the low spatial resolution. That is why several special approaches were proposed for the OV detection. For example, we can mention the procedures employing the Shack – Hartmann wavefront testing method [24, 25], 2D analytical signal representation and interpolation (‘OV metrology’) [26, 27], statistical algorithms based on reconstructed OV phase maps [28], etc. Since our primary goal is to precisely locate the OV cores in conditions where their existence is doubtless, these can be detected directly from the measured intensity patterns, searching the

amplitude zeros [20, 21]. Their coordinates were measured automatically by averaging over 150 consecutive snapshots, which facilitated a decrease in the fluctuation instabilities. To increase the spatial resolution, an interpolation procedure was employed: the near zero intensity distribution was approximated by an asymmetric paraboloid, and the exact zero position was recognized as a center of the equal-intensity ellipses with subpixel accuracy, $\sim 1 \mu\text{m}$.

This experimental scheme is adjusted for the study of a -dependent evolution of the diffracted beam pattern at several fixed distances z behind the screen; in the special conditions of [20, 21] $z = 30 \text{ cm}$, 60 cm and 82 cm .

13.3. Description of the Diffraction Model

13.3.1. General Principles of the Singular Skeleton Analysis

In conjunction with the experimental research, the diffraction of OV beams is analyzed by a mathematical model based on paraxial optics [2, 4, 5]. Let the incident monochromatic paraxial beam be described in the screen plane S (Fig. 13.1) by the slowly varying complex amplitude distribution $u_a(x_a, y_a)$; then in the observation plane at a distance z behind S the diffracted beam complex amplitude can be found via the Fresnel–Kirchhoff integral

$$u(x, y, z) = \frac{k}{2\pi iz} \int_{-\infty}^{\infty} dy_a \int_{-\infty}^a dx_a u_a(x_a, y_a) \exp\left\{\frac{ik}{2z}[(x-x_a)^2 + (y-y_a)^2]\right\}, \quad (13.4)$$

in any cross section, the instantaneous electric field of the paraxial beam field equals $\text{Re}[u(x, y, z) \exp(ikz - i\omega t)]$ with $\omega = ck$, c is the velocity of light. The integral (13.4) is calculated for different input distributions $u_a(x_a, y_a)$ characterizing the typical OV beams (see below) and the resulting distribution $u(x, y, z)$ is analyzed at given propagation distances z . Examples of the obtained intensity and phase patterns are presented in Fig. 13.3. The OV core positions are extracted from the calculated patterns in two mutually complementary ways: as the amplitude zeros (Fig. 13.3 (a, b)) and as points with indeterminate phase, where different contours of constant phase converge [15, 19] (Fig. 13.3 (c)). To enhance the visibility of the amplitude zeros, the calculated intensity distribution $I(x, y) = |u(x, y)|^2$ (Fig. 13.3 (a)) can be transformed according to the formula

$$I(x, y) \rightarrow I_t(x, y) = \left[|u(x, y)|^2\right]^{1/n}, \quad (13.5)$$

where n is a large enough positive integer, normally chosen $n = 15$. In addition to (or instead of) the transformation (13.5), the numerically calculated near-zero intensity distribution was approximated by a second-degree polynomial in x and y , and the OV-core position was identified as the center of the equal-intensity ellipses for this distribution (similarly to the interpolation procedure used for the experimental data interpretation, see Section 13.2).

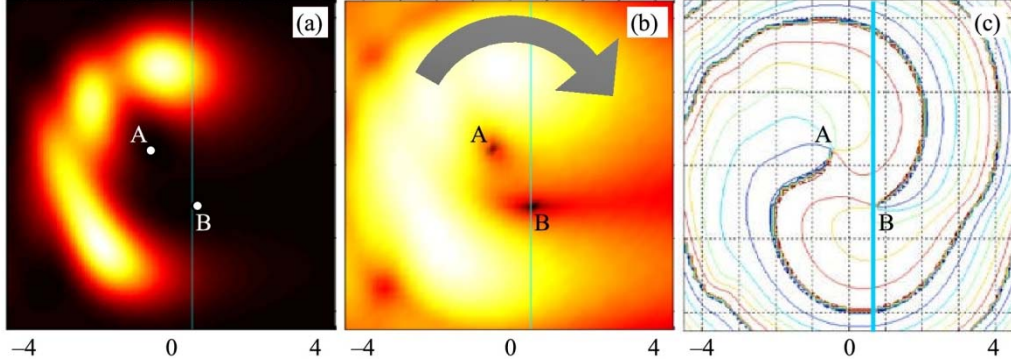


Fig. 13.3. Pseudo-color maps of the (a) intensity distribution $I(x, y)$, (b) transformed intensity distribution (13.5) with enhanced visibility of the amplitude zeros ($n = 15$) and (c) contours of constant phase (phase increment between adjacent contours is 1 rad). The images represent the profile of the diffracted beam with $m = -2$ at a distance 30 cm behind the obstacle, the screen edge is positioned at $a = 0.6b$; A and B mark the OV cores. All transverse scales are in units of b (13.3), the vertical cyan line is the screen edge projection, and the gray curved arrow shows the energy circulation in the incident OV beam.

In the phase maps (Fig. 13.3 (c)) the OV cores are situated at the ends of cuts (curvilinear ‘bundles’ of lines of different colors ‘tightly touching’ each other); along these cuts the phase experiences $\pm\pi$ ‘jumps’.

13.3.2. Description of the Incident OV Beams

Now let us specify more rigorously the incident OV beams whose diffraction is studied numerically. Addressing the experimental situation of Section 13.2 [22, 23], we will consider the case where the incident OV beam is described by the Kummer beam model [34] which is typical when an OV beam is formed from an initial Gaussian beam with the help of a VG element (see Fig. 13.1 (c)). In this case, the input complex amplitude distribution is $u_a(x_a, y_a) = u^K(x_a, y_a, z_h)$ where

$$u^K(x_a, y_a, z_h) = \frac{z_{he}}{z_h} \sqrt{\frac{\pi}{2}} (-i)^{|m|+1} \exp \left[\frac{ik}{2z_h} (x_a^2 + y_a^2) + im\phi_a \right]$$

$$\times \frac{z_R}{z_{he} - iz_R} e^{-A} \sqrt{A} \left[I_{\frac{|m|-1}{2}}(A) - I_{\frac{|m|+1}{2}}(A) \right], \quad (13.6)$$

where $\phi_a = \arctan(y_a/x_a)$ is the azimuth (polar angle) in the screen plane, m is the OV topological charge (corresponds to the phase increment $2m\pi$ upon the round trip near the beam axis), I_v denotes the modified Bessel function [36];

$$A = \frac{1}{4} \frac{z_R}{z_{he} - iz_R} \left[\frac{k}{z_{he}} (x_{ae}^2 + y_{ae}^2) \right], \quad z_R = kb^2, \quad (13.7)$$

$$z_{he} = \frac{z_h}{1 + z_h/R}, \quad x_{ae} = x_a \frac{z_{he}}{z_h}, \quad y_{ae} = y_a \frac{z_{he}}{z_h}, \quad (13.8)$$

b being the Gaussian beam radius at the VG plane, see Fig. 13.1 (c) and Eq. (13.3). Eqs. (13.6)–(13.8) admit the non-planar wavefront of the initial Gaussian beam, R is the wavefront curvature radius; equation for z_R in (13.7) just formally coincides with the Raleigh range definition [2] because for finite R , the quantity b is no longer associated with the beam waist.

Another beam type is the standard LG beam, which is more suitable in theoretical analysis; here, for simplicity, we restrict our consideration by the modes with zero radial index. In many situations, the LG beam model is considered universal and is used for approximate description of more complicated circular OV beams occurring in practice. In this case $u_a(x_a, y_a) = u^{LG}(x_a, y_a, z_c)$ where [2, 4, 5]

$$u^{LG}(x_a, y_a, z_c) = \frac{(-i)^{|m|+1}}{\sqrt{|m|!}} \left(\frac{z_{Rc}}{z_c - iz_{Rc}} \right)^{|m|+1} \left(\frac{x_a + i\sigma y_a}{b_0} \right)^{|m|} \exp\left(\frac{ik}{2} \frac{x_a^2 + y_a^2}{z_c - iz_{Rc}} \right), \quad (13.9)$$

where $\sigma = \text{sgn}(m) = \pm 1$, b_0 is the Gaussian envelope waist radius, z_c is the distance from the waist cross section to the screen plane (see Fig. 13.1 (b)), and $z_{Rc} = kb_0^2$ is the corresponding Rayleigh length (13.1); the current beam radius b_c and wavefront curvature radius R_c in the screen plane are determined by the known expressions

$$b_c^2 = b_0^2 \left(1 + \frac{z_c^2}{z_{Rc}^2} \right), \quad R_c = z_c + \frac{z_{Rc}^2}{z_c}. \quad (13.10)$$

Substituting (13.6) and (13.9) into (13.4) one can find the diffracted beam characteristics for arbitrary propagation distance z and arbitrary screen edge position a . This enables further determining the OV positions numerically as was described in Section 13.2, Fig. 13.3, and some examples are presented in Figs. 13.4–13.6. However, for beams (13.6)

and (13.9), the approximate analytical description of the diffracted beam structure is possible, which is considered in Section 13.4.

13.3.3. Migration of Singularities in the Diffracted OV Beams: Experimental Data Compared with Theory

The main results relating the OV positions and their behavior with varying incident beam screening (a -dependent evolution) are illustrated in Figs. 13.4–13.6. Each image shows the a -dependent migration of the OV cores over the fixed cross section of the diffracted beam for the incident Kummer beam with parameters (13.3) and $z_h = 11$ cm. In all cases, the transverse energy circulation is clockwise (see Fig. 13.3 (b)), and the OV cores describe spiral-like trajectories with opposite, counter-clockwise motion as is indicated by arrows. In compliance with the previous reports [7, 8, 19], as the screening grows (a diminishes to zero, and further down to negative values), the OVs move into the shadow region $x > a$ and eventually vanish (annihilate).

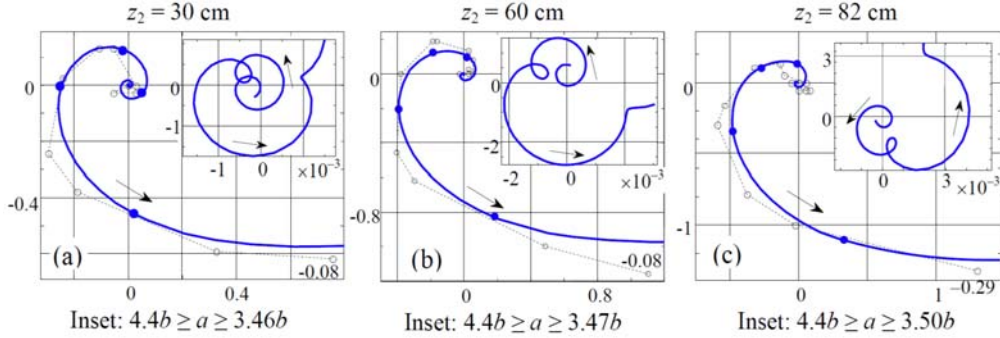


Fig. 13.4. Theoretical (solid) and experimental (dashed) trajectories of the OV cores within the cross sections of the diffracted Kummer beams with topological charge $m = -1$ (handedness of the energy circulation is as in Fig. 13.3 (b)); the cross section distance behind the obstacle is indicated above the images. All the transverse coordinates and the screen edge position a are given in units of the initial Gaussian beam radius b (13.3) at the VG plane. Arrows show the OV motion when the screen edge approaches the z -axis ($4.4b \geq a \geq 0.25b$, see Fig. 13.1 (a)). Filled circles mark the trajectories' points corresponding to decreasing half-integer $a/b = 2.0, 1.5, 1.0, 0.5$. Final positions of the screen edge where the experimental OVs apparently disappear (cannot be reliably identified) are indicated near the trajectories' ends. The insets show initial segments of the trajectories (for a within the ranges indicated below the images) that cannot be resolved in the main curves.

The detailed analysis and interpretation of these results will be performed in further sections of this Chapter; now we only comment on some general qualitative features of the OV behavior. First to note, Figs. 13.5 and 13.6 confirm earlier observations [14, 15, 19] that even very weak diffraction perturbation, when the screen edge merely 'touches' the far beam periphery and the diffraction practically does not affect the visible beam profile (which obviously takes place, for example, when $a > 4b$, cf. also Fig. 13.14 (a) below), is sufficient for the decomposition of an initial m -charged OV into a set of $|m|$ single-charged ones – the singular skeleton is really formed. Each singularity regularly

evolves with the diffracted beam propagation (growth of z) forming the OV filament [6]. Different individual filaments are marked by different colors: blue and red in case $|m| = 2$ (Fig. 13.5), and blue, red and green for $|m| = 3$ (Fig. 13.6).

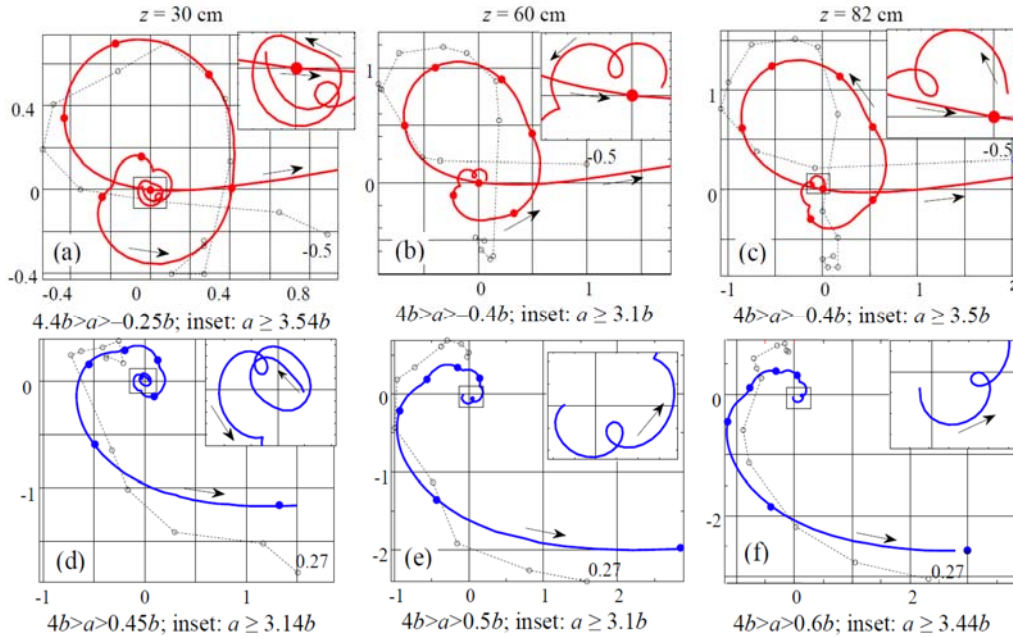
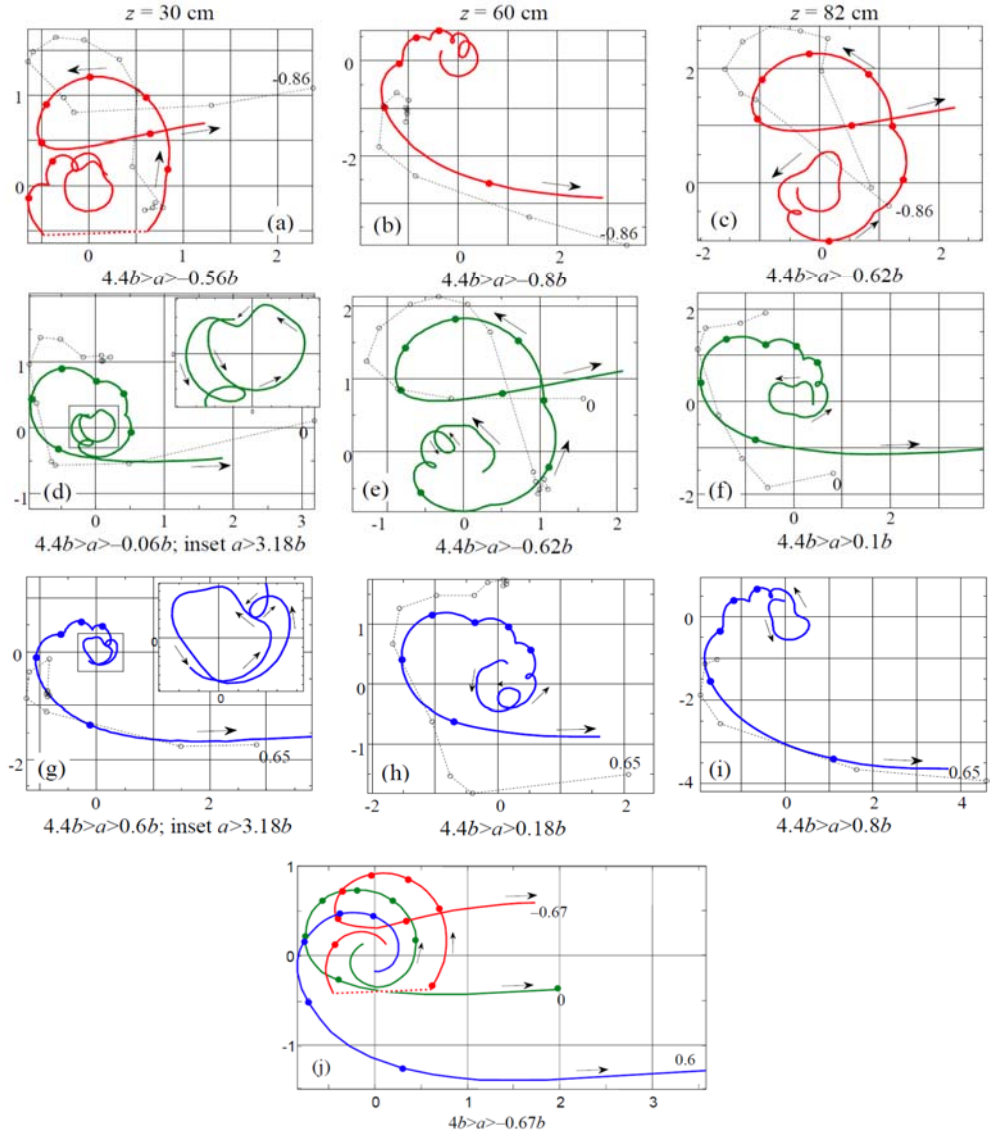


Fig. 13.5. Theoretical (solid colored) and experimental (dashed black) trajectories of the OV cores within the cross sections of the diffracted Kummer beams with topological charge $m = -2$ (handedness of the energy circulation is as in Fig. 13.3 (b)); the cross section distance behind the obstacle is indicated above each column. All the transverse coordinates and the screen edge position a are given in units of the initial Gaussian beam radius b (13.3) at the VG plane. Arrows show the OV motion when the screen edge approaches the z -axis (see Fig. 13.1 (a)); the range of the screen edge positions a accepted for calculations is indicated below each image. Filled circles mark the trajectories' points at which a/b accepts decreasing half-integer values starting from $a/b = 3.0$. The initial experimental points correspond to $a/b = 4.0$, final positions of the screen edge where the OVs could not be reliably identified are indicated near the final experimental points. The insets show magnified initial segments of the trajectories (marked by the rectangles), corresponding ranges of a are indicated below the images.

Neither in experiment nor by the numerical study could we detect the initial 'moment' (the value of a) at which the central OV of the multicharged incident beam starts to 'split'; likewise, we could not trace the earliest stages of the single-charged OV displacement from the axis z in case of the first-order beam diffraction (Fig. 13.4). With growing a (the screening weakens), the spirals of the OV trajectories seem to make infinite number of rotations around the axis z , gradually approaching the origin $x = y = 0$. For this reason, the initial points of all the trajectories correspond to certain finite OV displacement, after which its behavior becomes more regular and can be reliably traced both numerically and in the experiment.



Arrows show the OV motion when the screen edge approaches the z -axis (see Fig. 13.1 (a)); the range of the screen edge positions a assumed for calculations is indicated below each image. Filled circles mark the trajectories' points at which a/b accepts decreasing half-integer values starting from $a/b = 3.0$. The insets in (d) and (g) show magnified initial segments of the trajectories (marked by the rectangles), corresponding ranges of a are indicated below the images.

Fig. 13.6. (a)–(i) Theoretical (solid colored) and experimental (dashed black) trajectories of the OV cores within the cross sections of the diffracted Kummer beams, the cross section distance behind the obstacle is indicated above each column; (j) theoretical trajectories of the OV cores in the diffracted LG beam (13.9), (13.10) with parameters (13.11) for the distance behind the obstacle $z = 30$ cm. The topological charge of the incident OV is $m = -3$, all the transverse coordinates and the screen edge position a are given in units of b (13.3) or (13.11). The initial experimental points correspond to $a/b = 2.2$, final positions of the screen edge where the OVs could not be reliably identified are indicated near the final experimental points.

A good quantitative agreement between the experimental results and the simulation can be detected only for a single-charged OV (Fig. 13.4) whereas for two- or three-charged OVs just a qualitative correspondence is seen (see Figs. 13.5 and 13.6). This can be associated with the special character of the VG used in the experiment: With increasing order of the OV beams generated by the ‘fork’ hologram, the beams’ inclination with respect to the initial Gaussian beam axis also grows, which invokes the OV beam deformation [35] so that at the screen plane, the incident OV beam partly loses its symmetry, and a set of secondary OVs already exists instead of a single multicharged OV expected theoretically. Actually, Figs. 13.5 and 13.6 show that discrepancies between the experimental and theoretical trajectories grow with $|m|$.

Another reason for this is that the spiral-like OV evolution is complicated by multiple fine details distinctly seen in the theoretical curves in Figs. 13.4–13.6: the spirals experience oscillatory pulsations, sometimes resulting in trajectory self-crossings, regions of retrograde azimuthal motion (loops), etc. (further we will show that this is a characteristic feature of the Kummer beam diffraction, see Section 13.4). These fine details are the most articulate under the WDP conditions ($a \gg b$) but with decreasing a they disappear and the spirals become more regular. Presence of the fine details makes the experimental detection of the current OV positions difficult, as is the case of Figs. 13.5 and 13.6. In case of Fig. 13.4, these details only exist for very small radial deviation of the OV from the nominal beam axis ($\sim 0.001b$, see the insets in Figs. 13.4 (a–c)), and the observable part of the OV trajectory is already a regular spiral favorable for the experimental observations. Note that for $|m| = 2$ (Fig. 13.5) one of the OV trajectories always intersects the point $x = y = 0$, which confirms the earlier suggestion that when an even-charged circular OV beam diffracts on a sharp edge positioned at $a = 0$, one of the diffracted beam OVs is fixed at the incident beam axis [19].

An interesting peculiar feature is seen in Fig. 13.6 (a). For certain conditions, the OV trajectory experiences a discontinuity denoted by the dashed straight-line segment: a minute advance of the screen edge induces an articulate ‘jump’ of the OV core to a remote point of the diffracted beam cross section. This situation is rather typical and will be a subject of special consideration in Sections 13.6 and 13.7.

For comparison, the OV migration in the diffracted LG beam (13.9) is illustrated in Fig. 13.6 (j) for conditions corresponding to Figs. 13.6 (a, d, g) ($z = 30$ cm, $m = -3$). Numerical values of the LG beam parameters are chosen so that the transverse profile of the incident LG beam at the screen plane be in maximal possible similarity to the Kummer beam spatial profile of [20, 21]. This non-rigorous requirement is satisfied if we accept in Eqs. (13.9) and (13.10)

$$b_c = b = 0.232 \text{ mm}, R_c = 57 \text{ cm}, b_0 = 0.17 \text{ mm}, z_{Rc} = 28.5 \text{ cm}, z_c = 27 \text{ cm}. \quad (13.11)$$

The general pattern of the OV trajectories in Fig. 13.6 (j) looks qualitatively similar to that of the Kummer beam diffraction; even the discontinuity of the ‘red’ trajectory is confirmed. At the same time, the OV trajectories in the diffracted LG beam are much smoother and contain no radial pulsations, nor small-scale self-crossings (‘loops’).

13.4. Mathematical Model of the Singular Skeleton Evolution in Diffracted OV Beams

13.4.1. Asymptotic Analytical Model

If the incident beam is an LG beam, the integral (13.4) can, in principle, be evaluated analytically but when $|m| > 1$, the explicit representation is cumbersome and physically obscure [11, 12]; for the incident Kummer beams, an exact analytical representation is unknown. Nevertheless, the situation can be examined analytically by means of the simple model which is derived for $a \gg b$ (WDP case) but appears to be practically valid when the screen edge is separated by several b from the incident beam axis [21] (see Fig. 13.1). In this approximation, the diffracted beam (13.4) can be considered as a superposition of the unperturbed incident beam and the edge wave ‘emitted’ by the screen edge [2]. For any circular OV beam considered in this Chapter, near the z -axis its complex amplitude distribution can be presented in the form

$$E_{\text{inc}} = B_0 \left(\frac{r}{b} \right)^{|m|} \exp(im\phi) \exp(ikz), \quad (13.12)$$

where $r = \sqrt{x^2 + y^2}$ and $\phi = \arctan(y/x)$ are the polar coordinates in the observation plane. The quantity B_0 is a complex constant depending on the propagation distance and the beam type (e.g., Kummer or LG), as well as on its specific parameters, which can easily be derived from the explicit expressions (13.6) or (13.9). Near the origin of the observation plane, the edge-wave amplitude approximately amounts to

$$\begin{aligned} E_{\text{edge}} &= D_0(a, z) \exp \left[ik \left(z + \frac{a^2}{2z} - x \frac{a}{z} \right) \right] \\ &= D_0(a, z) \exp \left[ik \left(z + \frac{a^2}{2z} - r \frac{a}{z} \cos \phi \right) \right], \end{aligned} \quad (13.13)$$

with the complex coefficient $D_0(a, z)$ decreasing with growing $|a|$ and z . Eq. (13.13) differs from the similar expression used in Ref. [21] by the x -proportional term responsible for the wavefront inclination in the (xz) plane (see Fig. 13.1). The positions of the OV cores are determined by the condition $E_{\text{edge}} + E_{\text{inc}} = 0$, which entails

$$\frac{r}{b} = \left[\frac{|D_0(a, z)|}{|B_0|} \right]^{1/|m|}, \quad (13.14)$$

$$\phi + M \cos \phi = C_{N+1} + k \frac{a^2}{2mz}, \quad (13.15)$$

where

$$M = \frac{kra}{mz}, \quad (13.16)$$

and the coordinate-independent term C_{N+1} possesses its own value for each secondary OV numbered by $N=0, 1, \dots, |m|-1$,

$$C_{N+1} = \frac{1}{m} [\arg D_0(a, z) - \arg B_0 + (2N-1)\pi]. \quad (13.17)$$

Despite their very approximate character, Eqs. (13.14) and (13.15) enable efficient qualitative analysis of the OV trajectories. First, one observes that under the WDP conditions, the OV off-axis displacement $r \rightarrow 0$, and the second summand in the left-hand side of (13.15) can be neglected ($M \rightarrow 0$). Then Eq. (13.15), in full agreement with the experiment [20], predicts the monotonous behavior of the OV azimuth upon monotonous variation of a or z . Together with the monotonous nature of $D_0(a, z)$ in Eq. (13.14), this dictates the spiral character of the OV trajectory, which (theoretically) makes infinite number of rotations near the z -axis while $a \rightarrow \infty$ (the focus-type singularity [16]). Also, Eq. (13.15) with $M \rightarrow 0$ makes it obvious that the rate of the OV spiral evolution should slow down with a decrease of a and an increase of z , which is also confirmed by experiments and numerical calculations [20, 21].

13.4.2. Refined Analytical Model

This procedure can be substantiated and refined by employing the asymptotic representation of the diffracted beam field [21]. To this purpose, we resort to the asymptotic expression for the diffracted beam amplitude (13.4) derived in Appendix:

$$u(x, y, z) \approx B_1 \left(\frac{r}{b} \right)^{|m|} \exp(im\phi) - \frac{D_1}{a^3} \exp \left[\frac{ika^2}{2} \left(\frac{1}{z_h} + \frac{1}{z} \right) - \frac{ika}{z} x \right] + \frac{D_2}{a} \exp \left[\frac{ika^2}{2} \left(\frac{1}{z_d} + \frac{1}{z} \right) - \frac{ika}{z} x \right], \quad (13.18)$$

where z_d is defined in Eq. (13.A17),

$$B_1 = \sqrt{\frac{\pi}{2^{3|m|}}} \left(\frac{-iz_e}{z + z_h} \right)^{|m|+1} \frac{z_R^{|m|+1}}{z_e^{|m|/2} (z_e - iz_R)^{|m|/2+1}}, \quad (13.19)$$

$$D_1 = |m|(-i)^{|m|+1} \sqrt{\frac{i}{2\pi}} \frac{z_h}{z} \left[k \left(\frac{1}{z_h} + \frac{1}{z} \right) \right]^{-3/2}, \quad (13.20)$$

$$D_2 = -\frac{ik}{2^{2|m|}} \sqrt{\frac{i}{2\pi}} \frac{z_{he}}{z z_h} \frac{z_R^{2|m|+1}}{z_{he}^{2|m|} (z_{he} - iz_R)^{|m|+1}} \left[k \left(\frac{1}{z_d} + \frac{1}{z} \right) \right]^{-3/2}. \quad (13.21)$$

Equations (13.18)–(13.21) are valid if $a \gg b$ and only near the axis, $x \ll b$, $y \ll b$ (see Eq. (13.A4)) but they express the physical essence of the diffraction process [2]. The first term of Eq. (13.18) describes the unperturbed incident beam and is an analogue of Eq. (13.12), while the role of the screen is accumulated in the second and third terms that describe a sort of ‘edge wave’ (13.13) which interferes with the incident beam, thereby forming the diffraction pattern.

As a result, for the diffraction of the Kummer beam (13.6)–(13.8), instead of the simple relations (13.14), (13.15), the OV polar coordinates can be determined via equations

$$\frac{r}{b} = \left[\left| \frac{D_1}{a^3} \exp\left(\frac{ika^2}{2z_h}\right) - \frac{D_2}{a} \exp\left(\frac{ika^2}{2z_d}\right) \right| \cdot \frac{1}{|B_1|} \right]^{1/|m|}, \quad (13.22)$$

$$\phi + M \cos \phi = \frac{1}{m} \arg \left[\frac{D_1}{a^3} \exp\left(\frac{ika^2}{2z_h}\right) - \frac{D_2}{a} \exp\left(\frac{ika^2}{2z_d}\right) \right] + \frac{ka^2}{2mz} - \frac{1}{m} \arg B_1 + \frac{2N}{m} \pi, \quad (13.23)$$

where M is defined by (13.16).

Quite similarly, based on Eqs. (13.A1), (13.A8) and (13.A9), the analogs of Eqs. (13.18)–(13.21) for the LG beam (13.9), (13.10) for large enough $a \gg b_c$ can be derived:

$$u(x, y, z) \simeq B r^{|m|} \exp(im\phi) - D a^{|m|-1} \exp \left[\frac{ika^2}{2} \left(\frac{1}{z_c - iz_{Rc}} + \frac{1}{z} \right) \right], \quad (13.24)$$

where

$$B = \frac{1}{(z + z_c - iz_{Rc})^{|m|+1}}, \quad D = \sqrt{\frac{i}{2\pi}} \frac{k}{z} (z_c - iz_{Rc})^{-|m|-1} \left[k \left(\frac{1}{z} + \frac{1}{z_c - iz_{Rc}} \right) \right]^{-3/2}. \quad (13.25)$$

Hence, again, the explicit expressions for the OV cores’ polar coordinates can be easily found by equating (13.24) to zero:

$$r = \left[a^{|m|-1} \exp \left(-\frac{a^2}{2b_c^2} \right) \left| \frac{D}{B} \right| \right]^{1/|m|}, \quad (13.26)$$

$$\phi + M \cos \phi = \frac{1}{m} [\arg D - \arg B] + \frac{ka^2}{2mz} + \frac{ka^2}{2mR_c} + \frac{2N}{m} \pi. \quad (13.27)$$

The quality and descriptive abilities of the described asymptotic model are illustrated in Fig. 13.7 where, for simplicity, the OV azimuths are calculated discarding the cosine terms in Eqs. (13.23) and (13.27). The most noticeable model limitation is that it describes all the secondary OVs by the same equations differing only by N thus predicting that their radial coordinates are identical and the azimuthal ones differ only by additive constants (in the case of Fig. 13.7 with $|m|=3$ this additive constant defining the azimuthal ‘distance’ between different OV positions is $2\pi/3$). Nevertheless, one can see that the approximate solutions (13.22), (13.23) and (13.26), (13.27) fairly reflect main qualitative features of the a -dependent OV evolution in the diffracted beam cross sections: the dashed curve in Fig. 13.7 (a) explicitly shows the radial pulsations of the spiral trajectory (cf. Figs. 13.5 and 13.6); moreover, the dashed curves in Fig. 13.7 (b) contain segments where ϕ decreases with growing a explaining the retrograde evolution of the spirals and formation of the loops well seen in almost all images of Figs. 13.4–13.6. In fact, Eqs. (13.22) and (13.23), to a certain degree, disclose the physical reason for the radial pulsations and loops that are characteristic for all images of Figs. 13.5 and 13.6 except Fig. 13.6 (j): these are consequences of the interference between the two summands in brackets of the right-hand sides of these equations. In turn, these interfering terms appear due to the oscillatory behavior of the Kummer beam amplitude (‘ripple structure’) [34] and the slow amplitude decay for large transverse radius $r_a \rightarrow \infty$, so that the edge wave represented by the second line of Eq. (13.18) is formed as a superposition of contributions originating from different points of the screen edge with different amplitudes and initial phases that non-monotonously depend on a and/or z . With increasing z_h in Fig. 13.1 (c) and Eqs. (13.6)–(13.8), oscillations of the amplitude and phase at the Kummer beams’ periphery are gradually mitigated, and the pulsations of the OV trajectories in corresponding diffracted beams are softened and eventually disappear (this is not shown in figures but was confirmed by numerical calculations). The same reasoning perfectly explains the ‘smooth’ OV trajectories in the diffracted LG beams (Figs. 13.6 (j), 13.7 (c), 13.7 (d)) where the amplitude decays monotonously and very rapidly for $r_a \rightarrow \infty$.

Thus providing a good qualitative basis of the OV trajectories’ behavior for diffracted Kummer beams, the asymptotic model gives but a rather poor quantitative agreement with accurate numerical data, and even this poor agreement is only observed at large enough $a \gtrsim 3b$. In contrast, upon diffraction of LG beams, the good quantitative approximation for at least one of the OVs is realized up to $a \approx b$ (Fig. 13.7 (c)). This is rather surprising in view of the asymptotic character of the model that was derived assuming $a \gg b$. As to the azimuthal OV positions (Fig. 13.7 (d)), the main peculiarity of the model (dashed) curves is their perfectly smooth character in contrast to the accurate numerical data that demonstrate segments of relatively slow and rapid growth of ϕ . The ‘steep’ segments of the solid curves near the points where $\phi = -\pi/2, 3\pi/2, 7\pi/2$ (clearly seen in Fig. 13.7 (d) but masked by additional details in Fig. 13.7 (b)) as well as the jumps of the red curves near $a/b = 2.35$ in Fig. 13.7 (b) and $a/b = 2.6$ (corresponding to the discontinuities in Figs. 13.6 (a) and 13.6 (j)) in Fig. 13.7 (d) are associated with the discarded cosine terms in (13.23) and (13.27). The nature of these peculiarities will be discussed below in Section 13.6.

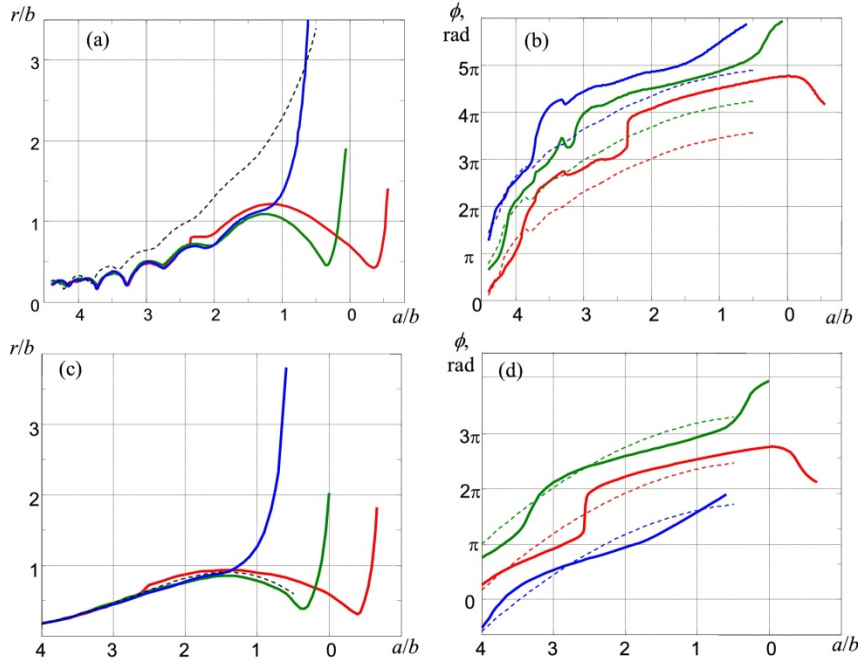


Fig. 13.7. Radial and angular polar coordinates of the OV cores in the diffracted OV beam with topological charge $m = -3$, cross section $z = 30$ cm. (a, b) Incident Kummer beam; (c, d) incident LG beam. Solid curves illustrate the numerically calculated evolution of r and ϕ when the screen edge moves from large positive x towards the beam axis and slightly further; inverse horizontal scale indicates the screen edge position a in units of b (13.2) or (13.11). Curves of different colors describe different secondary OVs (curves of the same colors in (a, b) and in Fig. 13.6 (b, e, i) describe the same OVs); thin dashed curves represent the asymptotic approximation of (a, b) (13.22), (13.23) and (c, d) (13.26), (13.27) with $M = 0$.

13.5. Theoretical Study of the OV Migration: LG Beams

The first experimental observations that have stimulated this research and have been described in Section 2 dealt with the Kummer beams. However, some general features of OV diffraction can more easily be understood with the help of simpler examples of LG incident beams (Fig. 13.1 (b)). In this section we analyze the OV localization and migration in case of diffraction of the simplest low-order OV beams, mostly under the WDP conditions which enable efficient and insightful interpretation based on the asymptotic analytical model described above. In particular, in this case the OV displacement from the z -axis is supposed to be small, due to which $M \ll 1$, and the cosine terms in the left-hand sides of Eqs. (13.23), (13.27) can be discarded. The WDP requirements impose natural limitations on the range of possible variations of the screen edge position a ; that is why in this section we mainly consider the z -dependent evolution of the diffracted beams for several fixed a . As everywhere in this Chapter, the transverse geometrical parameters are expressed in units of the incident beam transverse scale. For LG beams, this scale is represented by b_c (13.10). In order to preserve geometrical

conditions of Section 13.2 and for determinacy of numerical examples, we chose b_c to be equal to the Kummer beam transverse scale (13.3), i.e. in all further calculations we suppose

$$k = 10^5 \text{ cm}^{-1} \text{ and } b_c = 0.232 \text{ mm.} \quad (13.28)$$

13.5.1. OV Displacements: Incident Beam with Plane Wavefront

Following [14, 15, 19], we start with analyzing the single-charged incident LG beam ($m = -1$) with its waist at the screen plane, i.e. in equations (13.9) and (13.10)

$$z_c = 0, \quad b_c = b_0, \quad R_c = \infty, \quad (13.29)$$

which means

$$z_{Rc} = 53.8 \text{ cm.} \quad (13.30)$$

For a single-charged OV beam, the main consequence of the WDP is the OV displacement from its nominal position at the beam axis $x = 0, y = 0$. The main results relating the positions of the OV cores in the diffracted beam cross-section at different distances z behind the screen plane are given in Fig. 13.8 for four fixed screen-edge locations (see Fig. 13.1 (a)): $a = 3b_c$, $a = 2b_c$, $a = b_c$, and $a = 0.6b_c$. In order to divert from the trivial component of the OV migration associated with the beam divergence, the transverse OV coordinates are normalized by the current beam radius

$$b_d = b_0 \sqrt{1 + \frac{z^2}{z_{Rc}^2}}. \quad (13.31)$$

As expected, the distinct spiral-like trajectories are only observed under conditions of weak OV-beam perturbation ($a = 3b_c$ and $a = 2b_c$); the trajectories for $a = b_c$ and $a = 0.6b_c$ are given for comparison. Actually, these curves also contain certain spiral-like segments corresponding to the very small propagation distances, which are hardly available both to analysis and to observation because of the diffraction-fringes effects. On the other hand, under the WDP conditions, the OV displacement from the incident beam axis is rather small. This is noticeable, e.g., in Fig. 13.4 for the a -dependent evolution in the diffracted Kummer beam and is clearly seen in Fig. 13.8 (a) by matching the blue spiral for $a = 2b_c$ against the red and black curves for $a = b_c$, and $a = 0.6b_c$; in case of $a = 3b_c$ the spiral is so small that it is separately magnified in Fig. 13.8(b).

The images of Fig. 13.8 are actually transverse projections of the 3D OV filaments that evolve within the “body” of the propagating diffracted beam. Such 3D spirals for $a = 2b_c$ and $a = 3b_c$ are illustrated by Fig. 13.9. Here the longitudinal coordinate z varies within the range

$$10 \text{ cm} = 0.186z_{Rc} < z < 600 \text{ cm} \approx 11.15z_{Rc}.$$

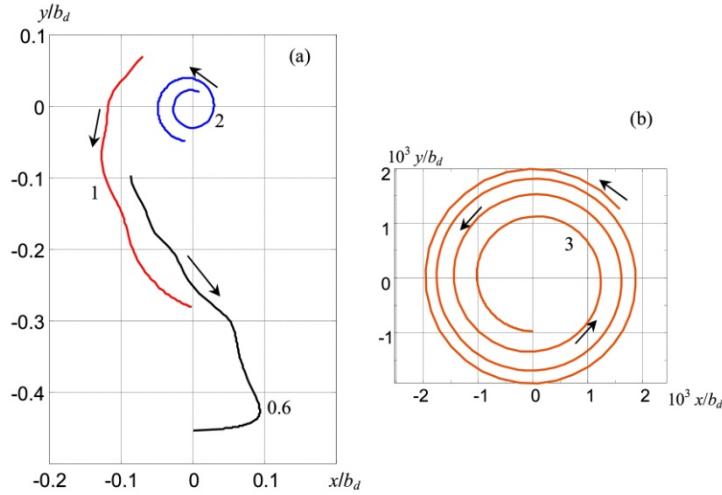


Fig. 13.8. z -Dependent evolution of the OV core position in the cross section of the diffracted LG beam of equations (13.9), (13.10), (13.29) with $m = -1$ and $R_c = \infty$ for (a) $a = 2b_c$ (blue), $a = b_c$ (red), $a = 0.6b_c$ (black) and (b) $a = 3b_c$ (brown); each curve is marked by the corresponding value of a in units of b_c . Arrows indicate directions of the OV motion, initial points of the curves correspond to $z = 10 \text{ cm} = 0.186z_{Rc}$ (near field), and final points correspond to $z = \infty$ (far field).

Note that the ‘motion’ of the OV core along its trajectory is not uniform: practically the entire observable evolution happens for the first 10 % of the full range of z variation (cf. Fig. 13.10 below). To present the 3D trajectories more conveniently, in Figs. 13.9 (a), 13.9 (b) the longitudinal scales are non-uniform: the real distance z is normalized by the scale factor (13.31). Remarkably, even with this precaution, the region $z < 10 \text{ cm}$ cannot be shown properly because with decreasing z , the rate of the spiral rotation (theoretically) infinitely grows while the spiral pitch (distance between consecutive coils) infinitely decreases. This is supported by the simplified asymptotic model (13.12)–(13.15) and (13.24)–(13.27) which at WDP conditions ($M \ll 1$) leads to the approximate rule

$$\phi \approx \text{const} + k \frac{a^2}{2mz}. \quad (13.32)$$

Obviously, with increasing z , the rotation practically stops whereas for small z the rate of rotation $d\phi/dz$ can be rather high. However, this implies no unphysical divergence for very small z because when $z \rightarrow 0$, the relation $D \rightarrow 0$ also takes place (see (13.25)) and Eqs. (13.14) and (13.24) dictate that with growing rate of rotation, the OV off-axial displacement becomes so small that the “theoretic” spiral motion is practically imperceptible.

Importantly, for $z \rightarrow \infty$ all the OV trajectories in Fig. 13.8 approach the vertical axis. This fact was noticed previously and interpreted based on the general concept that the diffracted beam intensity distribution rotates during propagation [9, 13, 19] (cf. Fig. 13.2). Just after the screen, the diffracted beam intensity distribution loses circular symmetry but preserves

the mirror symmetry with respect to an axis orthogonal to the screen edge. In the course of further propagation, the beam shape continuously changes, generally, in a rather complicated way. But this transformation possesses a sort of regular rotational component, which, in the far field, results in an intensity distribution with another symmetry axis, orthogonal to the initial one and parallel to the screen edge. Accordingly, in the far field, the OV of the diffracted beam eventually approach this symmetry axis that in normalized coordinates of Fig. 13.8 coincides with the y -axis.

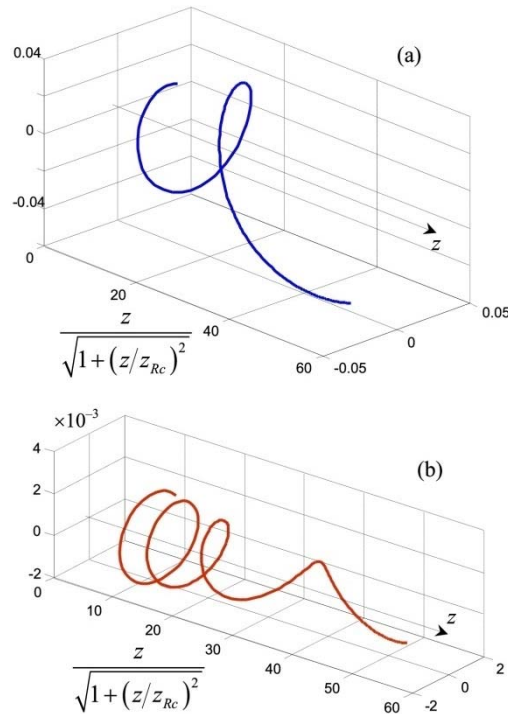


Fig. 13.9. 3D evolution of the OV core positions in the diffracted LG beam (13.9), (13.10), (13.29) with $m=-1$ and $R_c = \infty$ for (a) $a = 2b_c$ and (b) $a = 3b_c$; transverse coordinates are in units of b_d (13.31). The blue and brown spirals in Fig. 13.8 represent projections of these trajectories viewed against axis z .

13.5.2. OV Displacements: Incident Beam with Spherical Wavefront

The last conclusion of the previous section was drawn from the OV diffraction analysis based on an assumption that the wavefront of the incident OV beam is plane. Now consider what influence can be caused by the wavefront curvature. Of course, this can be carried out by calculating the diffraction integral (13.4) with the input amplitude distribution (13.9), (13.10) corresponding to a non-planar wavefront, or with the help of the model equations (13.25)–(13.27) with the necessary value of R_c but it would be suitable to involve another simple rule for transformation of the diffracted beam complex amplitude that occurs if the incident beam is modified according to the equation

$$u(x_a, y_a) \rightarrow u(x_a, y_a) \exp\left(ik \frac{x_a^2 + y_a^2}{2R_c}\right), \quad (13.33)$$

(e.g., the plane front is replaced by the spherical one with the same intensity profile) [14]. In this situation, if the initial distribution $u(x_a, y_a)$ produces a diffracted beam with complex amplitude $u(x, y, z)$, the modified initial beam (13.33) produces the diffracted beam distribution

$$u_e(x, y, z) = \frac{1}{1 + \frac{z}{R_c}} \exp\left[ik \frac{x^2 + y^2}{2(z + R_c)}\right] u\left(\frac{x}{1 + z/R_c}, \frac{y}{1 + z/R_c}, \frac{z}{1 + z/R_c}\right). \quad (13.34)$$

To elucidate the meaning of this rule, let us suppose that the incident beam $u(x_a, y_a)$ possesses a plane wavefront, except for the helical component associated with the term $(x_a + i\sigma y_a)^{|m|}$ in Eq. (13.9). Also, we suppose that this beam ('prototype beam'), being diffracted, produces in the cross section z_0 behind the screen such complex amplitude distribution, for which the OV core is situated in the point $(x = x_0, y = y_0)$. Hence, Eq. (13.34) dictates that the modified incident beam (13.33) produces the diffracted beam in whose cross section

$$z = \frac{z_0}{1 - z_0/R_c}, \quad (13.35)$$

the OV core is located in the point

$$x_r = \frac{x_0}{1 - z_0/R_c}, \quad y_r = \frac{y_0}{1 - z_0/R_c}. \quad (13.36)$$

In particular, the far field of the modified spherical-front beam (13.33) is realized when $z \rightarrow \infty$ which corresponds to the finite cross section of the prototype beam, $z_0 = R_c$. Note that transformation (13.36) affects only the off-axial distance of the OV position while its azimuth $\phi_r = \arctan(y_0/x_0) = \arctan(y_r/x_r)$ remains the same as in the prototype plane-front beam.

This reasoning suggests a simple procedure for determining the azimuthal far-field positions of the OV cores in the diffracted LG beam with non-planar wavefront. It requires knowledge of the OV trajectory for the prototype plane-front beam (see Fig. 13.10 where the blue spiral of Fig. 13.8 (a) is magnified and furnished with marks denoting the propagation distances behind the screen in centimeters). For example, when the incident beam wavefront possesses the curvature radius $R_c = 80$ cm, Eq. (13.35) shows that the far field for the diffracted beam is realized at $z_0 = R_c = 80$ cm. Accordingly, the corresponding

OV azimuth coincides with the azimuth of the red rectangle marked “80” in Fig. 13.10. And indeed, the independently calculated OV trajectory for the incident beam (13.9) with $R_c = 80$ cm (green curve) is oriented close to this direction, and the small discrepancy can be explained by the limited range of the propagation distances accepted for the green curve calculation (in fact, the ‘genuine’ far field with infinite propagation distance was never reached). The same is correct for other examples corresponding to $z_0 = R_c = 18$ cm (red curve), 30 cm (brown curve) and infinity (black curve). This reasoning distinctly shows that the far-field OV position belongs to the symmetry axis parallel to the screen edge only if the incident LG beam possesses a plane wavefront.

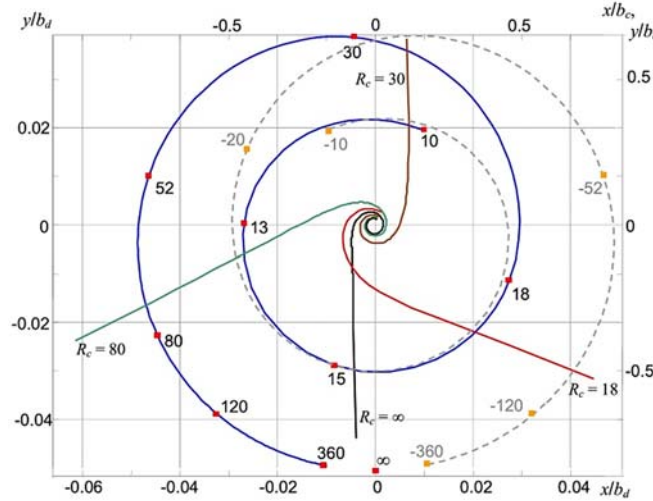


Fig. 13.10. Illustration of the far-field azimuthal OV positions in diffracted beams with non-planar wavefront for $m = -1$ and $a = 2b_c$. Blue curve is the transverse projection of the OV trajectory in the prototype plane-front diffracted beam (cf. Fig. 13.8 (a), blue curve, and Fig. 13.9 (a)), pale grey dotted curve represents its branch for the negative propagation distances. The left and bottom scales show the current OV coordinates in normalized units (13.31), red (yellow) rectangles denote the propagation distance marked in centimeters. Red, brown, green and black curves are the transverse projections of the OV trajectories in the modified spherical-front diffracted beams (13.33) (radius of curvature is indicated near each curve, the OV coordinates are measured in units of b_c , and marked in the right and top scales).

Further application of this procedure to cases with $R_c < 0$ (converging beams) requires knowledge of the plane-front diffracted beam behavior at $z_0 < 0$, which seems non-physical. However, the Fresnel–Kirchhoff integral (13.4) formally is valid for any z ; moreover, equations (13.4), (13.9) and (13.10) suggest that

$$u^*(x_0, y_0, -z_0) = u(x_0, -y_0, z_0),$$

where the asterisk denotes complex conjugation. This means that positions for the amplitude zeros at negative z_0 can be easily found once they are known for positive z_0 . Additionally, to get the far-field OV positions, we should take into account that in

Eq. (13.35) z tends to positive infinity, and for negative z_0 this implies that transformation (13.36) inverts the signs of both transverse coordinates. As a result, the ‘prototype beam’ OV trajectory for negative z_0 can be obtained by a mirror reflection of the blue spiral described above, and is presented in Fig. 13.10 as a pale grey dashed curve. This can be employed exactly in the same manner as the blue curve itself, with the help of corresponding distance marks, some of which are explicitly indicated.

Whereas for diverging beams (positive R_c), the symmetrical structure of the intensity distribution, typical for the plane-front beams, does not exist in the ‘physical’ range $z > 0$ (the far-field intensity pattern reproduces the prototype beam structure at finite z_0), for converging beams with $R_c < 0$ such a possibility is realized at finite z . According to Eqs. (13.35) and (13.36), this occurs when $z = -R_c > 0$, which corresponds to $z_0 = \infty$. It is expectable because the plane $z = -R_c > 0$ is actually a focal plane of the ‘lens’ performing transformation (13.33), and in this plane the Fraunhofer diffraction takes place which is equivalent to the far field propagation [1, 2].

The resulting azimuthal OV positions in diffracted LG beams with spherical wavefronts are illustrated in Fig. 13.11. Note that when $R_c \rightarrow \infty$, all the curves approach the azimuth values $\phi_r = (3/2)\pi$, $(7/2)\pi$ and $(15/2)\pi$, which means that the OV trajectories always end at the negative half-axis y , as was discussed above. This example also shows the advantages of the WDP conditions. Theoretically, in cases of strong beam screening, the wavefront curvature also affects the OV positions in the diffracted beam cross section but the red and black curves in Fig. 13.11 display a rather limited range of possible azimuthal deviations, and their sensitivity to the incident wavefront curvature appears to be low.

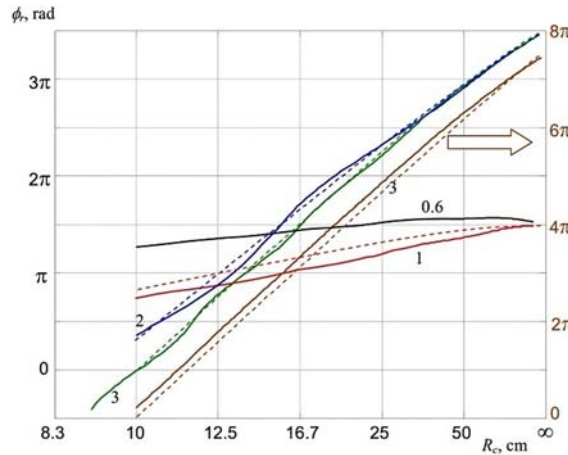


Fig. 13.11. Far-field azimuthal OV positions in the diffracted LG beam (13.9) with $m = -1$ vs. the wavefront curvature radius: (brown curve) $a = 3b_c$, right vertical scale; (blue curve) $a = 2b_c$ (cf. Fig. 13.10); (red curve) $a = b_c$ (cf. Fig. 13.8 (a)); (black curve) $a = 0.6b_c$ (cf. Fig. 13.8 (a)). The green curve represents the far-field orientation of the straight line connecting two secondary OVs in the diffracted LG beam (13.9) with $m = -2$, $a = 3b_c$ (cf. Fig. 13.13, dashed line). Each curve is marked by the corresponding value of a in units b_c ; dashed lines illustrate the analytical approximations (13.37) for the solid curves of the same colors.

All these conclusions are based on the numerical simulations but they are also justified by the approximate model of Eqs. (13.26), (13.27). First, we note that if we introduce a finite R_c by means of transformation (13.33), the set of parameters (13.28)–(13.30) becomes modified: b_c remains the same as in Eq. (13.28) but it is no longer equal to the new waist radius b_0 , and z_c is determined by the new waist position. Accordingly, the Rayleigh length of the modified LG beam $z_{Rc} = kb_0^2$ differs from the value of Eq. (13.30) but Eqs. (13.10) entail the relation

$$\frac{z_c}{z_{Rc}} = \frac{kb_c^2}{R_c}.$$

Then, with the help of Eqs. (13.25) and (13.27) with negligible M one can easily derive the far-field ($z \rightarrow \infty$) representation of the OV azimuth:

$$\phi_\infty = \frac{|m|}{m} \left(\frac{\pi}{2} - \arctan \frac{kb_c^2}{R_c} \right) + \frac{ka^2}{2mR_c} + \frac{1}{2m} \arctan \frac{kb_c^2}{R_c} + \frac{2N}{m} \pi. \quad (13.37)$$

The corresponding dependences $\phi_\infty(R_c)$ for $m = -1$ ($N = 0$) are presented in Fig. 13.11 by the brown, blue and red dashed curves (since the azimuth values differing by a complete angle are equivalent, upon constructing the dashed curves necessary integer numbers of 2π were added to the immediate results of (13.37)). One can see that expression (13.37) provides a rather good approximation for the precise numerical data, especially for large R_c and $a = 2b_c$ and $a = 3b_c$; when the screening grows (e.g., for $a = b_c$) and the WDP conditions are violated, the approximation loses its quantitative accuracy but remains qualitatively valid.

13.5.3. Incident LG Beam with the Second-Order OV

A similar WDP-induced behavior is typical for higher-order OV beams. As an example, we consider the diffraction of a LG beam (13.9) with $m = -2$ and the same parameters (13.28)–(13.30). The main difference from the charge-1 case is that diffraction makes the incident OV decompose into two single-charged secondary OVs which evolve separately within the diffracted beam ‘body’. At the WDP conditions, they form a ‘double spiral’ (Fig. 13.12), each component being quite similar to the OV trajectories observed in the diffracted beam with $|m| = 1$ (Fig. 13.9). Transverse projections of the trajectories shown in Fig. 13.12 are presented in more detail in Fig. 13.13; similarly to Fig. 13.10, the distance marks are added denoting the current longitudinal positions. Just as in Fig. 13.10, the distance marks can be used for prediction of the azimuthal orientation of the far-field OV displacement if the incident beam wavefront is not plane (see the black, green and cyan curves for $R_c = 50$ cm, $R_c = 85$ cm and $R_c = \infty$, respectively).

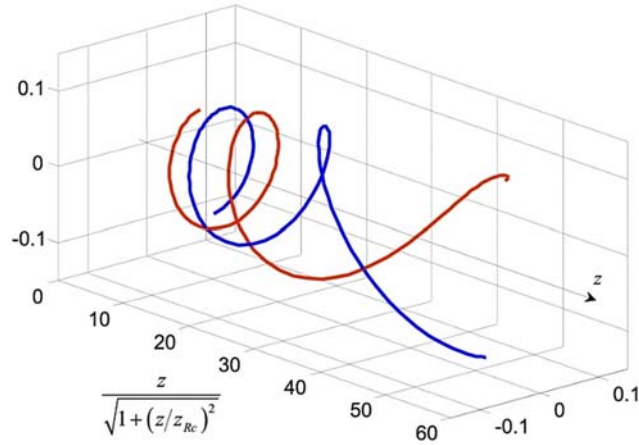


Fig. 13.12. 3D evolution of the OV core positions in the diffracted LG beam (13.9), (13.29) with $m = -2$ and $R_c = \infty$ for $a = 3b_c$. As in Fig. 13.9, the transverse coordinates are in units of b_d (13.31), and the longitudinal coordinate is normalized by the scale factor of (13.31).

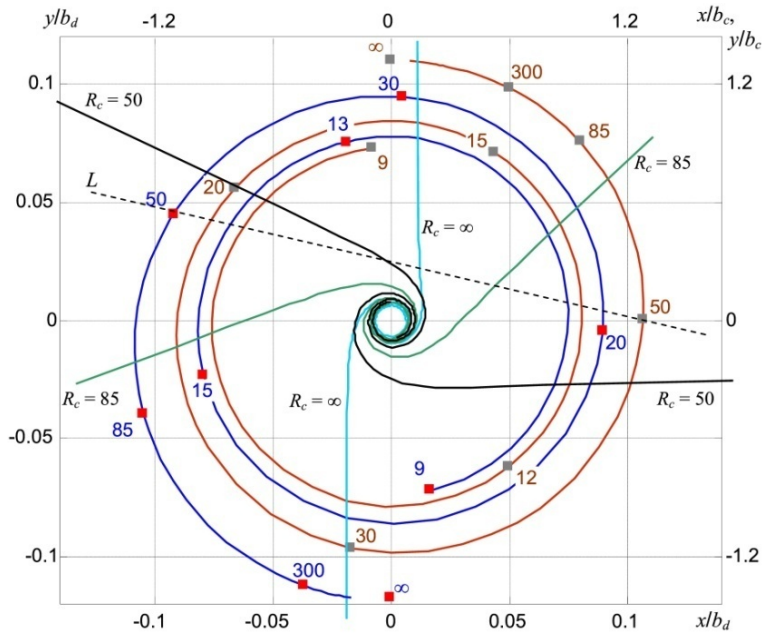


Fig. 13.13. Blue and brown curves represent trajectories of the two secondary OVs formed in the cross section of the diffracted LG beam with $m = -2$ and $a = 3b_c$ (cf. Fig. 13.12), red (grey) rectangles denote the propagation distance (longitudinal coordinate) marked in centimeters, left and bottom scales show the transverse OV coordinates in normalized units (13.31). Black, green and cyan curves are the transverse projections of the secondary-OV trajectories in the modified spherical-front diffracted beams (13.33) (radius of curvature is indicated near each curve, the OV coordinates are measured in units of b_c , and marked in the right and top scales). Dashed line L shows the far-field orientation of the OV pair at $R_c = 50$ cm.

Note that the two secondary OV's are always situated opposite with respect to the nominal beam axis z . However, they never form a perfect central-symmetric pair (with the centre at the transverse coordinate origin) dictated by the approximate model (13.27) with $M = 0$, except in the asymptotic case $z = \infty$, $R_c = \infty$; this is also a consequence of the slight symmetry perturbation upon the WDP. At the same time, the two secondary OV's define a straight line whose far-field orientation can be distinctly associated with the wavefront curvature of the incident LG beam (see, for example, the dashed line L in Fig. 13.13 that unites the two OV cores in case $R_c = 50$ cm). The rotation of this line with the incident wavefront curvature variation is illustrated by the green curve in Fig. 13.11. As in case of $|m| = 1$, its behaviour can be fairly approximated by Eq. (13.37), cf. the green dashed curve in Fig. 13.11. Its main difference from the simulation results is that it does not show the rotation accelerations near $\phi_r = \pi/2, 3\pi/2$ where one of the OV's crosses the negative y half-axis. These accelerations correspond to the trajectories' 'jumps' observed in the diffracted Kummer OV beams (Fig. 13.6 (a)) and can be explained with allowance for the cosine term in (13.27) (see Section 13.7).

In summary, the far-field pattern of the diffracted OV beam supplies an interesting example where the input wavefront curvature is transformed into the output azimuthal rotation of the secondary OV pair, which can possibly find applications for wavefront diagnostics and measurements.

13.6. Discontinuities of the OV Trajectories and Topological Reactions in the Diffracted OV Beams

The previous sections illustrate and explain the main regularities of the OV trajectories and the singular skeleton evolution in diffracted OV beams. Now we proceed with the study of apparent 'irregularities' which so far were beyond our attention. In essence, we will concentrate on the non-uniform velocity of the OV motion along its trajectory (while the screen edge moves uniformly), which is most articulately expressed by the enigmatic 'jump' of the a -dependent OV trajectory in Fig. 13.6 (a). In this section based mainly on the recent results of Ref. [33] we are going to elucidate the nature and conditions of this and similar effects.

13.6.1. The 'Jump' Description: Kummer Beams

To this end, we reconsider the a -dependent OV trajectories for the diffracted Kummer beam (13.6)–(13.8) with $m = -3$ and parameters

$$k = 10^5 \text{ cm}^{-1}, \quad b = 0.232 \text{ mm}, \quad R = 54 \text{ cm}, \quad z_h = 11 \text{ cm}, \quad (13.38)$$

observed in the cross section $z = 30$ cm behind the screen (see Fig. 13.1 (c)). The results are given in Fig. 13.14 representing a supplemented and re-arranged left column of Fig. 13.6. Panel (a) illustrates the phase distribution in the observed beam cross section; the lines of different colors indicate the constant-phase contours with increment 1 rad (cf. Fig. 13.3 (c)). In Fig. 13.14 (a), three single-charged OV's are seen that originate from

decomposition of the incident 3-charged OV; Figs. 13.14 (b–d) show the trajectories of OVs B–D, respectively.

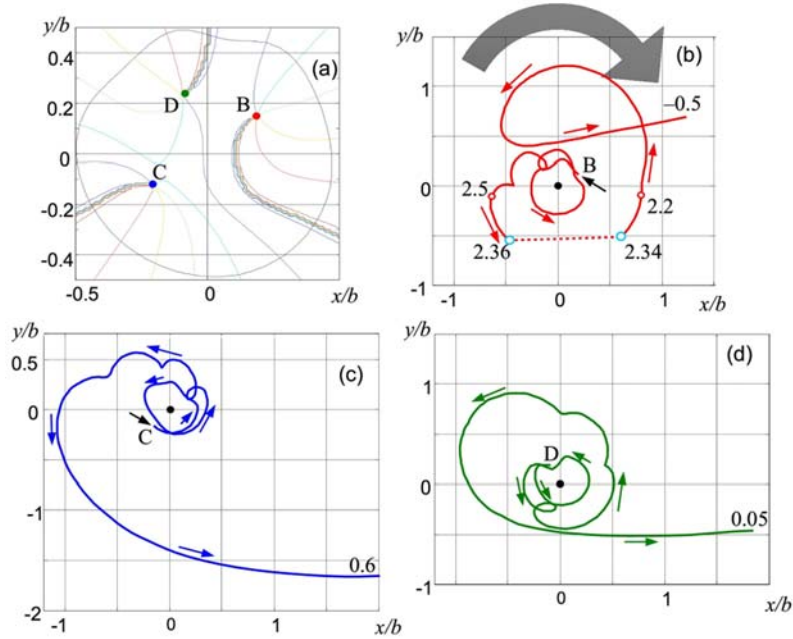


Fig. 13.14. Trajectories described by the OV cores in the cross section $z = 30$ cm behind the screen, the screen edge moving from $a = 4.4b$ to $a = -0.5b$ (see Fig. 13.1 (c)), for the incident Kummer beam with topological charge $m = -3$ and parameters (13.38). The transverse coordinates are expressed in units of b (13.38); large grey arrow shows the energy circulation in the incident beam (cf. Figs. 13.1(a) and 13.3 (b)), small arrows show the directions of the OV motion. (a) ‘Initial’ positions of the three secondary OVs marked B, C and D for $a = 4.4b$, the thin black curve denotes the constant intensity contour at a level 10 % of the maximum; (b)–(d) trajectories of OVs B, C and D while the screen edge advances (the final values of a/b at which the corresponding OV disappears are marked near the ends of the curves), the beam axis is denoted by the black circle. The dotted line in panel (b) illustrates the OV ‘jump’.

The non-uniformity of the OV motion along its trajectory is most impressively evident in the trajectory of the OV B (Fig. 13.14 (b)). While the screen performs a minute advance from $a = 2.36b$ to $a = 2.34b$, the OV abruptly ‘jumps’ between the points marked by cyan circles so that the trajectory apparently looks discontinuous (compare this with the adjacent trajectory segments where much larger screen shifts from $a = 2.5b$ to $a = 2.36b$ and from $a = 2.34b$ to $a = 2.2b$ cause noticeably smaller changes in the OV positions marked by the red circles). Also, while the OV B performs this ‘jump’, the positions of other OVs remain practically unchanged. In what follows, we intend to investigate the nature and mechanism of this effect.

To accomplish this, we return to the asymptotic analytical model of Eqs. (13.22), (13.23). We employed these equations in the previous sections but under the assumption that at

the WDP conditions, the off-axial displacements of the OV cores are small enough so that in Eq. (13.16) $M \ll 1$, and the cosine term in the left-hand side of (13.23) can be omitted. However, the trajectory details we are studying in this Section appear at not very small r -values when the cosine term cannot be discarded. Subsequently, the equations (13.15) or (13.23) for the azimuthal coordinate of the OV core become transcendent and, in contrast to what was made in Section 13.5, cannot be solved analytically. To inspect the main consequences of the trigonometric term, we investigate the simpler equation (13.15). Its qualitative analysis is illustrated in Fig. 13.15 (a). The left-hand side as a function of ϕ is imaged by the blue curve (for comparison, the thin light-blue line represents the left-hand side in the limiting case $M=0$), and each horizontal line expresses a certain value of the right-hand side depending on a and z for a certain secondary OV number N . The solution $\phi(a, z)$ is obtained as an intersection of the blue curve and the corresponding horizontal line. In the ‘normal’ situation, $M=0$, there is only one intersection point (see, e.g., points ϕ_1 and ϕ_4 in Fig. 13.15 (a)). When applied to the case of $m < 0$ presented in Fig. 13.14, with a decreasing monotonically the horizontal line moves upward, and the corresponding $\phi(a, z) = \phi_1$ also changes monotonically and continuously. However, due to the trigonometric term in Eq. (13.15), the left-hand side can be non-monotonic, and at certain values of a and z , the horizontal line reaches the region where the blue curve is nearly horizontal or decreases (e.g., between the red dashed lines in Fig. 13.15 (a), $\phi_2 < \phi < \phi_3$). Obviously, in this region $\phi(a, z)$ can change very rapidly; besides, there appear additional intersections that testify for nothing but emergence of additional OVs.

The graphical solution of Eqs. (13.22)–(13.16) is performed similarly. Although the evolution of the blue curve is more complicated than was discussed in the above paragraphs because of the variable M (13.16), which depends on a explicitly as well as implicitly, via r and Eq. (13.22), and due to the more complex a -dependence of the right-hand side of Eq. (13.23), the principal details remain the same [33].

The existence of several intersections of the horizontal line with the blue curve (as for the green line in Fig. 13.15 (a)) means that the smooth translational migration of the OV is no longer possible and is thus replaced by the topological reaction in which additional OVs emerge and annihilate [4]. Images of Figs. 13.15 (b–d) show the numerical example explaining the behavior of the OV B whose trajectory is depicted in Fig. 13.14 (b), within the ‘jump’ region. The OV positions are marked by the corresponding letters, as in Fig. 13.14 (b–d); additionally they are provided with curved arrows showing the local direction of transverse energy circulation, colored in agreement with the trajectory colors in Figs. 13.6 and 13.14. While a approaches the ‘jump’ region ($a = 2.36$ in Fig. 13.14 (b), point ϕ_2 in Fig. 13.15 (a)), there are three secondary OVs presented in Fig. 13.15 (b). At this moment, the small screen advancement towards the axis almost does not affect the OV positions but induces a topological event: in the area indicated by the black circle in Fig. 13.15 (b), the cut is torn and the dipole of oppositely charged OVs emerges (see Fig. 13.15 (c)). With further decrease of a , one of the new-born OVs, V, charged oppositely to all the other OVs (black curve arrow), rapidly moves against the ‘normal’ spiral OV motion. Subsequently it meets the OV B and annihilates with it, whereas the second member of the dipole pair, B', still remains and starts its migration as a

‘continuation’ of the OV B (Fig. 13.15 (d)). Note that singularities C and D are practically stable during this process, and the ‘virtual’ OV V moves from B' to B along the smooth arc looking as a natural ‘filling’ of the spiral-like trajectory between $a = 2.36$ and $a = 2.34$. This agrees with the approximate Eq. (13.22) that dictates that radial coordinates of all OVs, including the ‘virtual’ ones, are determined by a and z independently of the azimuth ϕ .

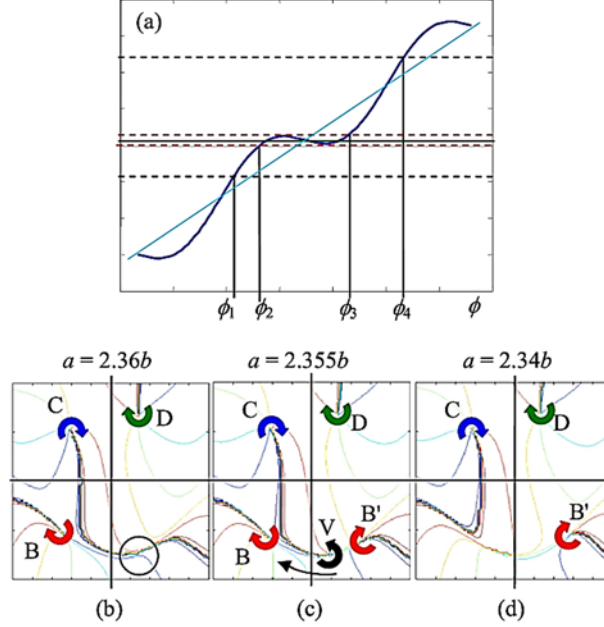


Fig. 13.15. (a) Illustration for the solution of Eqs. (13.15) and (13.23): The blue curve is the plot of the left-hand side expression for $|M| = 1.4$, horizontal lines symbolize different (a, z) -dependent values of the right-hand side. (b)–(d) Equiphase contours and the secondary OV positions in the cross section of the diffracted beam of Fig. 13.14; curve arrows show the local energy circulation near the OV cores; the screen-edge positions are indicated above each panel (further explanations in text).

The described anomalies of the OV trajectories in the diffracted beam are caused by the non-monotonic character of the left-hand side of Eq. (13.15) or (13.23), which takes place if the ‘jump criterion’ is realized,

$$|M| = \left| \frac{kra}{mz} \right| > 1, \quad (13.39)$$

and near the points where

$$\cos \phi = 0, \quad \frac{d}{d\phi} (M \cos \phi) < 0. \quad (13.40)$$

The latter condition explains why the jump of Fig. 13.14 (b), as well as the noticeable acceleration of the OV motion in Figs. 13.14 (c, d) [21] occur in the lower half-plane, near $\phi = 3\pi/2$ (remember that $m < 0$ and, consequently, $M < 0$); this is also the reason for the ‘steep’ segments of the solid curves in Fig. 13.7 (d) near the points where $\phi = -\pi/2, 3\pi/2, 7\pi/2$. In turn, Eq. (13.39) shows that the ‘jump’ can preferably take place at large enough a and not very high z ; in particular, this explains why the numerical analysis reveals the ‘jump’ anomalies at $z = 30$ cm but they cannot be detected, with the same incident beam, at $z = 60$ cm and $z = 82$ cm (see Fig. 13.6). In the present conditions of Eq. (13.38) and Fig. 13.14 (b) with $z = 30$ cm, $a = 2.35b$, $r \approx 0.72b$, one finds $|M| \approx 1.01$, which agrees with the ‘jump’ existence. Noteworthy, the trajectories of the OVs C and D differ from the considered OV-B trajectory by the values of a and r at which they traverse the vicinity of $\phi = 3\pi/2$. For the OV C this occurs at $a = 3.75b$, $r \approx 0.25b$ (Fig. 13.14 (c)), which gives $|M| = 0.56$; for the OV D – at $a = 3.1b$, $r \approx 0.4b$ (Fig. 13.14 (d)) whence $|M| = 0.74$. This completely agrees with the absence of jumps and accompanying topological events in trajectories C and D. Similarly, one can easily verify that $|M| < 1$ for all curves of Figs. 13.4–13.6, except Fig. 13.6 (a) and the red curve in Fig. 13.6 (j).

13.6.2. Discontinuities in the Laguerre-Gaussian Beams’ Diffraction

Now let us consider the singular skeleton evolution upon diffraction of the LG beam (13.9). Instead of the previously analyzed beam with parameters (13.11), here we choose another example that enables us to expose the special features of the OV trajectories more explicitly: we take the LG beam with $m = -3$ and assume in the calculations the following values for the beam parameters:

$$k = 10^5 \text{ cm}^{-1}, \quad b_c = b_0 = b = 0.232 \text{ mm}, \quad z_c = 0, \quad R_c = \infty, \quad (13.41)$$

(like in Section 13.5.1, the beam waist coincides with the screen plane). The numerically calculated OV trajectories in the diffracted beam cross section are presented in Fig. 13.16. As in the Kummer beam case (Fig. 13.14), there are three secondary OVs that evolve along the spiral-like trajectories and consecutively move to the shadow region where they vanish. The trajectories are marked by the same colors and the same letter notations as their counterparts in Fig. 13.14 (b–d). As expected, they show more regular and smooth behavior than in the case of a Kummer beam, which is associated with the slower decay and oscillations of the Kummer beam intensity at $r \gg b$ [21, 34] (see the remarks in the penultimate paragraph of Section 13.4.2). Accordingly, the analytical model of Section 13.4.2 not only provides qualitative but also a fair quantitative characterization of the trajectory B even if $a \approx b$, cf. Fig. 13.16(a) where the trajectory obtained analytically from Eqs. (13.26), (13.27) with $M = 0$ is presented as the thin dotted spiral; note that its final point corresponds to $a = 1.2b$.

Upon calculations, the ‘jumps’ were identified as events for which an additional pair of OVs emerge. For example, in Fig. 13.16 (a), while a decreases, the ‘red’ OV with topological charge -1 moves along the segment B_0B and at the moment it approaches point B, the OV dipole is distinguished with -1 -charged OV in point B'. This event takes place at $a = 1.98b$; then, the oppositely charged dipole member – ‘virtual’ OV V – rapidly

moves along the black arc against the main spiral evolution. Meanwhile, the ‘old’ OV still continues its slow motion to meet the ‘virtual’ one until the annihilation occurs in the point marked by the circle at $a = 1.94b$.¹ During the entire process, the OV radial coordinate remains approximately constant, $r = 0.44b$. Similar events happen to the OV C at $a = 2.92b$ to $2.90b$ (Fig. 13.16 (b), $r = 0.27b$) and to the OV D at $a = 2.52b$ to $2.48b$ (Fig. 13.16 (c), $r = 0.35b$). In contrast to the situation of Fig. 13.14, now all the OVs experience rather articulate ‘jumps’, which is explained by the high values of the jump factor (13.39): $|M| = 1.56, 1.40$ and 1.57 in cases of Fig. 13.16 (a–c), correspondingly.

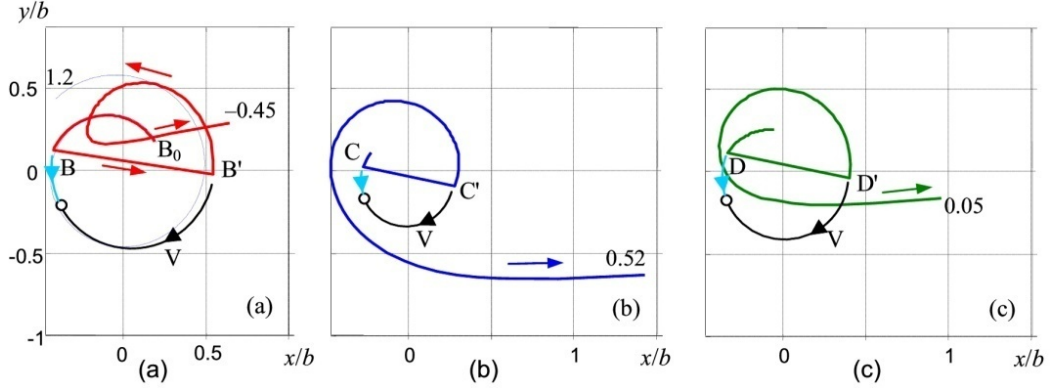


Fig. 13.16. Trajectories described by the OV cores in the cross section $z = 10$ cm behind the screen, the screen edge moving from $a = 3b$ to $a = -0.45b$ (see Fig. 13.1 (a)), for the incident LG beam with topological charge $m = -3$ and parameters (13.41). Each panel shows the trajectory of a single OV with additional explanatory details. The transverse coordinates are expressed in units of b (13.41), small arrows show the directions of the OV motion; the final values of a/b at which the corresponding OV disappears are marked near the ends of the curves. The trajectories experience ‘jumps’ between points B and B', C and C', D and D', respectively; the black (cyan) arcs represent the motion of ‘virtual’ (‘old’) OVs before their annihilation in points marked by circles. In panel (a), the trajectory calculated analytically via Eqs. (13.26), (13.27) for $3b > a > 1.2b$ with $M = 0$ is depicted by the thin dotted curve for comparison.

13.7. OV Jumps in the z -Dependent Singular Skeleton Evolution

According to the general physical arguments based on the analytical suggestions supplied by Eqs. (13.14), (13.15), (13.22), (13.23), (13.26) and (13.27), the discussed mechanisms determining the OV trajectories are still in charge of the z -dependent evolution, and the trajectory discontinuities and topological reactions of the above-described type are expected to occur in this situation as they do in the a -dependent trajectories studied in Section 13.6. Some examples below will illustrate the relevant processes.

¹Note that the ‘virtual’ OV distantly resembles the virtual particles in quantum theory [37]: it is short-lived, and its only role is to implement the reaction transforming B into B'

13.7.1. Kummer Beams

Fig. 13.17 represents the z -dependent evolution of the secondary OV's in the same diffracted beam that was analyzed in Section 13.3.3, Fig. 13.6 (left column) and Section 13.6.1, Fig. 13.14 but for the fixed screen-edge position $a = 4b$ illustrated in the panel (a). Note that, to make the beam structure better visible, the transverse amplitude distribution $|u^K(x_a, y_a, z_h)|$ is presented instead of the more common intensity $|u^K(x_a, y_a, z_h)|^2$. Anyway, the screen barely ‘touches’ the beam periphery, which, nevertheless, induces rich of details and quite observable perturbations of its singular skeleton displayed in Fig. 13.17 (b–d). In case of a propagating beam, there always is present the trivial component of the OV migration associated with the overall beam divergence; to abstract from this non-informative component, in Fig. 13.17 (b–d) the OV trajectories are displayed in the normalized transverse coordinates

$$x_e = x \left(1 + \frac{z}{R}\right)^{-1}, \quad y_e = y \left(1 + \frac{z}{R}\right)^{-1}. \quad (13.42)$$

In general, the OV trajectories of Fig. 13.17 (b–d) are similar to those of Fig. 13.14 (b–d) and show the same character of pulsating spirals. In the course of diffracted beam propagation (growing z), the pulsation period increases and in the far field the pulsations vanish. In contrast to the trajectories of Figs. 13.6 and 13.14, here are no self-crossings (‘loops’ as in Figs. 13.14 (b–d)); the apparent self-crossings near $z = 20$ cm in Fig. 13.17 (d) are illusive and appear only in the normalized coordinates (13.42). The most important feature is that in case of the z -dependent evolution there also exist regions of very rapid OV migration (the trajectories’ segments between the white-filled circles). In full agreement with the model of Section 13.6.1 (see Eq. (13.40) and Fig. 13.15 (a)), these regions are in the lower half-plane (near the OV core azimuth $\phi = 3\pi/2$). However, the ‘true’ jump only happens to the OV D in the panel (d). This agrees with the criterion (13.39) that can be checked based on the presented trajectories: in Fig. 13.17 (b), $r = 0.18b$, $z = 14$ cm, and $|M| = 0.97$; in Fig. 13.17 (c), $r = 0.226b$, $z = 17.4$ cm, and $|M| = 0.93$; and only in Fig. 13.17 (d) $r = 0.171b$, $z = 11.4$ cm, $|M| = 1.08$ – the conditions for the jump are realized, and it is indeed observed.

13.7.2. Laguerre-Gaussian Beams

Diffraction of an LG beam provides additional and rather conspicuous illustrations for the 3D singular skeleton evolution [23, 33]. Like in Section 13.6.2, we consider the incident LG beam (13.9) with its waist in the screen plane and the Gaussian envelope parameters (13.41) but with the topological charge $m = -2$ (Fig. 13.18).

Despite that the chosen screen edge position $a = 2b$ can hardly be treated as a far periphery of the incident beam profile and the expected perturbation of its structure is rather strong, the OV migration looks remarkably regular (Fig. 13.18 (b)). As in Section 13.5, to remove the trivial migration component associated with the beam divergence, the coordinates are

normalized by the Gaussian envelope radius (13.31) of the supposed unperturbed incident beam, where, in view of Eq. (13.41), $z_{Rc} = z_R = 53.8$ cm is the Rayleigh length of the incident beam. Again, as was noted during comparison of Figs. 13.16 and 13.14, the OV trajectories in the diffracted LG beam form almost perfect spirals, without pulsating irregularities observed in Figs. 13.17 (b–d) for the diffracted Kummer beam.

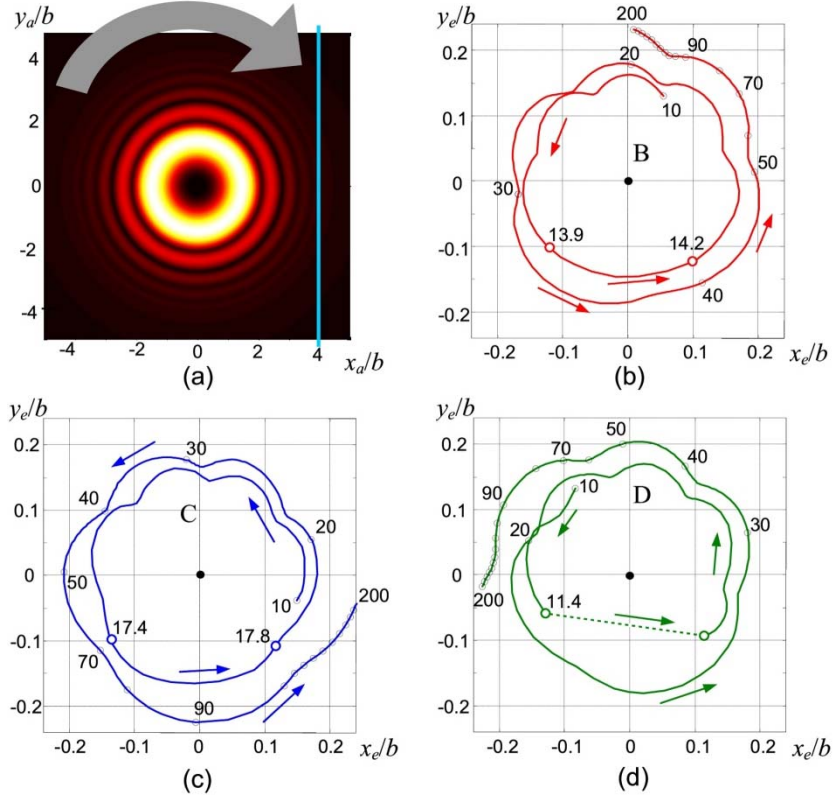


Fig. 13.17. Transverse projections of the OV trajectories behind the screen whose edge is fixed at $a = 4b$ (see Fig. 13.1(a)), for the incident Kummer beam with topological charge $m = -3$ and parameters (13.38) (cf. Fig. 13.14). (a) The screen edge position (blue line) against the incident beam amplitude distribution in the screen plane, the large arrow shows the energy circulation direction. (b)–(d) Separate OV trajectories for z growing from 10 cm to 200 cm, letters B, C and D denote the same secondary OVs that are shown in left column of Fig. 13.6 and in Fig. 13.14; thin black empty circles correspond to z values multiple of ten in centimeters, some of them are provided with corresponding numerical marks; colored white-filled circles mark the segments of rapid evolution. The horizontal and vertical coordinates are in normalized units of (13.42); small arrows show the directions of the OV motion. The trajectory ‘jump’ is seen only in panel (d) at $z = 11.4$ cm (dotted line).

In Fig. 13.18 (b) the OV B trajectory (red) experiences the jump at $z = 7.05$ cm while the OV C (blue) only shows the rapid evolution between $z = 13.1$ cm and $z = 13.7$ cm. This, again, is in full compliance with the criteria (13.39) and (13.40): for the OV C, $r = 0.234b$,

and with $m = -2$, $a = 2b$, $z = 13.1$ cm this entails $|M| = 0.96$ whereas for the OV B, $r = 0.191b$, $z = 7.05$ cm, and $|M| = 1.46$. The jump mechanism is completely the same as in other examples: the OV dipole is born in point B' after which its oppositely charged 'virtual' member V rapidly moves 'backward' towards the 'old' B and annihilates with it in point A corresponding to $z = 7.65$ cm. This example supplies a spectacular dynamical illustration of the topological reactions and the 'virtual' OV migration accompanying the jump.

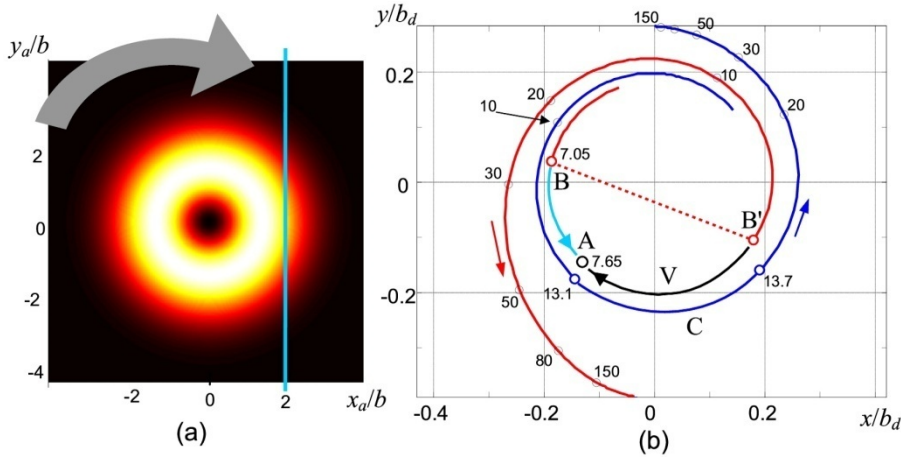


Fig. 13.18. Transverse projections of the OV trajectories behind the screen whose edge is fixed at $a = 2b$ (see Fig. 13.1 (a)), for the incident LG beam with topological charge $m = -2$ and parameters (13.41). (a) The screen edge position (blue line) against the incident beam amplitude distribution in the screen plane, the large arrow shows the energy circulation direction. (b) Red (B) and blue (C) curves represent the trajectories of the two secondary OVs for z growing from 5.6 cm to 530 cm ($9.85z_{Re}$); black empty circles denote the intermediate z values (marked in centimeters); colored white-filled circles mark the segments of rapid evolution. The transverse coordinates are given in units normalized by (13.31); small arrows show the directions of the OV motion. At $z = 7.05$ cm, the OV B experiences the 'jump' into B' position shown by the dotted line; the cyan and black arcs represent the evolution of the 'old' OV B and of the 'virtual' OV V after the jump until they annihilate in the point A marked by the black empty circle.

13.7.3. 3D Trajectories and the Nature of Discontinuities

To elucidate in more detail the discontinuous trajectory of the OV B in Fig. 13.18 (b), we present it as a 3D graph together with the trajectories of the 'old' OV B after the jump and of the virtual OV (cyan and black curves of Fig. 13.18 (b)). The result given in Fig. 13.19 reveals that the three trajectories of Fig. 13.18 (b) are actually fragments of the single 'full' curve that is perfectly continuous and smooth, so the jumps and topological reactions appear only in its projections (in particular, the red, cyan and black curves of Fig. 13.18 (b) are projections of the corresponding segments of the curve of Fig. 13.19 viewed from the positive end of axis z). This agrees with the usual concepts of the OV filaments [6, 38, 39] and discloses the nature of the intriguing effects considered in the previous sections.

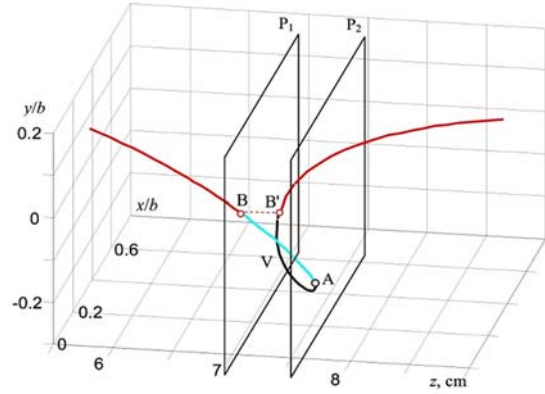


Fig. 13.19. 3D trajectory of the ‘red’ (B) OV of Fig. 13.18 (b) (incident LG beam with $m = -2$ and parameters (13.41), screen edge position $a = 2b$) in the near-jump region ($5.6 \text{ cm} < z < 9 \text{ cm}$). The transverse coordinates are given in units of b (13.41); plane P_1 ($z = 7.05 \text{ cm}$) crosses the trajectory in point B and is tangent to it in point B', plane P_2 ($z = 7.65 \text{ cm}$) is tangent to the trajectory in the annihilation point A (black empty circle); the red, cyan and black segments correspond to the red, cyan and black arcs in Fig. 13.18 (b).

Let the ‘full’ OV trajectory of Fig. 13.19 be represented in parametrical form, i.e. the coordinates of a current trajectory point are expressed as functions of the trajectory length s measured from the starting point, say, at $z = 5.6 \text{ cm}$:

$$x_v = x_v(s), \quad y_v = y_v(s), \quad z_v = z_v(s). \quad (13.43)$$

In a given transverse plane, the OV position is determined as an intersection between the plane and the trajectory. The ‘normal’ evolution implies that $dz_v/ds > 0$ everywhere, and then in each observation plane only one intersection point can exist; however, in some configurations of the diffracted beam singular skeleton, regions of a ‘retrograde’ evolution, where

$$dz_v/ds < 0, \quad (13.44)$$

may occur. It is such a situation that is depicted in Fig. 13.19 between the transverse planes P_1 and P_2 . When the observation plane approaches P_1 from the left, it ‘touches’ the trajectory at the additional point B' (a local minimum of the function $z_v(s)$), which corresponds to the dipole emergence. With further advance, the observation plane will contain three intersection points with the curve, which are interpreted as the ‘teleported’ OV B', ‘old’ OV B and the ‘virtual’ oppositely charged OV V. In the position P_2 the observation plane again touches the trajectory, now in point A with the local maximum of $z_v(s)$, and the intersections corresponding to B and V disappear: the two OVs annihilate.

This picture completely explains the discontinuous trajectories of the OV cores not only in case of the z -dependent evolution (Sections 13.7.1, 13.7.2) but also in case of the screen edge translation (Section 13.6). In the latter situation, the observation plane is fixed but

the ‘full’ 3D curve is smoothly deformed with variation of a , and the 2D trajectory jump takes place if in the observation plane the condition (13.44) becomes true. In fact, the ‘jump criterion’ (13.39) is equivalent to (13.44), and this is why it is equally applicable to both the z -dependent and a -dependent variations of the diffracted beam singular skeleton.

Here we are touching the aspect in which the theory of OV diffraction becomes entangled into the rich and stimulating field of the vortex lines and their geometry (see, e.g., [6] and references therein). This aspect deserves a special investigation; now we only remark that the intricate and at first glance artificial patterns of the OV lines that are deliberately generated by means of special procedures [6, 39] can naturally exist in the edge-diffracted circular OV beams.

13.8. Conclusion

The main results presented in this chapter testify that the simple and ubiquitous situation of edge diffraction provides additional impressive manifestations of the helical nature of light beams with OV. The diffraction obstacle introduces the beam perturbation that causes the OV displacement from its original axial position (for an m -charged OV, $|m|$ single-charged displaced OVs are formed around the axis). Then, while the diffracted beam freely propagates (z -dependent evolution), the displaced OVs migrate over the beam cross section along spiral-like trajectories, initially (at small post-screen distances) with high rotation rate, which rapidly decreases and practically stops far enough behind the screen. The similar OV migration takes place in a fixed cross section when the screen edge moves towards the beam axis (a -dependent evolution), in compliance with the approximate relations (13.32) or (13.15). The most articulate spiral-like trajectories occur under conditions of weak diffraction perturbation (WDP), when the screen edge is separated from the incident beam axis by two or more beam radii, i.e. when the beam circular shape is almost unaffected by the diffraction. This puts certain limitations to the practical employment of the predicted behavior, since the off-axial OV displacements are small and their measurement may impose difficulties. However, we emphasize the principal importance of the predicted features and their close relation with the physical nature of OVs. Besides, at least for a multicharged OV diffraction, even under the WDP conditions, the typical OV displacement can reach $\sim 10\%$ of the current beam radius b (see Figs. 13.5, 13.6, 13.12, 13.13–13.18), which is quite measurable.

It should be noted that the observed and predicted peculiar details of the singular skeleton behavior are rather common for light beams with well-developed singular structure, e.g. speckle fields [4, 6]. In this view, the diffracted OV beams can be considered as their simplified models and, possibly, produce efficient means to create controllable singular-optics structures with prescribed properties, which can be useful in diverse research and technology applications.

In particular, the presence of the well-developed, regular and easily interpretable singular structure makes the diffracted OV beams suitable objects for the general study of the OV filaments and their geometric regulations, evolution of individual singularities, their transformations, topological reactions and interactions. On the other hand, the OV

trajectories' discontinuities, 'jumps', birth and annihilation events described in this Chapter are, as a rule, highly sensitive to the incident beam parameters and the diffraction conditions. For example, the OV positions in the diffracted beam cross section can be sensitive indicators of the screen edge position with respect to the incident beam axis, which can be employed for precise distant measurements of small displacements and deformations [15, 19]. From Figs. 13.16 (b) and 13.18 one can easily see that near the 'threshold' conditions of topological reactions, the screen edge displacement of $0.01b$ induces a two orders of magnitude larger OV jump in the diffracted beam cross section. Note that such sensitivity is predicted without any special consideration; undoubtedly, a detailed analysis aimed at the search of the diffraction parameters most favorable for the distant metrology will improve these figures. This aspect of the present work enables us to suggest its applications for the problems of the precise OV metrology [29–32] as well as for the incident OV diagnostics, which can be prospective in the fields of laser beam shaping and analysis and in optical probing systems.

Additional interesting possibilities are associated with relations between the incident beam wavefront curvature and the diffracted beam structure. First, we have refined the earlier statements [15, 19] that in the far field, the diffracted circular OV beam acquires the symmetry with respect to an axis parallel to the screen edge. In fact, this symmetry is only realized in the beam's Fraunhofer (Fourier) plane, which can be real (if the beam converges, radius of curvature $R_c < 0$) or imaginary (if $R_c > 0$), and only for beams with plane wavefront it occurs at the propagation infinity (in the extreme far field). Importantly, the actual far-field pattern of the diffracted beam essentially depends on the incident beam wavefront curvature. In terms of the OV displacements, any change of the incident beam wavefront curvature—while preserving the intensity profile—induces the azimuthal rotation of the far-field OV position in the diffracted beam. This sensitivity may be useful for the OV beam diagnostic and for the wavefront measurements.

Most of the quantitative results of this Chapter are obtained numerically but their interpretation is based on the asymptotic analytical model of Eqs. (13.15)–(13.17) with refinements (13.22), (13.23) and (13.26), (13.27). Remarkably, the model derived for the condition $a \gg b$ appears to be valid in the much larger and physically interesting domain; at least, for the LG beam diffraction it does not fail even at $a \approx 2b$, and the model-based criterion (13.39) works perfectly well in all the considered examples. However, the model predicts monotonic behavior of the OV radial displacement r with growing z for Kummer beams, i.e. does not explain the radial pulsations of the spirals in Fig. 13.17 (b–d). Nevertheless, we hope that despite its approximate character, the model will give a reliable analytical basis for further research of the vortex beams' diffraction. At least, all the conclusions concerning the spiral-like character of the OV trajectories and their jumps when the criteria (13.39) and (13.40) are satisfied, are absolutely reliable and supported by experiment [20, 21]. The fine details of the OV trajectories in diffracted Kummer beams (self-crossings and pulsations in Figs. 13.14 (b–d) and 13.17 (b–d)), that appear due to the slow fall-off of the Kummer beam amplitude, are expected to be sensitive to the incident beam behavior at the far transverse periphery. In this view, even the 'routine' approximations usually employed in the numerical simulations can become sources of errors, e.g., the integration domain limitation in the Fresnel–Kirchhoff integral (13.4). In

such situations, the explicit allowance for the specific conditions of the Kummer beam preparation and for the optical system it passes would be necessary.

A possible direction for further research can be related with a more comprehensive characterization of the separate OV in the diffracted beam. So far we were only interested in their positions; but no less informative can be their morphology and anisotropy parameters [5, 6]: the orientation and the axes ratio of the equal-intensity ellipses in the nearest vicinity of the OV core. Especially, under conditions close to topological reactions, the OVs are highly anisotropic, and this supplies additional markers to characterize the qualitative discontinuities in the singular skeleton evolution. Another way of possible further development of ideas and approaches introduced in the present work can be oriented at the search of special conditions of the OV beam preparation and diffraction, which provide high sensitivity for the metrological and diagnostic applications outlined two paragraphs above.

Finally, we would like to note that the edge diffraction can be considered a special case of the OV beam transformation with symmetry violation. The specific response of OV beams to their symmetry breakdown is interesting from a fundamental point of view and supplies efficient means for their investigation and diagnostics [5, 14, 16, 18, 40, 41]. There is a wide variety of such transformations; in particular, the astigmatic focusing and telescopic transformations of OVs are studied in much detail (see, e.g., [42–44]). But they show quite different behavior of the secondary OVs within the transformed beam cross section, most likely due to preserving the central symmetry of the beam transverse profile. However, one can expect that some features of the OV migration, similar to those considered in this chapter, can be detected in OV beams subjected to the symmetry violation which destroys the central symmetry. It would be meaningful and instructive to inspect the response of an OV beam to the unilateral beam constraint, e.g., by a “soft” diaphragm, or even unilateral phase modification in some peripheral part of the beam cross section. These transformations are, properly, special cases of diffraction but they admit more ‘gentle’ conditions suppressing the edge waves, diffraction fringes [2], etc., which will be favourable for precise measurements.

Acknowledgements

The authors are grateful to Lidiya Mikhaylovskaya, Anna Khoroshun and Aleksey Chernykh for the valuable support and assistance. This work was supported, in part, by the Ministry of Science and Education of Ukraine.

References

- [1]. M. Born, E. Wolf, Principles of Optics, *Cambridge University Press*, 1999.
- [2]. S. Solimeno, B. Crosignani, P. DiPorto, Guiding, Diffraction and Confinement of Optical Radiation, *Academic Press*, 1986.

- [3]. H. Rubinsztein-Dunlop, A. Forbes, M. V. Berry, M. R. Dennis, D. L. Andrews, M. Mansuripur, C. Denz, C. Alpmann, P. Banzer, et al., Roadmap on structured light, *J. Opt.*, Vol. 19, Issue 1, 2017, 013001.
- [4]. M. S. Soskin, M. V. Vasnetsov, Singular optics, *Prog. Opt.*, Vol. 42, 2001, pp. 219–276.
- [5]. A. Ya. Bekshaev, M. S. Soskin, M. V. Vasnetsov, Paraxial Light Beams with Angular Momentum, *Nova Science Publishers*, 2008.
- [6]. M. R. Dennis, K. O’Holleran, M. J. Padgett, Singular optics: optical vortices and polarization singularities, *Prog. Opt.*, Vol. 53, 2009, pp. 293–364.
- [7]. I. G. Marienko, M. V. Vasnetsov, M. S. Soskin, Diffraction of optical vortices, *Proceedings of SPIE*, Vol. 3904, 1999, pp. 27–34.
- [8]. M. V. Vasnetsov, I. G. Marienko, M. S. Soskin, Self-reconstruction of an optical vortex, *JETP Lett.*, Vol. 71, Issue 4, 2000, pp. 130–133.
- [9]. J. Arlt, Handedness and azimuthal energy flow of optical vortex beams, *J. Mod. Opt.*, Vol. 50, Issue 10, 2003, pp. 1573–1580.
- [10]. V. N. Gorshkov, A. N. Kononenko, M. S. Soskin, Diffraction and self-restoration of a severe screened vortex beam, *Proceedings of SPIE*, Vol. 4403, 2001, pp. 127–137.
- [11]. J. Masajada, Gaussian beams with optical vortex of charge 2- and 3-diffraction by a half-plane and slit, *Optica Applicata*, Vol. 30, Issue 2–3, 2000, pp. 248–256.
- [12]. J. Masajada, Half-plane diffraction in the case of Gaussian beams containing an optical vortex, *Opt. Commun.*, Vol. 175, Issue 4–6, 2000, pp. 289–294.
- [13]. H. X. Cui, X. L. Wang, B. Gu, Y. N. Li, J. Chen, H. T. Wang, Angular diffraction of an optical vortex induced by the Gouy phase, *J. Opt.*, Vol. 14, Issue 5, 2012, 055707.
- [14]. A. Ya. Bekshaev, K. A. Mohammed, I. A. Kurka, Transverse energy circulation and the edge diffraction of an optical-vortex beam, *Appl. Opt.*, Vol. 53, Issue 10, 2014, pp. B27–B37.
- [15]. A. Bekshaev, K. A. Mohammed, Transverse energy redistribution upon edge diffraction of a paraxial laser beam with optical vortex, *Proceedings of SPIE*, Vol. 9066, 2013, 906602.
- [16]. A. Bekshaev, K. Bliokh, M. Soskin, Internal flows and energy circulation in light beams, *J. Opt.*, Vol. 13, Issue 5, 2011, 053001.
- [17]. K. Volke-Sepulveda, R. A. Terborg, Can diffraction provide quantitative information about energy flux in an optical vortex?, in *Proceedings of the Frontiers in Optics/Laser Science Conference (FIO/LS’11)*, OSA Technical Digest, San Jose, CA, United States, 16–20 Oct. 2011, pp. JTuA38.
- [18]. A. Y. Bekshaev, I. A. Kurka, K. A. Mohammed, I. I. Slobodeniuk, Wide-slit diffraction and wavefront diagnostics of optical-vortex beams, *Ukr. J. Phys. Opt.*, Vol. 16, Issue 1, 2015, pp. 17–23.
- [19]. A. Ya. Bekshaev, K. A. Mohammed, Spatial profile and singularities of the edge-diffracted beam with a multicharged optical vortex, *Opt. Commun.*, Vol. 341, 2015, pp. 284–294.
- [20]. A. Chernykh, A. Bekshaev, A. Khoroshun, L. Mikhaylovskaya, A. Akhmerov, K. A. Mohammed, Edge diffraction of optical-vortex beams formed by means of the “fork” hologram, *Proceedings of SPIE*, Vol. 9809, 2015, 980902.
- [21]. A. Bekshaev, A. Chernykh, A. Khoroshun, L. Mikhaylovskaya, Localization and migration of phase singularities in the edge-diffracted optical-vortex beams, *J. Opt.*, Vol. 18, Issue 2, 2016, 024011.
- [22]. A. Bekshaev, L. Mikhaylovskaya, A. Chernykh, A. Khoroshun, Evolution of the phase singularities in edge-diffracted optical-vortex beams, in *Proceedings of the IEEE 7th International Conference on Advanced Optoelectronics and Lasers (CAOL’16)*, Odessa, Ukraine, 12–15 Sept. 2016, pp. 23–25.
- [23]. A. Bekshaev, A. Chernykh, A. Khoroshun, L. Mikhaylovskaya, Displacements and evolution of optical vortices in edge-diffracted Laguerre-Gaussian beams, *J. Opt.*, Vol. 19, Issue 5, 2017, 055605.

- [24]. M. Chen, F. S. Roux, J. C. Olivier, Detection of phase singularities with a Shack–Hartmann wavefront sensor, *J. Opt. Soc. Am. A*, Vol. 24, Issue 7, 2007, pp. 1994–2002.
- [25]. K. Murphy, D. Burke, N. Devaney, C. Dainty, Experimental detection of optical vortices with a Shack–Hartmann wavefront sensor, *Opt. Express*, Vol. 18, Issue 15, 2010, pp. 15448–15450.
- [26]. O. V. Angelsky, S. G. Hanson, A. P. Maksimyak, P. P. Maksimyak, On the feasibility for determining the amplitude zeroes in polychromatic fields, *Opt. Express*, Vol. 13, Issue 12, 2005, pp. 4396–4405.
- [27]. W. Wang, T. Yokozeki, R. Ishijima, A. Wada, Y. Miyamoto, M. Takeda, S. G. Hanson, Optical vortex metrology for nanometric speckle displacement measurement, *Opt. Express*, Vol. 14, Issue 1, 2006, pp. 120–127.
- [28]. A. M. S. Maallo, P. F. Almoro, Numerical correction of optical vortex using a wrapped phase map analysis algorithm, *Opt. Lett.*, Vol. 36, Issue 7, 2011, pp. 1251–1253.
- [29]. M. R. Dennis, J. B. Götte, Topological aberration of optical vortex beams: Determining dielectric interfaces by optical singularity shifts, *Phys. Rev. Lett.*, Vol. 109, Issue 18, 2012, 183903.
- [30]. A. Popiołek-Masajada, B. Sokolenko, I. Augustyniak, J. Masajada, A. Khoroshun, M. Bacia, Optical vortex scanning in an aperture limited system, *Optics and Lasers in Engineering*, Vol. 55, 2014, pp. 105–112.
- [31]. D. Wojnowski, E. Jankowska, J. Masajada, J. Suszek, I. Augustyniak, A. Popiołek-Masajada, I. Ducin, K. Kakarenko, M. Sypek, Surface profilometry with binary axicon-vortex and lens-vortex optical elements, *Opt. Lett.*, Vol. 39, Issue 1, 2014, pp. 119–122.
- [32]. Ł. Płociniczak, A. Popiołek-Masajada, J. Masajada, M. Szatkowski, Analytical model of the optical vortex microscope, *Appl. Opt.*, Vol. 55, Issue 12, 2016, pp. B20–B27.
- [33]. A. Bekshaev, A. Chernykh, A. Khoroshun, L. Mikhaylovskaya, Singular skeleton evolution and topological reactions in edge-diffracted circular optical-vortex beams, *Opt. Commun.*, Vol. 397, 2017, pp. 72–83.
- [34]. A. Ya. Bekshaev, A. I. Karamoch, Spatial characteristics of vortex light beams produced by diffraction gratings with embedded phase singularity, *Opt. Commun.*, Vol. 281, Issue 6, 2008, pp. 1366–1374.
- [35]. A. Bekshaev, O. Orlinska, M. Vasnetsov, Optical vortex generation with a “fork” hologram under conditions of high-angle diffraction, *Opt. Commun.*, Vol. 283, Issue 10, 2010, pp. 2006–2016.
- [36]. M. Abramovitz, I. Stegun, Handbook of Mathematical Functions, *National Bureau of Standards: Applied Mathematics Series*, Vol. 55, 1964.
- [37]. M. E. Peskin, D. V. Schroeder, An Introduction to Quantum Field Theory, *Westview Press*, 1995.
- [38]. Z. S. Sacks, D. Rozas, G. A. Swartzlander, Holographic formation of optical-vortex filaments, *J. Opt. Soc. Am. B*, Vol. 15, Issue 8, 1998, pp. 2226–2234.
- [39]. J. Leach, M. R. Dennis, J. Courtial, M. J. Padgett, Vortex knots in light, *New J. Phys.*, Vol. 7, 2005, 55.
- [40]. A. Ya. Bekshaev, M. S. Soskin, M. V. Vasnetsov, Optical vortex symmetry breakdown and decomposition of the orbital angular momentum of light beams, *J. Opt. Soc. Amer. A*, Vol. 20, Issue 8, 2003, pp. 1635–1643.
- [41]. A. Mourka, J. Baumgartl, C. Shanor, K. Dholakia, E. M. Wright, Visualization of the birth of an optical vortex using diffraction from a triangular aperture, *Opt. Express*, Vol. 19, Issue 7, 2011, pp. 5760–5771.
- [42]. V. Denisenko, V. Shvedov, A. Desyatnikov, D. Neshev, W. Krolikowski, A. Volyar, M. Soskin, Y. S. Kivshar, Determination of topological charges of polychromatic optical vortices, *Opt. Express*, Vol. 17, Issue 26, 2009, pp. 23374–23379.

- [43]. A. Ya. Bekshaev, M. S. Soskin, M. V. Vasnetsov, Transformation of higher-order optical vortices upon focusing by an astigmatic lens, *Opt. Commun.*, Vol. 241, Issues 4–6, 2004, pp. 237–247.
- [44]. A. Ya. Bekshaev, A. I. Karamoch, Astigmatic telescopic transformation of a high-order optical vortex, *Opt. Commun.*, Vol. 281, Issue 23, 2008, pp. 5687–5696.
- [45]. F. W. J. Olver, Introduction to Asymptotics and Special Functions, *Academic Press*, 1975.
- [46]. A. Ya. Bekshaev, O. V. Orlinska, Transformation of optical vortex beams by holograms with embedded phase singularity, *Opt. Commun.*, Vol. 283, Issue 7, 2010, pp. 1244–1250.

Appendix

Under the WDP conditions, it is suitable to represent Eq. (13.4) in the form

$$u(x, y, z) = u^I(x, y, z) - \frac{k}{2\pi iz} \int_{-\infty}^{\infty} dy_a \int_a^{\infty} dx_a u_a(x_a, y_a) \exp \left\{ \frac{ik}{2z} \left[(x - x_a)^2 + (y - y_a)^2 \right] \right\}, \quad (13.A1)$$

where $u^I(x, y, z)$ describes the complex amplitude distribution of an unperturbed beam (what had occurred in the observation plane if the screen were absent), and for $a \gg b$, $a \gg b_c$ the integral can be evaluated by means of the asymptotic analysis [45].

Let us start with considering the LG beam diffraction, then $u^I(x, y, z) = u^{LG}(x, y, z_c + z)$ (cf. Eq. (13.9)) while, in the integral in the right-hand side of Eq. (13.A1), the expression (13.9) enters immediately. Then, omitting the coordinate-independent multipliers of Eq. (13.9), the integral in Eq. (13.A1) acquires the form

$$\exp \left[\frac{ik}{2z} (x^2 + y^2) \right] \int_{-\infty}^{\infty} dy_a P(y_a, y, z_c - iz_{Rc}) \times \int_a^{\infty} dx_a (x_a + i\sigma y_a)^{|m|} P(x_a, x, z_c - iz_{Rc}), \quad (13.A2)$$

where

$$P(x_a, x, d) = \exp \left[\frac{ikx_a^2}{2} \left(\frac{1}{d} + \frac{1}{z} \right) - \frac{ik}{z} x x_a \right]. \quad (13.A3)$$

Under conditions of WDP, the sought OV cores' positions are close to the beam axis, so one can assume

$$x \simeq y \simeq 0, \quad (13.A4)$$

and neglect the summands proportional to x^2 and y^2 when compared to the coordinate-independent terms but retain the terms in the first degree in x . Further, for large a , the internal integral in Eq. (13.A2) can be estimated with the help of an asymptotic formula valid for arbitrary function $f(x)$:

$$\int_a^\infty f(x) \exp(iKx^2) dx \simeq \frac{i}{2K} \frac{f(a)}{a} \exp(iKa^2) + O\left(\frac{1}{a^2}\right). \quad (13.A5)$$

Consequently,

$$\int_a^\infty dx_a (x_a + i\sigma y_a)^{|m|} P(x_a, x, z_c - iz_{Rc}) \simeq \frac{iP(a, x, z_c - iz_{Rc})}{ak \left(\frac{1}{z_c - iz_{Rc}} + \frac{1}{z} \right)} (a + i\sigma y_a)^{|m|}, \quad (13.A6)$$

and the external integral of Eq. (13.A2) is estimated by the method of stationary phase [45] which in connection with condition (13.A4) gives

$$\int_{-\infty}^\infty dy_a (a + i\sigma y_a)^{|m|} P(y_a, y, z_c - iz_{Rc}) \simeq \sqrt{\frac{2\pi i}{k \left(\frac{1}{z_c - iz_{Rc}} + \frac{1}{z} \right)}} a^{|m|}. \quad (13.A7)$$

Hence, we obtain the final representation for the integral term in the right-hand side of Eq. (13.A1):

$$-\frac{(-i)^{|m|+1}}{\sqrt{|m|!}} \sqrt{\frac{i}{2\pi}} \frac{k}{z} \left(\frac{z_{Rc}}{z_c - iz_{Rc}} \right)^{|m|+1} \left[k \left(\frac{1}{z} + \frac{1}{z_c - iz_{Rc}} \right) \right]^{-3/2} \frac{a^{|m|-1}}{b_0^{|m|}} P(a, x, z_c - iz_{Rc}). \quad (13.A8)$$

The first term of Eq. (13.A1), with allowance for the near-axis condition (13.A4), reads

$$u^I(x, y, z) = u^{LG}(x, y, z_c + z) \simeq \frac{(-i)^{|m|+1}}{\sqrt{|m|!}} \left(\frac{z_{Rc}}{z_c + z - iz_{Rc}} \right)^{|m|+1} \left(\frac{x + i\sigma y}{b_0} \right)^{|m|}. \quad (13.A9)$$

Now, gathering all terms of Eq. (13.A1) and omitting the common coordinate independent

multiplier $\frac{b_0}{\sqrt{|m|!}} \left(-\frac{iz_{Rc}}{b_0} \right)^{|m|+1}$, we find the asymptotic representation of the diffracted beam's complex amplitude distribution (13.26), (13.27).

In case of the Kummer beam diffraction, in Eq. (13.A1) $u^I(x, y, z) = u^K(x, y, z_h + z)$, and due to Eq. (13.6) and equations of Ref. [34] and the near-axis condition (13.A4),

$$u^K(x, y, z_h + z) \simeq \frac{z_e}{z + z_h} \sqrt{\frac{\pi}{2^{3|m|}}} (-i)^{|m|+1} \frac{z_R^{|m|+1}}{z_e^{|m|/2} (z_e - iz_R)^{|m|/2+1}} \left(\frac{x_e + i\sigma y_e}{b} \right)^{|m|}, \quad (13.A10)$$

where

$$z_e = \frac{z + z_h}{1 + (z + z_h)/R}, \quad x_e = x \frac{z_e}{z + z_h}, \quad y_e = y \frac{z_e}{z + z_h}.$$

In the integrand of Eq. (13.A1), we use the asymptotic expression for the Kummer beam amplitude (13.6) valid for $(x_a^2 + y_a^2)/b^2 \gg 1$,

$$u_a(x_a, y_a) = \frac{z_{he}}{z_h} \left(\frac{x_a + i\sigma y_a}{\sqrt{x_a^2 + y_a^2}} \right)^{|m|} \left\{ |m| (-i)^{|m|+1} \frac{z_{he}}{k(x_{ae}^2 + y_{ae}^2)} \exp \left[\frac{ik}{2z_h} (x_a^2 + y_a^2) \right] \right. \\ \left. + \frac{i}{2^{2|m|}} \frac{z_R^{2|m|+1}}{z_{he}^{|m|} (z_{he} - iz_R)^{|m|+1}} \exp \left[\frac{i}{2} \frac{k}{z_{he} - iz_R} (x_{ae}^2 + y_{ae}^2) + \frac{ik}{2(z_h + R)} (x_a^2 + y_a^2) \right] \right\}. \quad (13.A11)$$

This expression is similar to the formal asymptotic of the Kummer function [46] modified with account for the non-zero wavefront curvature of the incident Gaussian beam [14]; however, in the considered range of $(x_a^2 + y_a^2)/b^2 \simeq 10$, the formal asymptotic expansion is not sufficiently accurate. To improve the approximation, the numerical coefficient in the second line of the asymptotic (13.A11) is modified, from $-2^{-|m|}i$ in [45], to $2^{-2|m|}i$ in Eq. (13.A11); validity of this correction was checked numerically.

Now both summands of (13.A11) should be substituted into the integral term of Eq. (13.A1). The first summand yields the corresponding summand of the integral term, which with omitted coordinate-independent coefficients obtains the representation (cf. the expression (13.A2))

$$\exp \left[\frac{ik}{2z} (x^2 + y^2) \right] \int_{-\infty}^{\infty} dy_a P(y_a, y, z_h) \int_a^{\infty} dx_a \frac{(x_a + i\sigma y_a)^{|m|}}{(x_a^2 + y_a^2)^{|m|/2+1}} P(x_a, x, z_h), \quad (13.A12)$$

where function $P(x_a, x, z_h)$ is defined in Eq. (13.A3). Then, via the corresponding analogs of Eqs. (13.A6) and (13.A7), we obtain

$$\int_a^\infty dx_a \frac{(x_a + i\sigma y_a)^{|m|}}{(x_a^2 + y_a^2)^{|m|/2+1}} P(x_a, x, z_h) \simeq \frac{iP(a, 0, z_h)}{ak \left(\frac{1}{z_h} + \frac{1}{z} \right)} \frac{(a + i\sigma y_a)^{|m|}}{(a^2 + y_a^2)^{|m|/2+1}}, \quad (13.A13)$$

$$\int_{-\infty}^\infty dy_a \frac{(a + i\sigma y_a)^{|m|}}{(a^2 + y_a^2)^{|m|/2+1}} P(y_a, y, z_h) \simeq \sqrt{\frac{2\pi i}{k \left(\frac{1}{z_h} + \frac{1}{z} \right)}} \frac{1}{a^2}, \quad (13.A14)$$

and, after restoring the coordinate-independent coefficients, arrive at the total contribution of the first summand of (13.A11) to the integral term of Eq. (13.A1)

$$|m|(-i)^{|m|+1} \sqrt{\frac{i}{2\pi}} \frac{z_h}{z} \left[k \left(\frac{1}{z_h} + \frac{1}{z} \right) \right]^{-3/2} \frac{1}{a^3} P(a, x, z_h). \quad (13.A15)$$

Similar operations with the second summand of Eq. (13.A11) lead to the expression (cf. Eqs. (13.A2) and (13.A12))

$$\exp \left[\frac{ik}{2z} (x^2 + y^2) \right] \int_{-\infty}^\infty dy_a P(y_a, y, z_d) \int_a^\infty dx_a \frac{(x_a + i\sigma y_a)^{|m|}}{(x_a^2 + y_a^2)^{|m|/2}} P(x_a, x, z_d), \quad (13.A16)$$

where

$$\frac{1}{z_d} = \frac{1}{z_{he} - iz_R} \left(\frac{z_{he}}{z_h} \right)^2 + \frac{1}{z_h + R}, \quad (13.A17)$$

which, finally, results in the following contribution to the integral term of Eq. (13.A1):

$$-\frac{ik}{2^{2|m|}} \sqrt{\frac{i}{2\pi}} \frac{z_{he}}{zz_h} \frac{z_R^{2|m|+1}}{z_{he}^{|m|} (z_{he} - iz_R)^{|m|+1}} \left[k \left(\frac{1}{z_d} + \frac{1}{z} \right) \right]^{-3/2} \frac{1}{a} P(a, x, z_d). \quad (13.A18)$$

Then, combining Eqs. (13.A1), (13.A10), (13.A15) and (13.A18), we find the complex amplitude representation (13.18)–(13.21). Note that due to relation

$$\frac{1}{z_{he} - iz_R} \left(\frac{z_{he}}{z_h} \right)^2 + \frac{1}{z_h + R} = \frac{i}{kb^2(z_h)} + \frac{1}{R(z_h)},$$

expressions (13.A17) and (13.A18) contain the radius $b(z_h)$ and the wavefront curvature radius $R(z_h)$ that the initial Gaussian beam, incident onto the VG, would have possessed in the screen plane if it had propagated ‘freely’, without the VG-induced transformation.

Chapter 14

Wavefront Reconstruction with Rotational Fields

Travis W. Axtell and Roberto Cristi

14.1. Introduction

Closed-loop Adaptive Optics (AO) systems require an estimate of the wavefront phase to command the two-dimensional controllable Deformable Mirror (DM) actuators [1, 2]. The AO community has developed several computationally efficient and robust algorithms for wavefront phase reconstruction.

The original wavefront reconstruction techniques are based on matrix vector multiplies [3-6] with computational complexity of $O(n^2)$ or higher, with n the total data size. Later, a more efficient technique based on the Fast Fourier Transform was proposed by Freischlad [7] with further refinement by Poyneer [8] with complexity $O(n \log_2 n)$. During the same time period as Poyneer, Gilles [9] produced a multigrid preconditioned conjugate-gradient method with the same computational complexity.

The first $O(n)$ work in accurate wavefront reconstruction was performed by MacMartin where he developed a multi-resolution hierarchic reconstructor [10]. His work is based on downsampling the data by factors of 4 or more and results show that the larger downsampling factor decreases the relative performance and increases the noise sensitivity.

Additional $O(n)$ work followed, such as from Gilles [11]. This work was a direct application of a multi-grid solver of a minimum-variance reconstructor based on a sparse approximation of the wavefront inverse covariance matrix. Vogel also improved sparse matrix methods [12] and a multigrid least-squares algorithm [13]. Another minimum-variance solver that followed is the Fractal Iterative Method (FrIM) [14, 15] which performs a change of basis (using a fractal preconditioner). Minimum-variance

Travis W. Axtell
Department of the Navy, USA Emerson Cristiano Barbano, USA

reconstruction is an excellent choice, as it is optimal in the sense of maximizing the Strehl ratio [16].

In the last few years, several new wavefront reconstruction algorithms have been proposed. Rosensteiner has produced the Cumulative Reconstructor (CuRe), which is a direct integration reconstructor [17]. More recently, de Visser has shown the SABRE algorithm using B-spline basis functions [18].

Wavelets were first applied to wavefront reconstruction by Dowla [19]. This original work did not fully exploit the features of the discrete wavelet transform, and was an approximation. Peter Hampton and colleagues developed an algorithm that used the complete discrete wavelet transform and was able to perform reconstruction using the Haar wavelet [20, 21] in conjunction with a Poisson smoother or a de-noiser as a follow-on step [22].

In this research, we present a wavelet based phase reconstruction algorithm which can be extended to other wavelets other than the Haar. This results in an algorithm with complexity $O(n)$ and, for certain classes of wavelets (such Haar or Biorthogonal) all operations involve integer arithmetic. In the particular case of the simplest Haar wavelet the algorithm can be implemented with no multiplications, only additions and subtractions.

A major limitation of this approach is the assumption that the 2D data has to be irrotational (i.e. with zero curl) and arranged in a square matrix of dimensions $N \times N$, with N a power of 2. In most applications of interest data are collected by non-square apertures, usually circular, and also are based on measured wrapped observations of local phase variations.

The issue of non-square aperture is addressed in this research by defining the aperture by a binary correcting mask, applied to the square matrix assumed by the algorithm.

The problem of rotational data, (with non-zero curl) is of particular interest in systems which require long, horizontal paths through the atmosphere. In this case the cumulative turbulence can create branch points in the phase function, which are located in areas of low amplitude intensity of the Electric field. Branch points were first observed by Nye and Berry in 1974 [23].

The original work in reconstruction when branch points are present was done by Fried and Vaughn [24], followed by Fried [25] and Tyler [26] on the analysis work of branch points in wavefront reconstruction. In both works, the observed gradient phase field is decomposed into irrotational and rotational components, associated to a scalar (irrotational) and a vector (rotational) potentials. Both components are orthogonal to each other and the rotational component is hidden in the null space of the standard least squares reconstructor. Follow on works by Fried include [25] and [27]. A different approach by Tyler [26], is based on the Fourier Transform to characterize the phase in terms of scalar and vector potential functions.

In this chapter, we address the problem by adding a correcting factor to the rotational vector field, to make it irrotational, thus solving for the scalar potential function. It is shown in [28] that, in the case of wrapped phase observations, the phase error estimation is a multiple of π and in principle it can be eliminated with further processing.

14.2. Theoretical Framework

14.2.1. Two-Dimensional Signal Processing

The derivation for wavelet phase reconstruction requires some building blocks from multi-dimensional signal processing. We begin by defining a 2D sequence, $x[n_1, n_2]$, which is a known quantity at each lattice point $[n_1, n_2]$. We define the lattice to be finite in extent and have dimension lengths of powers of 2, usually of the form $2^N \times 2^N$.

We define two unit-shift operators, z_1 and z_2 , where the subscript denotes either the n_1 or n_2 direction. These definitions arise from the two-dimensional z -Transform

$$X(z_1, z_2) = \sum_{n_1=-\infty}^{\infty} \sum_{n_2=-\infty}^{\infty} x[n_1, n_2] z_1^{-n_1} z_2^{-n_2}. \quad (14.1)$$

A two-dimensional filter $H(z_1, z_2)$ is separable if it can be factored as two functions of a single variable as in $H(z_1, z_2) = H_1(z_1)H_2(z_2)$. To express the two-dimensional filtering convolution [29], we can write

$$y = H(z_1, z_2)x[n_1, n_2] = H_1(z_1)H_2(z_2)x. \quad (14.2)$$

To simplify notation, we use operator notation and disregard the indices. Eq. (14.2) is written with the sequence on the right side as the operand. When multiple operators are written, they are performed from right to left starting from the operand. The presence of a filter in operator notation implies the convolution operation (traditionally expressed as $h * x$ for two-dimensional convolution). Since the filter is separable, we observe that $H_1(z_1)H_2(z_2)x = H_2(z_2)H_1(z_1)x$ and that the order of operation does not change the final result.

The algorithm proposed here is performed on a multi-grid and it is multi-resolution [30]. To move between the resampled lattices, the downsample and upsample operators are used. We define downsampling by a factor of 2 as:

$$\begin{aligned} y_1 &= D_1 x & \Leftrightarrow & y_1[n_1, n_2] = x[2n_1, n_2], \\ y_2 &= D_2 x & \Leftrightarrow & y_2[n_1, n_2] = x[n_1, 2n_2], \\ y_3 &= D_1 D_2 x = D_2 D_1 x & \Leftrightarrow & y_3[n_1, n_2] = x[2n_1, 2n_2]. \end{aligned} \quad (14.3)$$

The downsampling operator discards the odd-index entries for the *downsampling phase* of 0 and discards the even-index entries for downsampling phase of 1, the latter of which is not shown in Eq. (14.3). The use of phase in the terminology is not wavefront phase – rather, it is in the sense of periodicity in the signal processing literature.

Likewise, the upsampling operator can be defined as

$$\begin{aligned} y_1 = U_1 x &\Leftrightarrow \begin{cases} y_1[2n_1, n_2] = x[n_1, n_2] \\ y_1[2n_1 + 1, n_2] = 0 \end{cases} \\ y_2 = U_2 x &\Leftrightarrow \begin{cases} y_2[n_1, 2n_2] = x[n_1, n_2] \\ y_2[n_1, 2n_2 + 1] = 0 \end{cases} \end{aligned} \quad (14.4)$$

for each dimension. Eq. (14.4) shows the relationship for *upsampling phase* of 0; swap the $x[n_1, n_2]$ and 0 to show upsampling phase of 1.

Eq. (14.4) is also separable and the relationship $y = U_1 U_2 x = U_2 U_1 x$ shows that the order of operations can be swapped. Although downsampling and upsampling by any positive integer factor is possible, this algorithm only uses a factor of 2. Thus, both operator forms do not show the resampling factor, only the dimension in which the operator performs.

Separable filters have useful properties to simplify implementation by dealing with only one dimension at a time. However, when a serial grouping of operators perform in the same dimension, in general the order of operations cannot be changed arbitrarily. This restriction is shown in Fig. 14.1. We must use the Noble identities to correctly establish the relationship in instances such as

$$\begin{aligned} y_1 &= D_1 H(z_1^2) x = H(z_1) D_1 x, \\ y_2 &= D_2 H(z_2^2) x = H(z_2) D_2 x. \end{aligned} \quad (14.5)$$

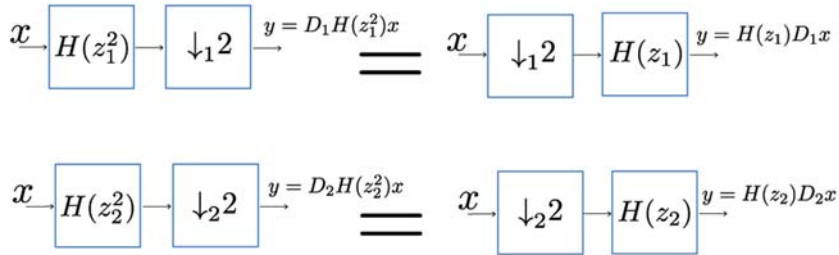


Fig. 14.1. Noble Identities in block diagram showing equivalence of Eq. (14.5) for both dimensions. The serial combination of two operations that both occur in the same dimension cannot change order without changing the filtering operation.

14.2.2. Tree Structure and Two-Dimensional Quadrature Mirror Filters

A tree structure is a signal processing technique to decompose a single-dimensional signal, $x[n]$, into multiple channels. In the case of this algorithm, each branch has two channels. In multi-rate systems, the channels are usually a low-pass and high-pass filters.

A special circumstance of the tree structure is referred as a Quadrature Mirror Filter (QMF) shown in Fig. 14.2. It is comprised of four filters, two used for analysis (with tildes) and two used for synthesis (without tildes). We can express the output as a function of the input using

$$y = \left(G(z)UD\tilde{G}(z) + H(z)UD\tilde{H}(z) \right) x. \quad (14.6)$$

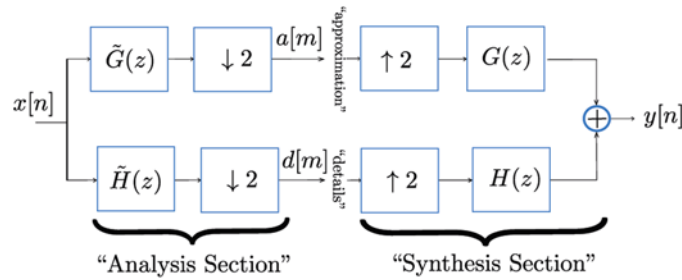


Fig. 14.2. Tree structure of a Quadrature Mirror Filter for single dimensional signal $x[n]$. Tree structures with the perfect reconstruction property result in $y[n]$ being equivalent to $x[n]$.

The channel with $\tilde{G}(z)$ is a low-pass filter and the channel with $\tilde{H}(z)$ is a high-pass filter.

The filters that are used in a QMF system are specifically designed in a manner to cancel out aliasing effects due to non-ideal frequency response of the filters such that

$$G(z)UD\tilde{G}(z) + H(z)UD\tilde{H}(z) = 1, \quad (14.7)$$

which means that the output y is a perfect reconstruction of x . With the perfect reconstruction property, the addition in Eq. (14.7) cancels the aliasing. Details on designing Quadrature Mirror Filters and their applications are found in numerous references such as [31].

In the case of a 1D signal $x[n]$ the two channels $x_L[m]$ and $x_H[m]$ are associated to the lower half and the upper half of the spectrum respectively and they are usually named “approximation” (x_L) and “details” (x_H) at half the number of samples.

In the case of a 2D signal $x[n_1, n_2]$ the decomposition is performed along the two dimensions independently and it yields four channels $x_{LL}[n_1, n_2], x_{LH}[n_1, n_2], x_{HL}[n_1, n_2], x_{HH}[n_1, n_2]$ as shown in Fig. 14.3.

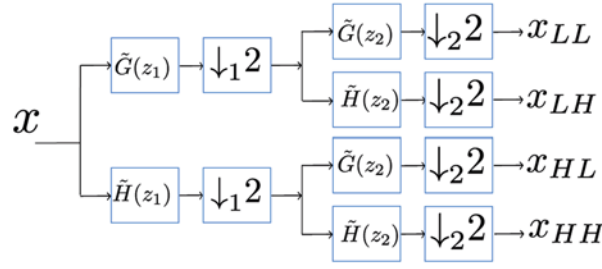


Fig. 14.3. Two-dimensional Quadrature Mirror Filter for the analysis section.

The four output channels can be described similar to the QMF terminology where x_{LL} is the approximation and x_{LH} , x_{HL} and x_{HH} are the details. Fig. 14.3 depicts the two-dimensional discrete wavelet transform.

Likewise, if the four channels are known, we can perform the two-dimensional inverse discrete wavelet transform to perfectly reconstruct the original two-dimensional sequence.

14.2.3. Wavelets in Tree Structure and Factoring Wavelets

The condition expressed in (14.7) relating the four filters guarantees perfect reconstruction of the signal by the QMF structure. Filters satisfying this condition belong to a number of classes having different properties and are associated to the Wavelet decomposition of a signal. There are many classes of wavelets based on their inherent properties including orthogonal, vanishing moments and others. The proposed algorithm relies on the orthogonal property, which means the Haar, Daubechies family and Biorthogonal family of wavelets can be used.

The Haar wavelet is the simplest and the low-pass and high-pass filters are defined as:

$$\begin{aligned} g(z) &\triangleq \frac{1}{\sqrt{2}}(1 + z^{-1}) \\ g(-z) &\triangleq \frac{1}{\sqrt{2}}(1 - z^{-1}) \end{aligned} \quad (14.8)$$

To define the wavefront reconstruction algorithm, we will make use of the following factorization property of the the Haar high-pass filter:

$$\begin{aligned} g(-z^N) &= \left(\sum_{\ell=0}^{N-1} z^{-\ell} \right) g(-z) \quad \forall N \geq 1 \\ &= \left(\sum_{\ell=0}^{\frac{N}{2}-1} z^{-2\ell} \right) \sqrt{2} g(z) g(-z) \quad \text{if } N \text{ is even.} \end{aligned} \quad (14.9)$$

In general it can be shown that the filters associated to all classes of wavelets have the Haar filters as factors, as follows

$$\begin{aligned} G(z) &= g(z)G_0(z), \\ H(z) &= g(-z)H_0(z), \\ \tilde{G}(z) &= g(z)\tilde{G}_0(z), \\ \tilde{H}(z) &= g(-z)\tilde{H}_0(z). \end{aligned} \tag{14.10}$$

In the particular case of the Haar wavelet, the factoring can be $G_0(z) = -H_0(z) = 1$, $\tilde{G}_0(z) = \tilde{H}_0(z) = z$. The factoring of Eq. (14.10) makes phase reconstruction possible for wavelet families with the orthogonal property.

14.2.4. Relationship of Wavelets to Shack-Hartmann Measurements

The Fried model of a Shack-Hartmann sensor [3] is shown in Fig. 14.4. The model defines the relationship between slope measurements and phase point values as

$$\begin{aligned} X_F[m, n] &= \frac{1}{2}(\Phi[m, n-1] - \Phi[m-1, n-1] + \Phi[m, n] - \Phi[m-1, n]), \\ Y_F[m, n] &= \frac{1}{2}(\Phi[m-1, n] - \Phi[m-1, n-1] + \Phi[m, n] - \Phi[m, n-1]). \end{aligned} \tag{14.11}$$

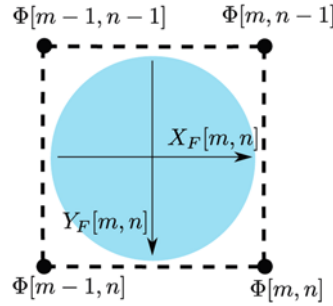


Fig. 14.4. The Fried geometry relationship between phase points and their measured slopes lattice for a single lenslet. A Shack-Hartmann sensor will have an array of these lenslets. Note that the vector field with magnitudes X_F and Y_F has a relationship to the scalar field Φ .

Eq. (14.11) was rewritten from its original form in [3] to appear in causal form. By doing so, close observation of Eq. (14.11) reveals that the slope measurements can be written as a separable filtering operation on the phase points. It involves one low pass filter and one high pass filter and is stated in operator form as

$$\begin{aligned} X_F &= g(-z_2)g(z_1)\Phi, \\ Y_F &= g(-z_1)g(z_2)\Phi. \end{aligned} \tag{14.12}$$

The relationship of Eq. (14.12) is the connection of the Haar wavelet to the Fried geometry; X_F and Y_F are the slope measurements. Fig. 14.5 depicts the block diagram relationship of the Hudgin [4] and Fried geometries to the wavefront. However, as we will see in Section 14.4.1, there is not a direct solution to solve for Φ based on these equations. In order to reconstruct Φ from the slope measurements, the tree structure 2D QMF is employed. We are then able to specify some additional information to reconstruct the wavefront.

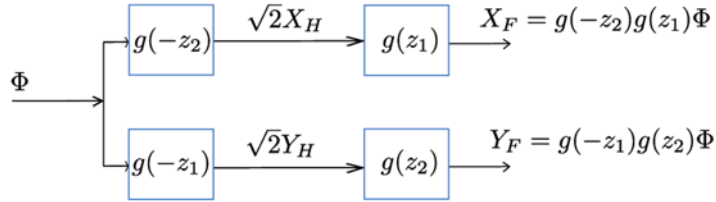


Fig. 14.5. The block diagram relationship between the phase points and the Hudgin and Fried geometries.

14.3. Phase Reconstruction Algorithm

In this section, we review a wavelet phase reconstruction algorithm that we first presented in [32] as an extension of the work done by Hampton, et al [20, 21].

14.3.1. Iteration for Level 1

From observation of the 2D QMF in Fig. 14.3, we can write equations which relate the phase points to the four channels of the 2D QMF as

$$\begin{aligned}
 \phi_{LL}^1 &= D_2 D_1 \tilde{G}(z_2) \tilde{G}(z_1) \Phi, \\
 \phi_{LH}^1 &= D_2 D_1 \tilde{H}(z_2) \tilde{G}(z_1) \Phi, \\
 \phi_{HL}^1 &= D_2 D_1 \tilde{G}(z_2) \tilde{H}(z_1) \Phi, \\
 \phi_{HH}^1 &= D_2 D_1 \tilde{H}(z_2) \tilde{H}(z_1) \Phi.
 \end{aligned} \tag{14.13}$$

We have swapped the order of the operators in Eq. (14.13) for a convenience of notation and use the superscript 1 to denote the first iteration, not an exponent. Using the factoring of Eq. (14.10) in Eq. (14.13), we can make the substitutions

$$\begin{aligned}
 \phi_{LH}^1 &= D_2 D_1 \tilde{H}_0(z_2) \tilde{G}_0(z_1) (g(-z_2)g(z_1)\Phi) \\
 &= D_2 D_1 \tilde{H}_0(z_2) \tilde{G}_0(z_1) X_F, \\
 \phi_{HL}^1 &= D_2 D_1 \tilde{G}_0(z_2) \tilde{H}_0(z_1) (g(z_2)g(-z_1)\Phi) \\
 &= D_2 D_1 \tilde{G}_0(z_2) \tilde{H}_0(z_1) Y_F.
 \end{aligned} \tag{14.14}$$

The remaining two quantities, ϕ_{LL}^1 and ϕ_{HH}^1 , do not have a simple relationship to the slope measurements X_F and Y_F , because the filters of Eq. (14.13) are both low pass or high pass, whereas the slope definitions require one of each.

If they did have a trivial relationship, then reconstruction would be a simple 2D inverse discrete wavelet transform. We need to further explore ϕ_{LL}^1 and ϕ_{HH}^1 in order to determine if such a reconstruction is possible.

14.3.2. Iteration for Level 2

In this algorithm, when there is an unknown quantity, we apply the 2D DWT and break apart the unknown quantity into 4 new channels. We now explore the second iteration to determine if any further substitution of Φ is possible with known slope measurements.

Because we have two unknown quantities, we must do this for both the ϕ_{LL}^1 and ϕ_{HH}^1 channels. First we will only consider ϕ_{LL}^1 ; we start by writing out the expressions for the four channels in the analysis:

$$\begin{aligned}
 \phi_{LL/L}^2 &= D_2 D_1 \tilde{G}(z_2) \tilde{G}(z_1) \phi_{LL}^1 \\
 &= D_2 D_1 \tilde{G}(z_2) \tilde{G}(z_1) D_2 D_1 \tilde{G}(z_2) \tilde{G}(z_1) \Phi, \\
 \phi_{LH/L}^2 &= D_2 D_1 \tilde{H}(z_2) \tilde{G}(z_1) \phi_{LL}^1 \\
 &= D_2 D_1 \tilde{H}(z_2) \tilde{G}(z_1) D_2 D_1 \tilde{G}(z_2) \tilde{G}(z_1) \Phi, \\
 \phi_{HL/L}^2 &= D_2 D_1 \tilde{G}(z_2) \tilde{H}(z_1) \phi_{LL}^1 \\
 &= D_2 D_1 \tilde{G}(z_2) \tilde{H}(z_1) D_2 D_1 \tilde{G}(z_2) \tilde{G}(z_1) \Phi, \\
 \phi_{HH/L}^2 &= D_2 D_1 \tilde{H}(z_2) \tilde{H}(z_1) \phi_{LL}^1 \\
 &= D_2 D_1 \tilde{H}(z_2) \tilde{H}(z_1) D_2 D_1 \tilde{G}(z_2) \tilde{G}(z_1) \Phi.
 \end{aligned} \tag{14.15}$$

We are using the superscript 2 for the second iteration and add the /L subscript for the ϕ_{LL}^1 data. There is additionally a /H analysis for the ϕ_{HH}^1 data. For brevity, we will only state the HH results at the end of this section since their development follows the same analysis, but we emphasize the resulting expression is not exactly the same. Examining Eq. (14.15) reveals that while $\phi_{LL/L}^2$ is only comprised of low pass filters, the remaining three channels have a combination of low and high pass filters. If we factor the filters as in Eq. (14.10), we will again find substitutions with the measured slope data. For $\phi_{LH/L}^2$ we obtain

$$\begin{aligned}
 \phi_{LH/L}^2 &= D_2 D_1 \tilde{H}_0(z_2) g(-z_2) \tilde{G}(z_1) D_2 D_1 \tilde{G}(z_2) \tilde{G}_0(z_1) g(z_1) \Phi \\
 &= D_2 D_1 \tilde{H}_0(z_2) \tilde{G}(z_1) D_2 D_1 \tilde{G}(z_2) \tilde{G}_0(z_1) (g(-z_2^2) g(z_1) \Phi) \\
 &= D_2 D_1 \tilde{H}_0(z_2) \tilde{G}(z_1) D_2 D_1 \tilde{G}(z_2) \tilde{G}_0(z_1) g(z_2) (\sqrt{2} X_F),
 \end{aligned} \tag{14.16}$$

where the final step uses the high-pass filter simplification of Eq. (14.9). Using the same procedure, we can also solve for $\phi_{HL/L}^2$ as

$$\begin{aligned}\phi_{HL/L}^2 &= D_2 D_1 \tilde{G}(z_2) \tilde{H}_0(z_1) g(-z_1) D_2 D_1 \tilde{G}_0(z_2) g(z_2) \tilde{G}(z_1) \Phi \\ &= D_2 D_1 \tilde{G}(z_2) \tilde{H}_0(z_1) D_2 D_1 \tilde{G}_0(z_2) \tilde{G}(z_1) \left(g(-z_1^2) g(z_2) \Phi \right) \\ &= D_2 D_1 \tilde{G}(z_2) \tilde{H}_0(z_1) D_2 D_1 \tilde{G}_0(z_2) \tilde{G}(z_1) g(z_1) \left(\sqrt{2} Y_F \right).\end{aligned}\quad (14.17)$$

The final channel, $\phi_{HH/L}^2$, yields two possible simplifications

$$\begin{aligned}\phi_{HH/L}^2 &= D_2 D_1 \tilde{H}_0(z_2) g(-z_2) \tilde{H}(z_1) D_2 D_1 \tilde{G}(z_2) \tilde{G}_0(z_1) g(z_1) \Phi \\ &= D_2 D_1 \tilde{H}_0(z_2) \tilde{H}(z_1) D_2 D_1 \tilde{G}(z_2) \tilde{G}_0(z_1) \left(g(-z_2^2) g(z_1) \Phi \right) \\ &\text{or similarly} \\ &= D_2 D_1 \tilde{H}_0(z_1) \tilde{H}(z_2) D_2 D_1 \tilde{G}(z_2) \tilde{G}_0(z_1) \left(g(-z_1^2) g(z_2) \Phi \right).\end{aligned}\quad (14.18)$$

The two possible substitutions arise from the flexibility of having two high pass filters. Either simplification is exact when using noise-free data. Rather than choose one definition over the other, we take an average

$$\begin{aligned}\phi_{HH/L}^2 &= \frac{1}{2} D_2 D_1 \tilde{H}_0(z_2) \tilde{H}(z_1) D_2 D_1 \tilde{G}(z_2) \tilde{G}_0(z_1) g(z_2) \left(\sqrt{2} X_F \right) \\ &\quad + \frac{1}{2} D_2 D_1 \tilde{H}(z_2) \tilde{H}_0(z_1) D_2 D_1 \tilde{G}_0(z_2) \tilde{G}(z_1) g(z_1) \left(\sqrt{2} Y_F \right).\end{aligned}\quad (14.19)$$

The averaging of Eq. (14.19) allows for some robustness to noise in the slope measurements at very little computational cost. The 1/2 coefficient simply assumes additive white Gaussian noise. Correlation statistics analysis may provide a better coefficient. Using the same process for the HH data, we now state the results as

$$\begin{aligned}\phi_{LH/H}^2 &= D_2 D_1 \tilde{H}_0(z_2) \tilde{G}(z_1) D_2 D_1 \tilde{H}(z_2) \tilde{H}_0(z_1) g(-z_2) \left(\sqrt{2} Y_F \right), \\ \phi_{HL/H}^2 &= D_2 D_1 \tilde{G}(z_2) \tilde{H}_0(z_1) D_2 D_1 \tilde{H}(z_1) \tilde{H}_0(z_2) g(-z_1) \left(\sqrt{2} X_F \right), \\ \phi_{HH/H}^2 &= \frac{1}{2} D_2 D_1 \tilde{H}_0(z_2) \tilde{H}(z_1) D_2 D_1 \tilde{H}(z_2) \tilde{H}_0(z_1) g(-z_2) \left(\sqrt{2} Y_F \right) \\ &\quad + \frac{1}{2} D_2 D_1 \tilde{H}(z_2) \tilde{H}_0(z_1) D_2 D_1 \tilde{H}_0(z_2) \tilde{H}(z_1) g(-z_1) \left(\sqrt{2} X_F \right).\end{aligned}\quad (14.20)$$

We have now completed the derivation of the second iteration and show it in Fig. 14.6. While the equations look complex on paper, actual implementations are straightforward and efficient in processing performance. Every expression is simply a serial grouping of the filter-filter-downsample-downsample block.

14.3.3. Further Iterations

We are able to generalize the formulation for additional iterations and include the necessary information in Appendix Section B.

By developing this level k implementation, we are able to scale the algorithm for any power of 2 sized data quickly. This algorithm is possible due to the flexibility of the high pass filter combined with the simplification of Eq. (14.9); hence, the nomenclature wavelet phase reconstruction is appropriate. Due to these features, we are able to swap out definitions of channels with Φ using the known, measured quantities X_F and Y_F .

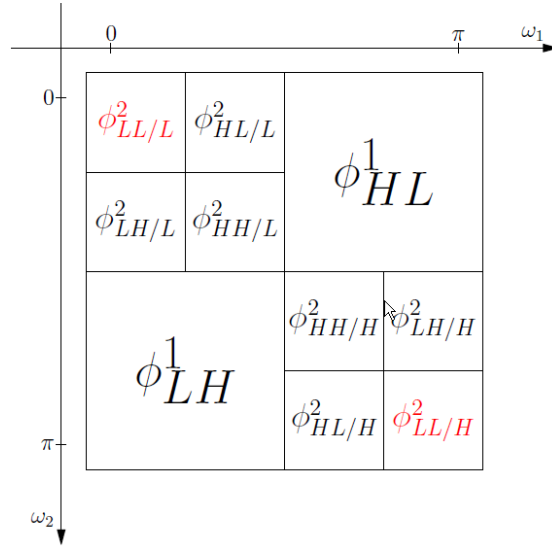


Fig. 14.6. The 2D QMF diagram of the channels at the second iteration. The upper left and lower right are each divided into four channels. The axes show digital frequencies of the Discrete Fourier Transform.

14.3.4. Setting the Mean and Waffle Values

At the final iteration, we have two sets of 2×2 matrices and no further downsampling is required. The upper-left scalar value of each is significant and represents the ϕ_{LL} channel that has gone undetermined for all prior iterations. Each value represent undetected modes of the Fried geometry: the piston and waffle modes. The piston represents the mean of the entire Φ data set. Since the Shack-Hartmann sensor only measures differences between phase points and not absolute values, the mean value cannot be known. We can assign it a value of zero and accept that we are within a constant value of the actual piston of the wavefront. A separate sensor is required for measuring piston. The waffle mode represents a nuisance checkerboard pattern along the phase points with a mean of zero. We show the completed 2D QMF analysis section structure in Fig. 14.7, where all values are known.

14.3.5. Synthesis Section

Up until this section, all of the previous algorithm code has been used to iteratively create the four channel blocks of the analysis section. We separated each unknown channel into four sub-channels. While we did not have the direct information for each channel, we were able to substitute for it using the measurements that were available. The analysis section is now complete and must now take the four channel blocks and perform the inverse discrete wavelet transform as shown in Fig. 14.8. In doing so, we recreate the unknown channel of the previous iteration. We recursively perform this until we have no more four channel blocks, which is the final solution of the wavefront phase surface.

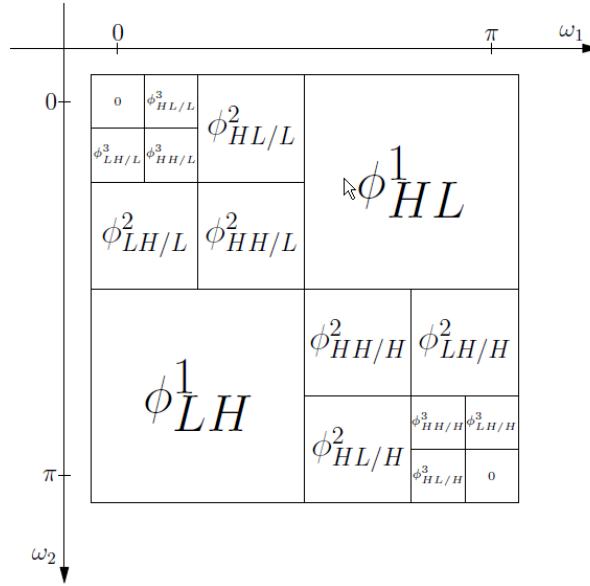


Fig. 14.7. The 2D QMF is performed iteratively on the upper left and lower right until scalar values remain.

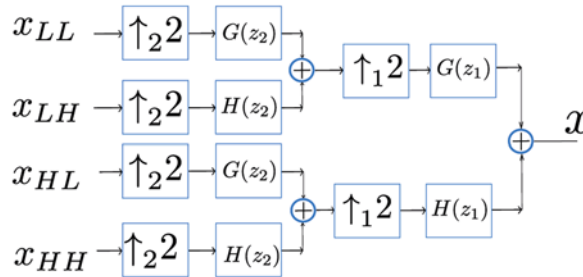


Fig. 14.8. Synthesis section of the 2D QMF.

14.4. Rotational Phase Fields

When the relative phase measurements have a non-zero curl, the standard wavefront reconstruction algorithms have a failure condition. Since this time, other approaches have addressed the rotational phase problem [24-27, 33, 34]. In this section, we will cover a new correction technique that enables standard wavefront reconstruction algorithms to function properly.

The key insight into this work is the following: The original definition of a Fried gradient in Eq. (14.11) did not include the realization of a phase wrapping operation. When the measured relative phase differences are small, the wrapping function is not necessary. However, in the presence of phase singularities, the wrapping can occur. By adding the wrapping operation into the definition of a Fried gradient, we obtain a more general purpose Fried gradient that models the correct behavior.

14.4.1. Continuous Vector Fields

The Helmholtz decomposition states that a vector field $\psi \in \mathbb{R}^3$ in the three dimensional space is represented by a gradient component and a rotational component as

$$\psi = \nabla\Phi + \nabla \times A, \quad (14.21)$$

where $\nabla = [\nabla_x, \nabla_y, \nabla_z]^T$ represents the gradient operator, $\Phi(x, y, z) \in \mathbb{R}$ is the scalar potential and $A(x, y, z) \in \mathbb{R}^3$ is the vector potential, with the outer product $\nabla \times$ defining the curl operation of the vector. In particular, in the case of a two-dimensional vector field in the x, y plane

$$\psi(x, y) = \begin{bmatrix} \psi_x(x, y) \\ \psi_y(x, y) \end{bmatrix}, \quad (14.22)$$

where we assume the component along the z axis to be identically zero, the scalar potential is $\Phi(x, y)$ and the vector potential is along the z axis as $A(x, y, z) = [0, 0, c(x, y)]^T$. This leads to a simple expression of the decomposition (14.22) in matrix form as

$$\begin{bmatrix} \psi_x(x, y) \\ \psi_y(x, y) \end{bmatrix} = \begin{bmatrix} \nabla_x & \nabla_y \\ \nabla_y & -\nabla_x \end{bmatrix} \begin{bmatrix} \Phi(x, y) \\ c(x, y) \end{bmatrix}. \quad (14.23)$$

14.4.2. Discrete Vector Fields

Like the previous section, the derivation for a discretely sampled vector field can be stated as:

$$\begin{bmatrix} X_F[n_1, n_2] \\ Y_F[n_1, n_2] \end{bmatrix} = \begin{bmatrix} \nabla_X & \nabla_Y \\ \nabla_Y & -\nabla_X \end{bmatrix} \begin{bmatrix} \Phi(n_1, n_2) \\ c(n_1, n_2) \end{bmatrix}. \quad (14.24)$$

Up to this point we assume that the measured gradients X_F and Y_F come from irrotational data, without singularities. However this is not always the case, and it has been shown that with rotational data (i.e. with nonzero curl) the phase cannot be reconstructed by simple least squares – this indicates that $c[n_1, n_2] \neq 0 \forall n_1, n_2$. In fact, the two-dimensional curl is defined as

$$c[n_1, n_2] \equiv \nabla_Y X_F[n_1, n_2] - \nabla_X Y_F[n_1, n_2]. \quad (14.25)$$

where we define the gradient operators to be

$$\begin{aligned} \nabla_X(z_1, z_2) &= \frac{1}{2} (1 - z_1^{-1}) (1 + z_2^{-1}), \\ \nabla_Y(z_1, z_2) &= \frac{1}{2} (1 + z_1^{-1}) (1 - z_2^{-1}) \end{aligned} \quad (14.26)$$

Notice how each definition of ∇_X and ∇_Y has a single difference. Any time there is a relative phase difference measurement, there is an implied phase wrapping operation. When the phase difference is less than 2π , the operator has no discernible change on the numerical difference; this is what explains why the consideration of the phase wrapping to be overlooked. In this section, we now define a broader definition of the Fried gradient which does have a wrapping operator after the difference operator. By operator linearity we can move the wrapping to the last step. We define the phase wrapping operation W as

$$\theta[\mathbf{n}] = W(\alpha[\mathbf{n}]) \equiv \alpha[\mathbf{n}] + 2\pi\ell[\mathbf{n}],$$

with $\ell[\mathbf{n}]$ indicating a generic 2D sequence of integer values only. In this case, choose $\ell[\mathbf{n}]$ values such that $-\pi \leq \theta[\mathbf{n}] < \pi \forall \mathbf{n}$. When the W operator is applied in a wavefront sensing context, it implies that phase differences of 2π are not sensed.

From Eq. (14.25), we can make an important observation: the curl is exclusively defined by measurements X_F and Y_F – from this information, we can seek a method to enforce $c[n_1, n_2] = 0 \forall n_1, n_2$ thus making the wavefront reconstruction algorithm work as intended. Stated another way, we want to move $c[n_1, n_2]$ to the left hand side of Eq. (14.24), but in working through the details a special mathematical trick must be applied. The general concept is that only one of the sets of measurements $X_F[n_1, n_2]$ or $Y_F[n_1, n_2]$ needs to be modified to make the curl equal to zero.

As stated previously, the mean and waffle modes are not detectable by the Fried geometry. Using the gradient operators it can be shown that, if for $n_1, n_2 \geq 0$:

$$\nabla_X \nabla_Y \Phi[n_1, n_2] = \nabla_X \nabla_Y \hat{\Phi}[n_1, n_2],$$

then

$$\Phi[n_1, n_2] = \hat{\Phi}[n_1, n_2] + C_0 + C_1(-1)^{n_1+n_2},$$

for arbitrary constants C_0 and C_1 . Thus, we can cast the original problem into an alternate form where we find a wrapped-equivalent wavefront $\hat{\Phi}$ of wavefront Φ . The vector field ψ can be expressed as

$$\begin{bmatrix} \psi_X(n_1, n_2) \\ \psi_Y(n_1, n_2) \end{bmatrix} = \begin{bmatrix} \nabla_X & 0 \\ \nabla_Y & -\nabla_Y \end{bmatrix} \begin{bmatrix} \Phi(n_1, n_2) \\ c(n_1, n_2) \end{bmatrix}.$$

with \bar{c} defined as

$$c[n_1, n_2] \equiv \nabla_X \nabla_Y \bar{c}[n_1, n_2]. \quad (14.27)$$

Call ∇W the wrapped Fried derivative gradient so that the two X and Y components are now defined as

$$\begin{bmatrix} X_F[n_1, n_2] \\ Y_F[n_1, n_2] \end{bmatrix} = \begin{bmatrix} \nabla W_X \\ \nabla W_Y \end{bmatrix} \Phi[n_1, n_2].$$

From the definition of the wrapping operator W and the factor $1/2$ in the averaging second block we can relate ∇ and ∇W as

$$\nabla W \Phi[\mathbf{n}] = \nabla \Phi[\mathbf{n}] + \pi \ell[\mathbf{n}], \quad (14.28)$$

with $\ell[\mathbf{n}]$ again a generic sequence of integers.

What distinguishes between rotational and irrotational fields in the previous sections is that we assumed the field to be irrotational, when the curl is zero. When this is the case the vectors X_F , Y_F are gradients of a function. When the curl is not zero, then there are singularity points and the arguments developed cannot be applied without further considerations. The following result is shown in [28]:

Main Result. Let $X_F[n_1, n_2]$, $Y_F[n_1, n_2]$ be the vector field of the wrapped Fried gradient of the phase $\Phi[n_1, n_2]$, defined as

$$\begin{bmatrix} X_F[\mathbf{n}] \\ Y_F[\mathbf{n}] \end{bmatrix} = \begin{bmatrix} \nabla W_X \\ \nabla W_Y \end{bmatrix} \Phi[\mathbf{n}].$$

Let the terms of its curl, $c[\mathbf{n}]$, due to phase wrapping, be zeros or integer multiples of π , as

$$c[\mathbf{n}] = \pi \ell[\mathbf{n}],$$

and call $\psi[n_1, n_2]$ the gradient vector augmented by the curl as follows

$$\begin{bmatrix} \psi_X[\mathbf{n}] \\ \psi_Y[\mathbf{n}] \end{bmatrix} = \begin{bmatrix} X_F[\mathbf{n}] \\ Y_F[\mathbf{n}] - c[\mathbf{n}] ** q[\mathbf{n}] \end{bmatrix}, \quad (14.29)$$

with $**$ denoting 2D convolution and $q[\mathbf{n}]$ defined as

$$q[n_1, n_2] = 2(-1)^{n_1} u[n_1] u[n_2],$$

where $u[n]$ is the discrete unit step function.

Then the following holds:

1. The curl of the vector fields ψ_X, ψ_Y is zero, i.e.

$$\nabla_Y \psi_X[n_1, n_2] - \nabla_X \psi_Y[n_1, n_2] = 0.$$

2. Therefore, the problem can be stated as a gradient:

$$\begin{bmatrix} \psi_X[\mathbf{n}] \\ \psi_Y[\mathbf{n}] \end{bmatrix} = \begin{bmatrix} \nabla_X \\ \nabla_Y \end{bmatrix} \hat{\Phi}[\mathbf{n}]. \quad (14.30)$$

3. There exist constants C_0, C_1 for which

$$\Phi[\mathbf{n}] = \hat{\Phi}[\mathbf{n}] + C_0 + C_1 (-1)^{n_1 + n_2} + 2\pi \ell[\mathbf{n}]. \quad (14.31)$$

Since the “waffle” term is zero, i.e. $C_1 = 0$, then this implies that there exists a constant C_0 such that

$$W\Phi[\mathbf{n}] = W(\hat{\Phi}[\mathbf{n}] + C_0). \quad (14.32)$$

Although a detailed explanation is given in [28], the fact that the vector field ψ_X, ψ_Y is irrotational (i.e. its curl is zero), is very simple, once we realize that the 2D z -Transform of $q[\mathbf{n}]$ is given by

$$q(z_1, z_2) = 2(1 - z_1^{-1})(1 + z_2^{-1}), \quad (14.33)$$

and therefore $\nabla_x q[\mathbf{n}] = \delta[\mathbf{n}]$, with $\delta[\mathbf{n}]$ the unit impulse.

The comparison of this algorithm with the traditional approach is given in Fig. 14.9.

14.5. Conclusion

In this chapter, we addressed the problem of estimating a phase signal based on observation of wrapped local measurements. This approach is based on a particular representation of the vector field in terms of a non-orthogonal basis which seems to be better suited than the standard orthogonal basis associated to scalar and potential fields.

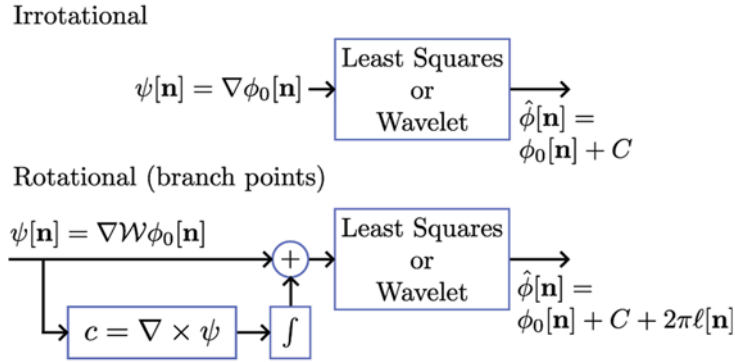


Fig. 14.9. The block diagram comparison of the traditional least-squares approach to the new form that is capable of handling branch points. When the curl is equal to zero, the output is exactly the same for both forms.

It is shown that, correcting the observed gradient with a filtered curl, the overall phase (including what has been called the “hidden phase”) is estimated by standard least-squares solver.

This approach is able to efficiently determine a wavefront surface that is a member of the ensemble of wavefronts that all have the same wavefront measurements. The approach is as computationally efficient as the least-squares or equivalent reconstructor chosen. The approach does not unwrap the phase, as we leave that as a follow on step to the output of our algorithm presented here.

References

- [1]. F. Roddier, *Adaptive Optics in Astronomy*, Cambridge University Press, 1999.
- [2]. P. Y. Bely (Ed.), *The Design and Construction of Large Optical Telescopes*, Springer, 2003.
- [3]. D. L. Fried, Least-square fitting a wave-front distortion estimate to an array of phase-difference measurements, *J. Opt. Soc. Am.*, Vol. 67, 1977, pp. 370–375.
- [4]. R. H. Hudgin, Wave-front reconstruction for compensated imaging, *J. Opt. Soc. Am.*, Vol. 67, 1977, pp. 375–378.

- [5]. W. H. Southwell, Wave-front estimation from wave-front slope measurements, *J. Opt. Soc. Am.*, Vol. 70, 1980, pp. 998–1006.
- [6]. J. Herrmann, Least-squares wave front errors of minimum norm, *J. Opt. Soc. Am.*, Vol. 70, 1980, pp. 28–35.
- [7]. K. Freischlad, Wavefront reconstruction from noisy slope or difference data using the discrete Fourier transform, *Proceedings of SPIE*, Vol. 551, 1985, pp. 74–80.
- [8]. L. A. Poyneer, D. T. Gavel, J. M. Brase, Fast wave-front reconstruction in large adaptive optics systems with use of the Fourier transform, *J. Opt. Soc. Am. A*, Vol. 19, 2002, pp. 2100–2111.
- [9]. L. Gilles, C. R. Vogel, B. L. Ellerbroek, Multigrid preconditioned conjugate-gradient method for large-scale wave-front reconstruction, *J. Opt. Soc. Am. A*, Vol. 19, 2002, pp. 1817–1822.
- [10]. D. G. MacMartin, Local, hierarchic, and iterative reconstructors for adaptive optics, *J. Opt. Soc. Am. A*, Vol. 20, 2003, pp. 1084–1093.
- [11]. L. Gilles, Sparse minimum-variance open-loop reconstructors for extreme adaptive optics: order N multigrid versus preordered Cholesky factorization, *Proceedings of SPIE*, Vol. 5169, 2003, pp. 201–205.
- [12]. C. R. Vogel, Sparse matrix methods for wavefront reconstruction, revisited, *Proceedings of SPIE*, Vol. 5490, 2004, pp. 1327–1335.
- [13]. C. R. Vogel, Q. Yang, Multigrid algorithm for least-squares wavefront reconstruction, *Appl. Opt.*, Vol. 45, 2006, pp. 705–715.
- [14]. C. Béchet, M. Tallon, E. Thiébaud, FRIM: minimum-variance reconstructor with a fractal iterative method, *Proceedings of SPIE*, Vol. 6272, 2006, pp. 62722U–62722U.
- [15]. E. Thiébaud, M. Tallon, Fast minimum variance wavefront reconstruction for extremely large telescopes, *J. Opt. Soc. Am. A*, Vol. 27, 2010, pp. 1046–1059.
- [16]. J. Herrmann, Phase variance and Strehl ratio in adaptive optics, *J. Opt. Soc. Am. A*, Vol. 9, 1992, pp. 2257–2258.
- [17]. M. Rosensteiner, Cumulative Reconstructor: fast wavefront reconstruction algorithm for Extremely Large Telescopes, *J. Opt. Soc. Am. A*, Vol. 28, 2011, pp. 2132–2138.
- [18]. C. C. de Visser, M. Verhaegen, Wavefront reconstruction in adaptive optics systems using nonlinear multivariate splines, *J. Opt. Soc. Am. A*, Vol. 30, 2013, pp. 82–95.
- [19]. F. U. Dowl, Fast Fourier and wavelet transforms for wavefront reconstruction in adaptive optics, *Proceedings of SPIE*, Vol. 4124, 2000, pp. 118–127.
- [20]. P. J. Hampton, P. Agathoklis, C. Bradley, A new wave-front reconstruction method for adaptive optics systems using wavelets, *IEEE J. Sel. Topics Signal Process*, Vol. 2, 2008, pp. 781–792.
- [21]. P. J. Hampton, Robust order N wavelet filterbanks to Perform 2-D numerical integration directly from partial difference or gradient measurements, PhD Thesis, *University of Victoria*, 2009.
- [22]. P. J. Hampton, P. Agathoklis, R. Conan, C. Bradley, Closed-loop control of a woofer-tweeter adaptive optics system using wavelet-based phase reconstruction, *J. Opt. Soc. Am. A*, Vol. 27, 2010, pp. A145–A156.
- [23]. J. F. Nye, M. V. Berry, Dislocations in wave trains, *Proceedings of the R. Soc. A*, Vol. 336, 1974, pp. 165–190.
- [24]. D. L. Fried, J. L. Vaughn, Branch cuts in the phase function, *Appl. Opt.*, Vol. 31, 1992, pp. 2865–2882.
- [25]. D. L. Fried, Branch point problem in adaptive optics, *J. Opt. Soc. Am. A*, Vol. 15, 1998, pp. 2759–2768.
- [26]. G. Tyler, Reconstruction and assessment of the least-squares and slope discrepancy components of the phase, *J. Opt. Soc. Am. A*, Vol. 17, 2000, pp. 1828–1839.
- [27]. D. L. Fried, Adaptive optics wave function reconstruction and phase unwrapping when branch points are present, *Optics Communications*, Vol. 200, 2001, pp. 43–72.

- [28]. R. Cristi, T. W. Axtell, Least-squares phase estimation with wrapped measurements and branch points, *J. Opt. Soc. Am A*, Vol. 30, 2013, pp. 2225–2236.
- [29]. D. E. Dudgeon, R. M. Mersereau, Multidimensional Digital Signal Processing, *Prentice-Hall*, 1984.
- [30]. W. L. Briggs, V. E. Henson, S. F. McCormick, A Multigrid Tutorial, *SIAM*, 2000.
- [31]. P. P. Vaidyanathan, Multirate Systems and Filter Banks, *Prentice-Hall*, 1993.
- [32]. T. W. Axtell, R. Cristi, Generalized orthogonal wavelet phase reconstruction, *J. Opt. Soc. Am. A*, Vol. 30, 2013, pp. 859–870.
- [33]. R. Goldstein, H. A. Zebker, C. L. Werner, Satellite radar interferometry: Two-dimensional phase unwrapping, *Radio Science*, Vol. 23, 1988, pp. 713–720.
- [34]. D. C. Ghiglia, M. D. Pritt, Two-Dimensional Phase Unwrapping: Theory, Algorithms, and Software, *Wiley*, 1998.

Appendix. Further Proof Details

A. High-Order Wavelet Simplification Proof

This proof shows how the results of Eq. (14.9) are determined. We start with the definition

$$g(-z^N) \triangleq \frac{1 - z^{-N}}{\sqrt{2}}, \quad (14.34)$$

and the definition of the geometric series

$$\sum_{\ell=0}^{N-1} z^{-\ell} \triangleq \frac{1 - z^{-N}}{1 - z^{-1}}. \quad (14.35)$$

We immediately observe that Eqs. (14.34) and (14.35) can be combined.

$$\begin{aligned} g(-z^N) &= \frac{1 - z^{-N}}{\sqrt{2}}, \\ &= \frac{\left(\sum_{\ell=0}^{N-1} z^{-\ell} \right) (1 - z^{-1})}{\sqrt{2}}, \\ &= \left(\sum_{\ell=0}^{N-1} z^{-\ell} \right) g(-z). \end{aligned} \quad (14.36)$$

We have now shown the first result. The second result takes some manipulation similar to the concept of polyphase decomposition where we split the sequence up into an even and odd component. We proceed from the result of Eq. (14.36) in

$$\begin{aligned}
 \left(\sum_{\ell=0}^{N-1} z^{-\ell} \right) g(-z) &= \left(\sum_{\ell=0}^{\frac{N-1}{2}} z^{-2\ell} + z^{-2\ell-1} \right) g(-z), \\
 &= \left((1+z^{-1}) \sum_{\ell=0}^{\frac{N-1}{2}} z^{-2\ell} \right) g(-z), \\
 &= \left(\sum_{\ell=0}^{\frac{N-1}{2}} z^{-2\ell} \right) \sqrt{2} g(z) g(-z).
 \end{aligned} \tag{14.37}$$

The first simplification is the realization that the sum of the two sequences can be factored to $1 + z^{-1}$. The final factoring swaps with the Haar scaling function and needs a $\sqrt{2}$ to cancel the denominator.

B. Iteration for Level k

After the second iteration, there are two unknown quantities ϕ_{LLL}^2 and ϕ_{LLH}^2 . We now seek to generalize the formulas for each level $k \geq 2$. Until the final iteration, there will still be two unknown quantities. We first write out the formulas

$$\begin{aligned}
 \phi_{LH/L}^k &= \left(D_2 D_1 \tilde{H}(z_2) \tilde{G}(z_1) \right) \left(D_2 D_1 \tilde{G}(z_2) \tilde{G}(z_1) \right)^{k-2} \\
 &\quad \left(D_2 D_1 \tilde{G}(z_2) \tilde{G}(z_1) \right) \Phi, \\
 \phi_{HL/L}^k &= \left(D_2 D_1 \tilde{G}(z_2) \tilde{H}(z_1) \right) \left(D_2 D_1 \tilde{G}(z_2) \tilde{G}(z_1) \right)^{k-2} \\
 &\quad \left(D_2 D_1 \tilde{G}(z_2) \tilde{G}(z_1) \right) \Phi, \\
 \phi_{HH/L}^k &= \left(D_2 D_1 \tilde{H}(z_2) \tilde{H}(z_1) \right) \left(D_2 D_1 \tilde{G}(z_2) \tilde{G}(z_1) \right)^{k-2} \\
 &\quad \left(D_2 D_1 \tilde{G}(z_2) \tilde{G}(z_1) \right) \Phi.
 \end{aligned} \tag{14.38}$$

The intent of the exponential notation is that the operations inside occur $k - 2$ times. We choose to express these equations as three groups since the left group will be used to modify the right group. Again we will factor the $\tilde{H}(z)$ on the left side, then move the $g(-z)$ to the right. As it is swapped position with the downsampling operators, the Noble identities will apply, resulting in $g(-z^{2^{k-1}})$. Then the high-order filter will be simplified

to a delayed summation of the first-order filter. The relationship to the slope measurements can then be made. The end result for the LL data is

$$\begin{aligned}
 \phi_{LH/L}^k &= \left(D_2 D_1 \tilde{H}_0(z_2) \tilde{G}(z_1) \right) \left(D_2 D_1 \tilde{G}(z_2) \tilde{G}(z_1) \right)^{k-2} \\
 &\quad \dots \left(D_2 D_1 \tilde{G}(z_2) \tilde{G}_0(z_1) \right) \\
 &\quad \dots \left(\left(\sum_{\ell=0}^{2^{k-1}-1} z_2^{-\ell} \right) X_F \right), \\
 \phi_{HL/L}^k &= \left(D_2 D_1 \tilde{G}(z_2) \tilde{H}_0(z_1) \right) \left(D_2 D_1 \tilde{G}(z_2) \tilde{G}(z_1) \right)^{k-2} \\
 &\quad \dots \left(D_2 D_1 \tilde{G}_0(z_2) \tilde{G}(z_1) \right) \\
 &\quad \dots \left(\left(\sum_{\ell=0}^{2^{k-1}-1} z_1^{-\ell} \right) Y_F \right),
 \end{aligned} \tag{14.39}$$

with a combination summation for the HH / L channel

$$\begin{aligned}
 \phi_{HH/L}^k &= \frac{1}{2} \left(D_2 D_1 \tilde{H}_0(z_2) \tilde{H}(z_1) \right) \left(D_2 D_1 \tilde{G}(z_2) \tilde{G}(z_1) \right)^{k-2} \\
 &\quad \left(D_2 D_1 \tilde{G}(z_2) \tilde{G}_0(z_1) \right) \left(\left(\sum_{\ell=0}^{2^{k-1}-1} z_2^{-\ell} \right) X_F \right) \\
 &+ \frac{1}{2} \left(D_2 D_1 \tilde{H}(z_2) \tilde{H}_0(z_1) \right) \left(D_2 D_1 \tilde{G}(z_2) \tilde{G}(z_1) \right)^{k-2} \\
 &\quad \left(D_2 D_1 \tilde{G}_0(z_2) \tilde{G}(z_1) \right) \left(\left(\sum_{\ell=0}^{2^{k-1}-1} z_1^{-\ell} \right) Y_F \right).
 \end{aligned} \tag{14.40}$$

The HH data is again developed through the same manner and results in definitions with some slight differences

$$\begin{aligned}
 \phi_{LH/H}^k &= \left(D_2 D_1 \tilde{H}_0(z_2) \tilde{G}(z_1) \right) \left(D_2 D_1 \tilde{G}(z_2) \tilde{G}(z_1) \right)^{k-2} \\
 &\quad \dots \left(D_2 D_1 \tilde{H}(z_2) \tilde{H}_0(z_1) \right) \\
 &\quad \dots \left(\left(\sum_{\ell=0}^{2^{k-2}-1} z_2^{-2\ell} \right) \sqrt{2} g(-z_2) Y_F \right),
 \end{aligned}$$

$$\begin{aligned}\varphi_{HL/H}^k &= \left(D_2 D_1 \tilde{G}(z_2) \tilde{H}_0(z_1) \right) \left(D_2 D_1 \tilde{G}(z_2) \tilde{G}(z_1) \right)^{k-2} \\ &\quad \dots \left(D_2 D_1 \tilde{H}_0(z_2) \tilde{H}(z_1) \right) \\ &\quad \dots \left(\left(\sum_{\ell=0}^{2^{k-2}-1} z_1^{-2\ell} \right) \sqrt{2} g(-z_1) X_F \right),\end{aligned}\tag{14.41}$$

and the final channel, HH / H , is again defined as a combination

$$\begin{aligned}\varphi_{HH/H}^k &= \frac{1}{2} \left(D_2 D_1 \tilde{H}_0(z_2) \tilde{H}(z_1) \right) \left(D_2 D_1 \tilde{G}(z_2) \tilde{G}(z_1) \right)^{k-2} \\ &\quad \dots \left(D_2 D_1 \tilde{H}(z_2) \tilde{H}_0(z_1) \right) \\ &\quad \dots \left(\left(\sum_{\ell=0}^{2^{k-2}-1} z_2^{-2\ell} \right) \sqrt{2} g(-z_2) Y_F \right) \\ &\quad + \frac{1}{2} \left(D_2 D_1 \tilde{H}(z_2) \tilde{H}_0(z_1) \right) \left(D_2 D_1 \tilde{G}(z_2) \tilde{G}(z_1) \right)^{k-2} \\ &\quad \dots \left(D_2 D_1 \tilde{H}_0(z_2) \tilde{H}(z_1) \right) \\ &\quad \dots \left(\left(\sum_{\ell=0}^{2^{k-2}-1} z_1^{-2\ell} \right) \sqrt{2} g(-z_1) X_F \right).\end{aligned}\tag{14.42}$$

The summations represent either a zero-padded shift or a circular shift of the data, and it should match the preferred implementation of how the sequences are treated for boundary conditions.

C. Branch Point Boundary Condition Proof

Let $w(x, y)$ be such that

$$\nabla_x \nabla_y w(x, y) = 0$$

for all x, y real. Then $w(x, y)$ can be written as

$$w(x, y) = a(x) + b(y).$$

To show this, define

$$g(x, y) = \nabla_x w(x, y) \quad (14.43)$$

Then $\nabla_y g(x, y) = 0$ and therefore,

$$g(x, y) = g(x, 0),$$

i.e., independent of y . Substitute in (14.43) to obtain

$$w(x, y) = w(0, y) + \int_0^x g(\lambda, 0) d\lambda,$$

which shows the result.

Chapter 15

Fresnel Nearfield Space-Grating Optics in the Human Retina Explains Human Color and Dimlight Vision

Norbert Lauinger

15.1. The Gap in the Physics of Colors

Newton demonstrated that the sunlight is split up into its rainbow colors due to refraction in a prism. At diffraction of white light on a line grating the colors of the rainbow will occur in the Fresnel nearfield behind the grating and this color sequence will also appear in the Fraunhofer farfield behind a one-layered planar grating. This is a sound stock of the physics of the colors. When at the beginning of the 20th century crystal optics demonstrated that by means of diffracting white x ray light into the farfield, the colors would again occur; however, in this instance in a completely new sequence and in the case of a high number of gratings in the crystal in a monochromatic manner, the triumph of the farfield space grating optics was able to get started with the help of the von-Laue-equation. Nearfield and farfield are of importance in acoustics and/or telecommunications engineering and display different physical characteristics. The nearfield is located directly behind the source of radiation and is marked by the interference effects of the waves, whereas the farfield is located at a greater distance behind and displays planar waves in space. It was possible to explain the structure of the elements and of many inorganic and organic molecules successfully by means of x-ray crystallography. Although Ewald [1] emphasized the validity of crystal optics also for visible light (380–760 nm) at an early stage in 1965 with the title of his essay “Crystal optics for visible light and x rays”, space grating optics did not achieve any success in this spectral range. Bergmann-Schäfer [2] described the situation as follows: “Diffraction at a space grating is of no importance in actual optics since there have been hardly any successes in creating a space grating of sufficient preciseness”. The gap in the physics of colors is located - as shown in Fig. 15.1 - in the space grating optics for the Fresnel nearfield. With it, it can be demonstrated that the RGB-colors in daylight vision and the luminosities in dimlight vision represent the product of a cortical layer of cell bodies located lightwards before the photoreceptors in

the ‘inverted’ retina. It will be possible to demonstrate that it is not photochemistry, but rather space grating optics which transforms the visible spectrum into the RGB-color space.

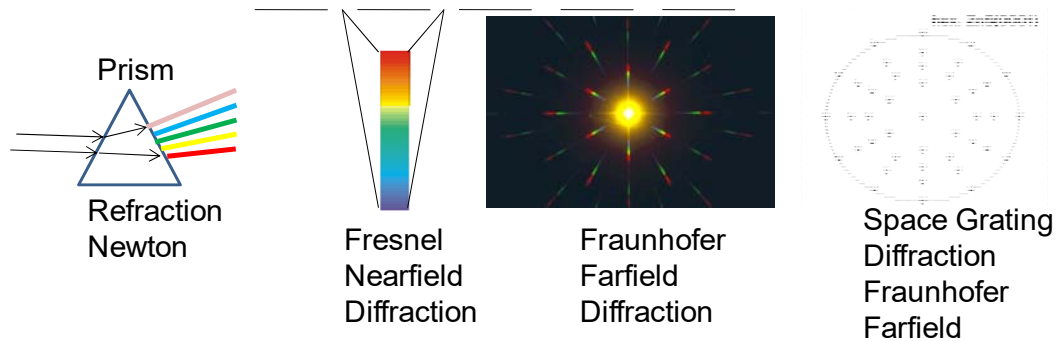


Fig. 15.1. The gap in the physics of colors. (left) refraction of white light in the prism; (center left and right) diffraction of white light on a one-layered planar grating in the nearfield and farfield; (right) diffraction of white x ray light by a crystal space grating into the Fraunhofer farfield.

The simple observation that the luminosity is not reduced to half upon closing one eye in binocular vision proves that we are concerned with ‘diffraction and interference ‘of light in this case. “The non-additivity of luminosities is an indication of interference” [Bergmann-Schäfer, 2]. It appears to be consistent that due to the immediate arrangement of the photoreceptors directly behind the retina, we can only be dealing with Fresnel nearfield interferences.

15.2. Human Color and Dimlight Vision

The tri-chromatic additive RGB- (RED-GREEN-BLUE) color vision at daylight in which colors complement each other to become white (Fig. 15.2, left) as well as the light-dark dimlight vision are to be explained first of all by means of space grating nearfield optics. The basis for this is represented by a new interpretation of the brain layer of the retina in which three layers of cell bodies achieve a layered image information processing which in turn is far more common in other cortical visual centers (Corpus Geniculatum Laterale CGL and V1).

The peak values of the photometrically measured absorption curves of the cone photo pigments were indicated by Rushton [3] in the visible spectrum at $R = 570 \text{ nm}$, $G = 535 \text{ nm}$ and $B = 445 \text{ nm}$. On this basis, it was proven that there are three types of cones in the photoreceptor layer of the retina which in turn are each arranged on one of these three RGB-components. Thus, it was possible to ascertain that photo chemistry is a determining value for RGB-color vision. With this it was also possible to reason that the color of things is only present “in the eye” and not on the objects themselves; that “the colors are a result of the activity of the retina” and “color being a function of the eye is of a subjective nature and consequently adheres to the eye directly and only secondarily and

indirectly to the objects” [Schopenhauer, 4]. It also became possible to understand dimlight vision on the basis of a comparable photometric measurement of the photo pigments in the outer segments of the rods. Thus, it appears to have been clarified that photo chemistry is responsible for the transformation of visible light into the RGB-color space. The same was valid for the perception of brightness in dimlight vision.

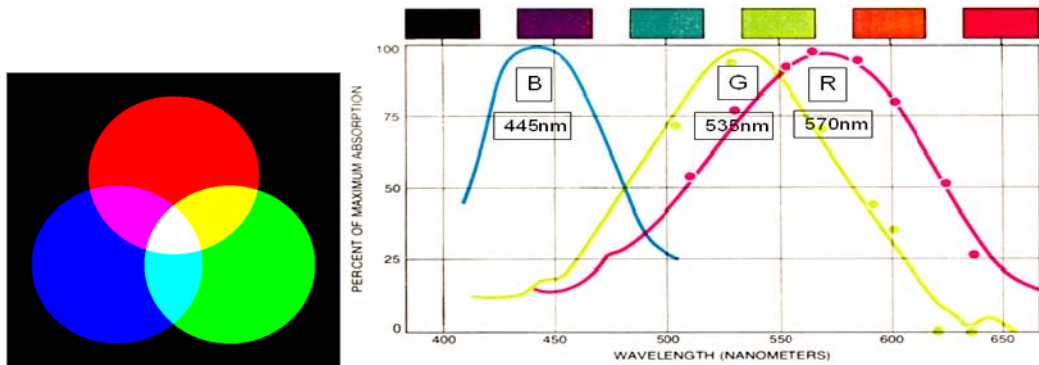


Fig.15.2. (left) the tri-chromatic additive RGB-color vision (Th. Young) and (right) the spectral brightness sensitivity curves of the cones (Rushton).

15.3. What a Space Grating Optical Explanation of Vision Ought To Be Able To Achieve

In order to supplement the explanations so far with a space grating optical component, it ought to be able to explain three central aspects of vision: the RGB-colors in daylight vision, the luminosities in dimlight vision and the relativization of the local colors on to the global light conditions in object space. While Fig. 15.2 provided the photometrically measured curves of the spectral brightness sensitivity of the cones, Fig. 15.3 (left) now shows the photometrically measured spectral brightness sensitivity curve of the rods in dimlight vision with a peak at 512nm. The so-called Purkinje-Shift describes how during the transition from daylight to dimlight vision, by means of the shift of the peak of the curves towards shorter wavelengths a blue flower will appear brighter and a red rose darker. Fig. 15.3 (right) documents the relativization of the local to the global RGB-values in a partial manner which E. Land [5] studied in the Retinex experiments. He was able to demonstrate that the eye perceives a gray or white upon any random Mondrian-Patch provided the RGB-values of the total illumination achieve the same percentages as the RGB-values of the radiation reflected from the patch to the eye and an RGB-equilibrium of 33 % of each is present. All percentage RGB-inequalities however, result in the perceptions of color. By means of the intensification of green in the total illumination (18.8 units) for example, the previously gray patch is perceived in the complementary color of green to be red ($R = 65\%$) provided the RGB-values reflected towards the eye are kept constant [Lauinger, 6].

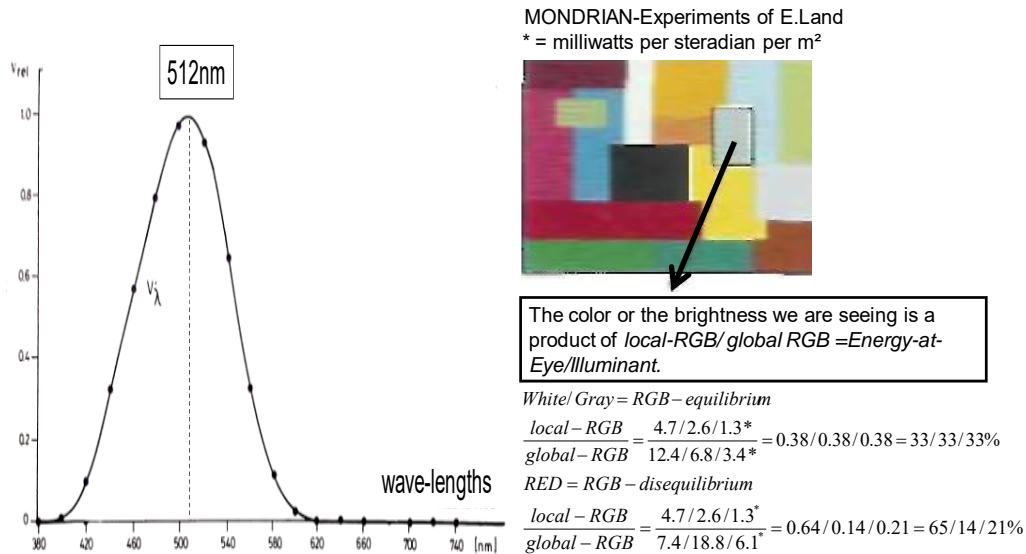


Fig. 15.3. (left) Spectral brightness sensitivity curve of the rod receptors; (right) the relativization of the local RGB-values on to the global RGB-values in object space.

15.4. The Development of the Cortical Retina to Become a Three-Layered Cell Body Grating

Up until the 7th pre-natal month there are no photoreceptors (cones and rods) in the eye; during this time, the brain is busy with the development of the cortical retina to become a cellular three-layered grating. Ida Mann [7] documented these developmental steps. Fig.15.4 (left) illustrates the development of two hemispheres at opposite sides of the embryonal forebrain which develop to become the eye stalk in each case. A cellular planar epithelium surrounds each eye stalk which is filled with liquid (Fig. 15.4, right). After the contact of the anterior ball of the eye stalk with the exterior skin of the body in the region of the face the invagination of the planar epithelium into the eye cavity follows. During this, a piece of tissue of the body is incorporated into the eye cavity which is then used for the development of the lens in the anterior part of the eye (Fig. 15.5, left and center). The retina develops from the invaginated planar epithelium. During this process, the cells of the planar epithelium produce further cells which while becoming smaller from layer to layer make up the second and third granular layer. Especially the third layer is extremely densely packed with cell bodies; it forms a hexagonal, cellular space grating which is present in all regions of the retina.

At an early stage, the optic nerves grow out of the cells of the inner layer - from the approx. one million ganglion cells. By means of cell divisions, a second layer with a larger number of smaller cell nuclei develops being the so-called MNL-layer of the bipolars. It in turn produces the third ONL-layer of even smaller cell bodies from which only after the seventh pre-natal month approx. 110 million rods and 6 million cones emerge into the space between the outer limiting membrane and the pigment epithelium (Fig. 15.5, right).

Thus far, this space had been empty. They are located behind the cortical cell gratings which form the so-called ‘inverted’ retina in a lightward direction in front of them. Fig. 15.6 shows the final retina development in the region of the fovea, the location with the most acute vision in daylight vision.

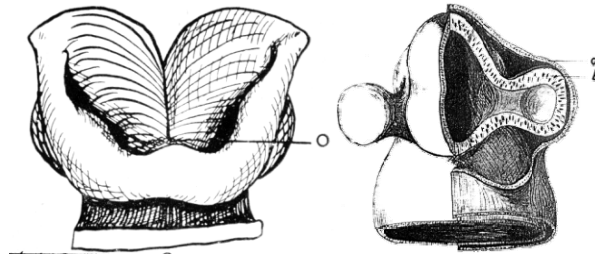


Fig. 15.4. (left) Forebrain with hemispheres (length of embryo 2.6mm) which (right) develop to become the eye stalk (embryo length 4mm = week 3-4) [I. Mann, 7, Fig.6, p.7 and Fig. 17, p.20]. © 2013 reproduced by permission of the British Medical Association.

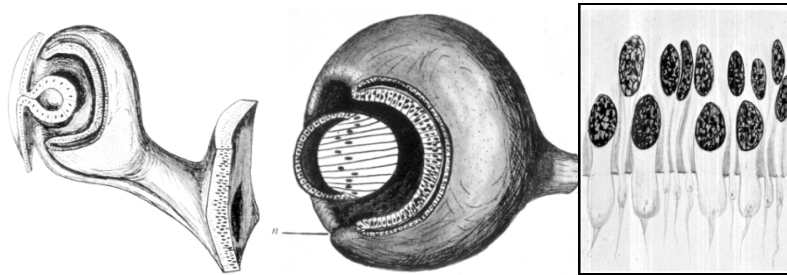


Fig. 15.5. (left) Invagination of the retina into the eye cavity taking a piece of boundary tissue of the body with it. (Center) Eye cavity with invaginated retina and developing lens [I. Mann, 7, Fig. 28, p. 31]; (right) Emerging of the photoreceptors from the ONL-layer of the retina [I. Mann, 7, Fig.83]. © 2013 reproduced by permission of the British Medical Association.

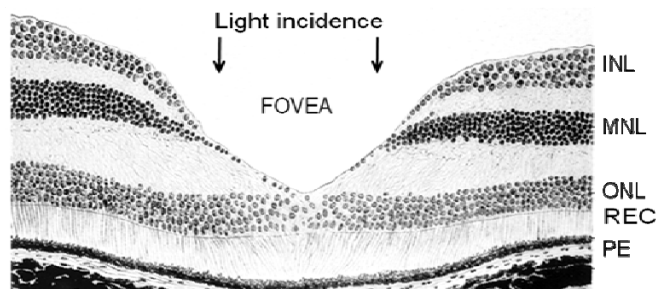


Fig. 15.6. Three-layered structure of the retina in the fovea region of a rhesus monkey ($V = 250\times$) (hematoxylin-eosine-staining) [Bargmann, 8, Fig.654, p.709]. INL = inner nuclear layer, MNL = middle nuclear layer, ONL = outer nuclear layer. REC = Photoreceptors, PE = pigment epithelium. Average thickness of the layer of the retina = 200-225 μm ; thickness of the photoreceptor layer approx. 50-70 μm . © 2013 with permission of Georg Thieme Verlag/Stuttgart.

Blechs Schmidt [9, Fig. 9] documented the differentiation of the layers during the prenatal developmental months. Fig. 15.7 shows the still undifferentiated accumulation of cell bodies (39 mm embryo length); in the center of the illustration there is the INL-layer (115 mm embryo length); in the right part of the illustration (retina of a new-born) all granular layers are differentiated and the photoreceptors ($V = 450\times$) are present in the central area of the retina perpendicular to the cell layers. Even in the Fovea the ONL-cell layer remains anterior in a lightward direction to the cones which are solely present there.

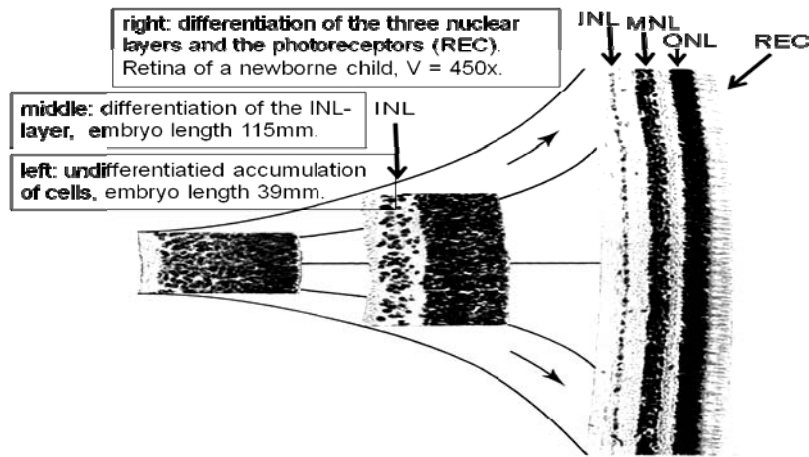


Fig. 15.7. Gradual differentiation of the retinal granular layers up until the time of birth.
Blechs Schmidt [9], (© 2013 with permission of S. Karger AG, Basel.

15.5. The Layered Processing of Information in the Cortical Visual Centers

It is proven that a layered processing of information takes place in the visual cortical centers CGL (Corpus Geniculatum Laterale) and V1 (area 17) which are located directly posterior to the retina. In the central optic tract, the right hemispheres of the two eyes are joined in the left and the left hemispheres in the right CGL after the optic chiasm and are now superimposed in $2 \times 3 = 6$ layers as shown in Fig. 15.8 (right). In V1, the information from both eyes and/or CGL-centers is joined and at the same time more layers and 'optical pillars' are differentiated. Still, also Ramon y Cajal provided evidence for three nuclear layers being located in V1 [Fig. 15.8, left]. It is noticeable in the CGL that $2 \times 2 = 4$ layers (3+5 as well as 4+6) of the sequence follow contra-ipsi, thus originating from the eye on opposite sides or the same side. Contrary to that, for the 3rd layer (1+2) the inverse sequence ipsi-contra from both eyes is valid.

There is no apparent reason, why the layered processing should not be valid in the brain layer of the retina also. Which functions can the individual layers take on? The assumption which for several reasons is the most plausible claims that the layers INL and MNL have functions which are already guaranteed in light-dark vision: the spatial frequency filtering (visual acuity performance), the 3D-depth map and the log-polar object classification [Lauinger, 6].

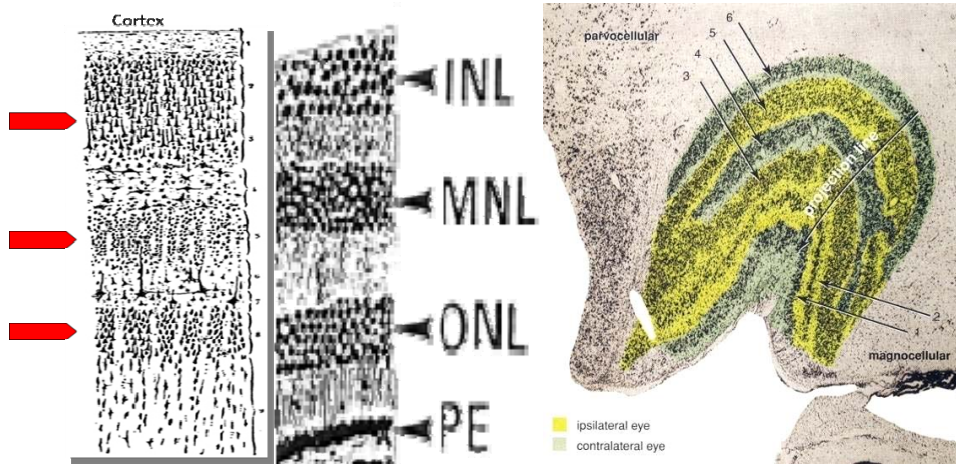


Fig. 15.8. (left) Three granular layers in the visual cortex (V1) [Ramon y Cajal, 10, Fig. 64, p. 151]. ©2013 by permission of Oxford University Press, USA). (Center) Three granular layers in the retina [Bargmann, 8, Fig. 654, p. 709], © 2013 permission by Georg Thieme Verlag/Stuttgart. (right) $2 \times 3 = 6$ granular layers in the CGL [Rodieck, 11, p. 282] © 2013 with permission of Sinauer Associates Inc., Sunderland, MA.

Thus, the functionality of color processing which is probably developed last in evolution can be attributed to the third granular layer, the cellular space grating ONL of the retina. To support it, the following Fresnel nearfield space grating calculation is presented.

15.6. The Calculation of the Achievements of the ONL Space Grating of the Retina

The densest packing of cell bodies having both a nucleus and cell plasma in the ONL layer of the retina leads to hexagonal gratings whose grating constants are located at the 2-10-fold of the wavelengths of the visible spectrum. The cell bodies are completely transparent for light. Diffraction of incident light in the space grating and interference of light in the Fresnel nearfield occur due to the fact that only differences in the refraction index between cell nucleus and plasma occur which transform the space grating into an optical phase grating. Fig. 15.9 illustrates how the Fresnel interferences on a simple line grating result in the so-called Talbot-light carpet. The interference levels nearest to the grating are the richest in contrast, the T/4-level displays a periodicity of interference maxima corresponding to a halving of the grating constants. In the case of the hexagonal space grating, the Fresnel carpet is replaced by a three-dimensional spatial interference-field. The fundamental periodicity is at $3g^2/2\lambda$.

Fig.15.10 (left) illustrates the hexagonal densest packing of cell bodies in a one-layered grating. The ONL-layer is shown in Fig. 15.10 (right) with 3 cell layers in the space grating. In contrast to the crystal gratings, only few grating layers are to be expected in the ONL-space grating so that also no strictly monochromatic interference maxima are to be expected in the diffraction image.

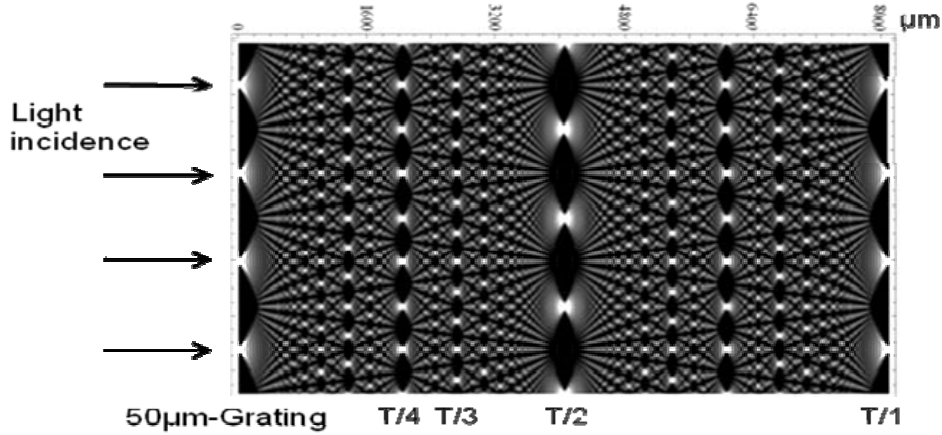


Fig. 15.9. Fresnel-/Talbot-/Lau-interference carpet at a line grating with the grating constant $g = 50 \mu\text{m}$ [Tomandl, 12]. © 2013 with permission of M. Tomandl, Vienna. In the case of a fundamental periodicity g^2/λ the $T/4$ level for the wavelength $\lambda = 625 \text{ nm}$ lies at $2000 \mu\text{m}$, the distance between the interference maxima there is at $25 \mu\text{m}$ being half of the grating constant.

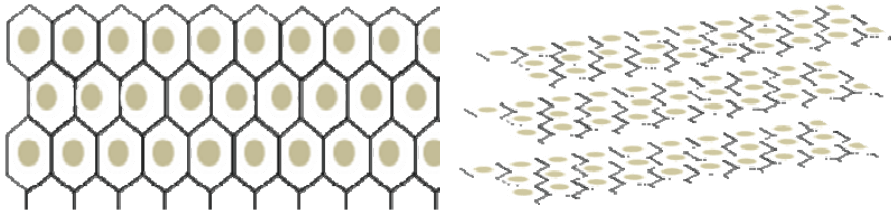


Fig. 15.10. (left) Hexagonal, cellular space grating; (right) hexagonal space grating as a 3rd grating in the construction of the granular layers of the retina.

The calculation of the diffraction orders of the hexagonal space grating starts with the double cone directional cosine in Fig. 15.11. The associated equations are:

$$\left(\cos \alpha - \cos \alpha^0\right)^2 + \left(\cos \beta - \cos \beta^0\right)^2 + \left(\cos \gamma - \cos \gamma^0\right)^2 = 1,$$

$$\left(\cos^2 \alpha + \cos^2 \beta + \cos^2 \gamma\right) - 2\left(\cos \alpha \cos \alpha^0 + \cos \beta \cos \beta^0 + \cos \gamma \cos \gamma^0\right) = 0,$$

$$\frac{2\left(\cos \alpha \cos \alpha^0 + \cos \beta \cos \beta^0 + \cos \gamma \cos \gamma^0\right)}{\cos^2 \alpha + \cos^2 \beta + \cos^2 \gamma} = 1.$$

In the cosine segments of the angles $\alpha^0, \beta^0, \gamma^0$ and α, β, γ and/or in their squared binomial segments, the light which is incident in focused light cones into the space grating is linked to the light which is diffracted into diffractive orders by the space grating. The diffractive orders are defined by means of whole number $h_1 h_2 h_3$ -triples. The fact that each resonant instrument must be tuned to a fundamental wave also applies to a space grating. The triple

for the diffraction order of the so-called fundamental wave therefore receives the triple-number $h_1h_2h_3 = 111$. If one chooses the angles $\alpha^\circ = \beta^\circ = 90^\circ$ and $\gamma^\circ = 0^\circ$ for $\alpha^\circ, \beta^\circ, \gamma^\circ$, then the directional cosine is reduced to the equation (15.1a), to the case of a light beam perpendicularly incident into the space grating. Equation (15.1b) is identical to (15.1a), but it emphasizes a different aspect of space grating optics which is to be proven using specific numbers.

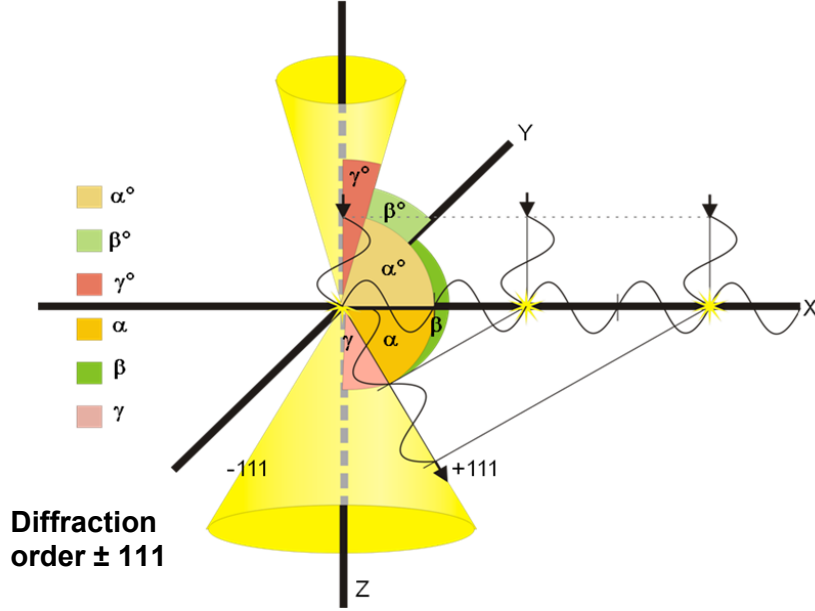


Fig. 15.11. Geometry of the double cone directional cosine in the Cartesian coordinate system.

$$\left(\cos^2 \alpha + \cos^2 \beta + \cos^2 \gamma \right) - 2 \cos \alpha \cos \beta \cos \gamma = 0, \quad (15.1a)$$

$$\frac{2 \cos \alpha \cos \beta \cos \gamma}{\cos^2 \alpha + \cos^2 \beta + \cos^2 \gamma} = 1. \quad (15.1b)$$

Another decisive step consists of the fact that instead of the classic von-Laue formulation of the $\cos \alpha, \beta, \gamma$ -segments, namely $\cos \alpha = h_1 \lambda / g_x$ etc., which is valid for the farfield, the reciprocal Fresnel periodicity with g/λ , thus $\cos \alpha = h_1 g_x / s$ ($s = k \lambda$, $k = \text{integer}$) is chosen. Fig. 15.12 (left) illustrates this.

The geometry of the hexagonal space grating cell is included in the equations with the three grating constants formulated in the Cartesian coordinate system (in each case in μm) $g_x: g_y: g_z = 2: \sqrt{3}: 1$. The spatial Pythagoras valid in the hexagonal space grating cell is illustrated in Fig. 15.12 (right).

$$\left(g_x^2 + g_y^2 + g_z^2 \right) - D^2 = (4 + 3 + 1) - 8 = 0$$

In the denominator of the equation (15.2a) λ is not selected, but rather s is taken to be the indicative reference. Its reference to λ and/or $v\lambda$ will soon be clarified. In equation (15.2b), the quantity s is in the numerator after the solution of (15.2a) according to s .

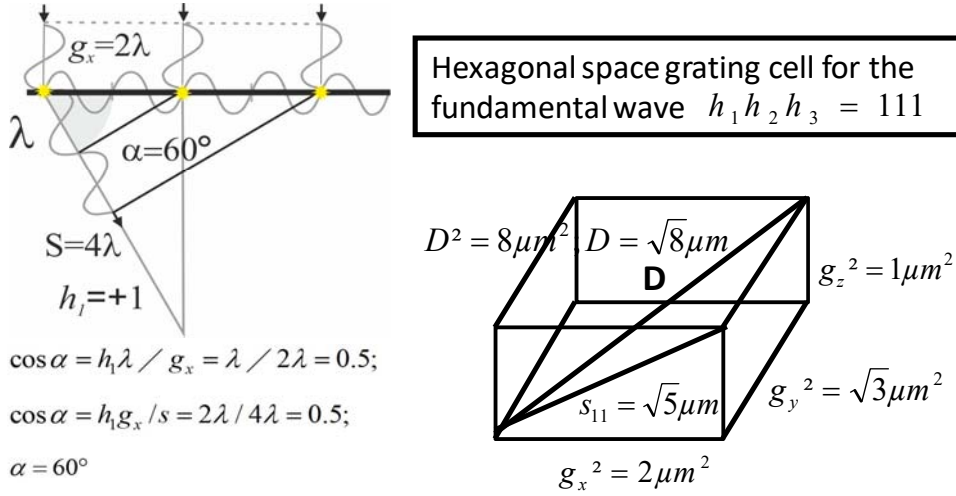


Fig. 15.12. (left) Fresnel formulation of the directional cosine instead of the von-Laue concept. (right) geometry of the hexagonal space grating cell.

With this Fresnel formulation, the double cone directional cosine (15.1a) and (15.1b) for the light beam incident perpendicularly into the space grating adopts the shape in (15.2a) and (15.2b).

$$\left(\frac{h_1 g_x}{s}\right)^2 + \left(\frac{h_2 g_y}{s}\right)^2 + \left(\frac{h_3 g_z}{s}\right)^2 - \frac{2h_3 g_z}{s} = 0, \quad (15.2a)$$

$$\frac{2h_3 g_z s}{h_1^2 g_x^2 + h_2^2 g_y^2 + h_3^2 g_z^2} = 1. \quad (15.2b)$$

The clarification of the reference quantity s is carried out by means of the crystal-optical resonance factor $v\lambda$ [Ewald, 1]. With the grating constant of the hexagonal geometry and the diffraction triple of the fundamental wave $h_1 h_2 h_3 = 111$, $s_{111} = 4 (\mu m)$ in (15.4) corresponds to the second term in $v\lambda = 1 (\mu m \mu m^{-1})$ in (15.3).

$$v\lambda = \frac{2h_3 g_z}{h_1^2 g_x^2 + h_2^2 g_y^2 + h_3^2 g_z^2} \times \frac{h_1^2 g_x^2 + h_2^2 g_y^2 + h_3^2 g_z^2}{2h_3 g_z} = 1, \quad (15.3)$$

$$s_{111} = \frac{h_1^2 g_x^2 + h_2^2 g_y^2 + h_3^2 g_z^2}{2h_3 g_z} = \frac{4 + 3 + 1}{2} = \frac{8}{2} = 4 \quad (15.4)$$

With $s = s_{111} = 4$ in (15.2a) and (15.2b) for the $h_1 h_2 h_3 = 111$ -triple of the fundamental wave λ_{111} the equations (15.5a) and (15.5b) result. In these, the two different aspects of the identical equation which were named (a) and (b) become clear. The a-version moves the spatial Pythagoras into the center whereas the b version centers upon the crystal-optical resonance factor. The spatial Pythagoras states that the three Cos^2 steps which lead from the original source of the interferences in the space grating to the location of the interference maxima in the Fresnel space must be identical with the Cosine^2 value above the space diagonals back to the source which the interference term $2h_3 g_z / s_{111} = 0.5$ in (15.5a) provides for the $h_1 h_2 h_3 = 111$ -triple of the fundamental wave λ_{111} . In (15.5b) however, it must always be: $v_{h_1 h_2 h_3} \lambda_{h_1 h_2 h_3} = v_{111} \lambda_{111} = 1$. Every admitted resonance wavelength must suffice to these two conditions in order to achieve a constructive interference with the help of the space grating.

$$\left(\frac{h_1 g_x}{s_{111}} \right)^2 + \left(\frac{h_2 g_y}{s_{111}} \right)^2 + \left(\frac{h_3 g_z}{s_{111}} \right)^2 - \frac{2h_3 g_z}{s_{111}} = (0.25 + 0.1875 + 0.0625) - 0.5 = 0.5 - 0.5 = 0 \quad (15.5a)$$

$$v_{111} \lambda_{111} = \frac{2h_3 g_z s_{111}}{h_1^2 g_x^2 + h_2^2 g_y^2 + h_3^2 g_z^2} = \frac{2}{8} \times \frac{8}{2} = 0.25 \times 4 = 1 \quad (15.5b)$$

With the results of these equations, the hexagonal space grating resonator has been fully described in its geometry. On this basis, the search for the resonator wavelength of the 111-fundamental wave and its harmonic waves may now begin. The geometry demands upon the space grating interference optics do not provide any information regarding the resonator or the resonance wavelengths in the space grating. As a resonant instrument, the space grating must first of all be tuned to a fundamental wave λ_{111} in order to disclose its harmonic waves. Since the hexagonal space grating g_y is strictly bound to g_x , only g_x and g_z represent free resonant parameters. Together they form the angled sender antenna of the resonator. Its product is the diagonal $s_{11} = \sqrt{(g_x^2 + g_z^2)} = \sqrt{5}$. If one transforms s_{111} upon λ_{111} and sets $s_{111} \lambda_{111} = s_{11}$, then the fundamental wavelength λ_{111} results in (15.6) with 559nm:

$$\lambda_{111} = \frac{2h_3 g_z s_{111} \lambda_{111}}{h_1^2 g_x^2 + h_2^2 g_y^2 + h_3^2 g_z^2} = \frac{2 \times 1 \times 1 \times 4 \lambda_{111}}{4 + 3 + 1} = \frac{2s_{11}}{8} = \frac{\sqrt{5}}{4} = 0.559 \mu\text{m} \quad (15.6)$$

In (15.6), if one varies the diffraction triples between the $h_1 h_2 h_3 = 111$ -333 (24 triples), then the R-fundamental wave $h_1 h_2 h_3 = 111$ is joined by only two harmonic waves with $h_1 h_2 h_3 = 123$ and $h_1 h_2 h_3 = 122$. These three resonance wavelengths make up the crucial RGB-triple. For these, the following is valid:

$$\lambda_{111} = s_{11} / s_{111} = \sqrt{5} / 4; \dots \lambda_{123} = s_{11} / s_{123} = \sqrt{5} / 4.167; \dots \lambda_{122} = s_{11} / s_{122} = \sqrt{5} / 5.$$

(Aside from the RGB-diffraction orders only one more $h_1h_2h_3 = 213$ -triple forms a resonance of poor intensity for the wavelength $\lambda = 479\text{nm}$. It is conceivable that it influences the RG-curves in Fig. 15.16 with their unexplained bend at the spectral location in an interfering manner).

The result that an RGB-triple is the sole determining factor of the chromatics of the diffractive orders, represents a completely unexpected result. However, it is commented for example in the space grating optics by Bergmann-Schäfer as follows: “If you have white light and thus a whole wavelength continuum at your availability then the space grating will select only the wavelength(s) fulfilling the resonance conditions” [Bergmann-Schäfer, 2. p.296].

$$R = \lambda_{111} = \frac{2 \times 1 \times 1 \times \sqrt{5}}{4 + 3 + 1} = \frac{2\sqrt{5}}{8} = \frac{\sqrt{5}}{4} = 0.25\sqrt{5} = 0.559\mu\text{m}$$

$$G = \lambda_{123} = \frac{2 \times 3 \times 1 \times \sqrt{5}}{4 + 12 + 9} = \frac{6\sqrt{5}}{25} = \frac{\sqrt{5}}{4.166} = 0.24\sqrt{5} = 0.537\mu\text{m}$$

$$B = \lambda_{122} = \frac{2 \times 2 \times 1 \times \sqrt{5}}{4 + 12 + 4} = \frac{4\sqrt{5}}{20} = \frac{\sqrt{5}}{5} = 0.20\sqrt{5} = 0.447\mu\text{m}$$

Fig. 15.13 illustrates the result of the calculations, the RGB-Fresnel nearfield transmission system with the space grating resonator as sender tuned to the R-fundamental wave $\lambda_{111} = \sqrt{5}/4 = 0.559\mu\text{m}$ and with the three RGB space waves in an RBG-sequence corresponding to the crystal-optical resonance in equation (15.6):

$$\nu_{111}\lambda_{111} = 0.25 \times 4 = 1, \nu_{123}\lambda_{123} = 0.24 \times 4.17 = 1 \text{ and } \nu_{122}\lambda_{122} = 0.20 \times 5 = 1.$$

The outer segments of the photoreceptors in the central region of the retina - exclusively cones - can be taken into consideration as RGB- receiver antennae.

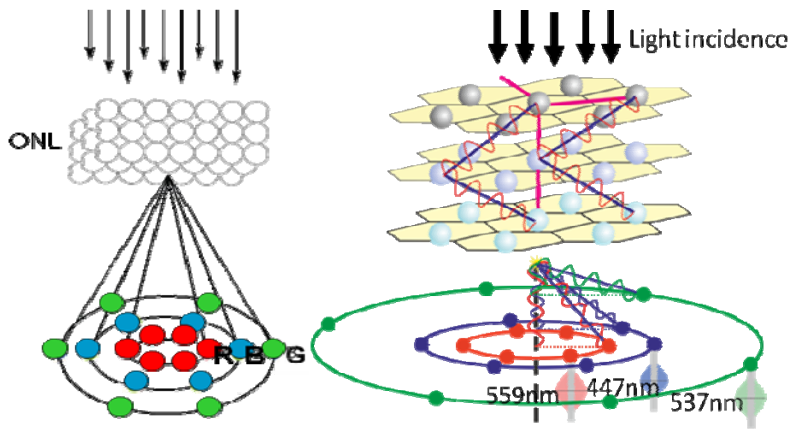


Fig. 15.13. (left) RGB-sequence of the diffraction orders; (right) the RGB-Fresnel nearfield transmission system with the space grating as RGB sender antenna, RGB-space waves and RGB receivers.

The four quantities forming the resonator have been drawn into the space grating: the three grating constants and the diagonal $s_{11} = \sqrt{(g_x^2 + g_z^2)} = \sqrt{5} = 4\lambda_{111}$. On the diagonal s_{11} , the sending antenna of the resonator, only R and B form integer waves with $s_{11} = 4\lambda_{111} = \sqrt{5} \mu\text{m}$; $s_{123} = 4.167\lambda_{123} = \sqrt{5} \mu\text{m}$; $s_{122} = 5\lambda_{122} = \sqrt{5} \mu\text{m}$. Only in the case of six-fold values $6\sqrt{5} = \sqrt{180} = 13.42 \mu\text{m}$ all the three standing RGB-waves consist of integers. The relationship of the RGB-wavelengths is at 1: 0.96: 0.8 = 25: 24: 20. The three RGB-wavelengths achieve an optimum transmission in the space grating, neighbored wavelengths are more strongly dimmed as the distance from the peak wavelengths increases. R is located on the most inner circle, B on the middle circle and G on the outer circle around the z-axis. Each circle has six maxima. As is valid in the von-Laue equation which applies to the Fraunhofer interferences, the $\cos \alpha, \beta, \gamma$ segments in the directional cosine describe three light cones centered upon the axes x, y, z which must intersect at a common point in the space. Accordingly, the interference maxima in the case of the hexagonal space grating are located in the intersection points of three pairs of hyperbolas which in turn are located on a circle. With reference to this, Bergmann-Schäfer [2, p.296] commented on a cubic space grating: "Brightness is only present where the two pairs of hyperbolas and the selected circles intersect at a point and this is generally not the case. Only if the wavelength λ has been selected correctly, this incident can occur; it can therefore no longer be selected at random". λ is "completely determined" by the geometric requirements of the directional cosine, i.e. the spatial Pythagoras. Fig. 15.14 (left) illustrates the product $v\lambda = 1$ as a constant and (right) the von-Laue cone requirement for farfield interferences which also applies to Fresnel nearfield interferences.

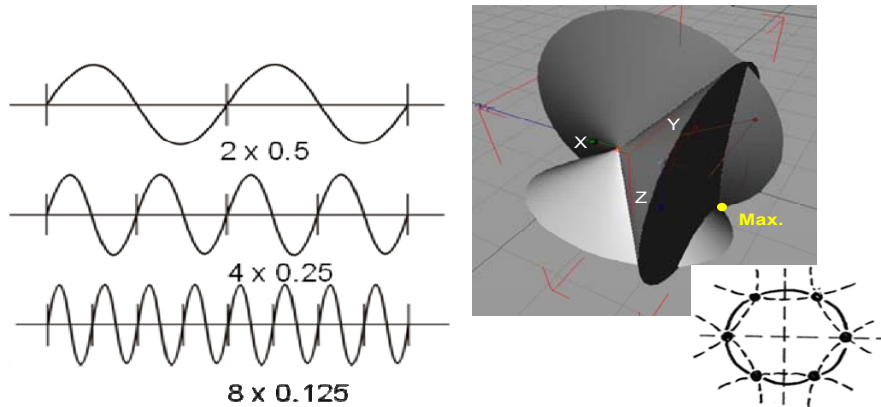


Fig. 15.14. (left) crystal-optical resonance factor as a $v\lambda$ -constant and (right) von-Laue requirement for the occurrence of resonance wavelengths.

Finally, Fig. 15.15 proves that the wavelength triple corresponds to the peak wavelengths of the spectral brightness sensitivity curves of the cones.

Just like the fundamental wave 559 nm (R) in equation (15.5a) and (15.5b), the two harmonic waves meet the requirements of the spatial Pythagoras and the crystal-optical resonance factor:

$$123(G): (0.2304+0.6912+0.5184) - 1.44 = 1.44 - 1.44 = 0,$$

$$122(B): (0.16 + 0.48 + 0.16) - 0.8 = 0.8 - 0.8 = 0 \text{ and}$$

$$\nu_{123}\lambda_{123}(G) = (2 \times 3 \times 4.166)/(4 + 12 + 9) = 0.24 \times 4.166 = 1.$$

$$\nu_{122}\lambda_{122}(B) = (2 \times 2 \times 5)/(4 + 12 + 4) = 0.20 \times 5 = 1$$

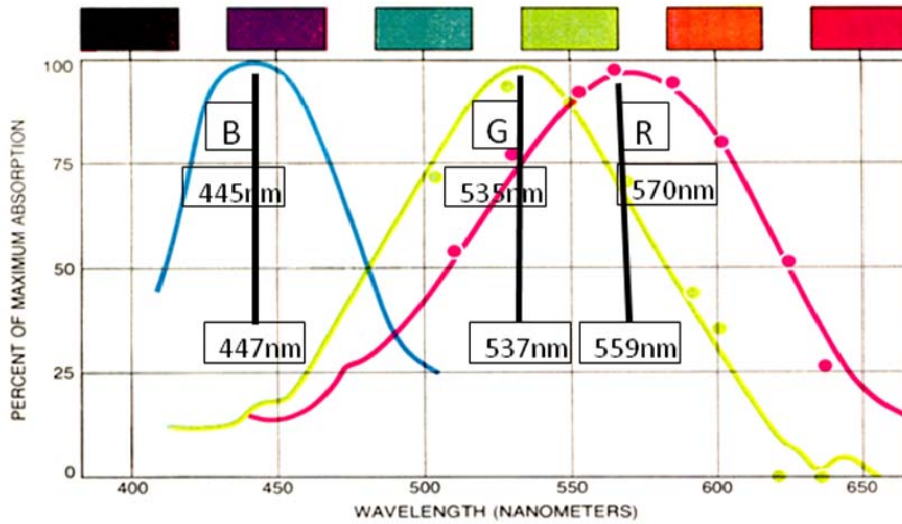


Fig. 15.15. The three RGB-resonance wavelengths correspond to the peak values of the spectral brightness sensitivity curves of the cones. The calculated RGB-values are indicated under the photometrically measured peak values.

The results of these calculations prove that the optical transformation of the visible spectrum into the RGB-color space can be realized via the space grating optics of the third cortical granular layer and not - as had been assumed up to date - by means of the chemistry of the photoreceptor pigments. Instead, the pigments are programmed by means of the space grating optics. In this way, the geometric and resonance requirements of the space grating interference optics in the Fresnel nearfield are met. Every resonant wavelength must correspond to the spatial Pythagoras and the crystal-optical $\nu\lambda$ resonance factor. Fig. 15.16 demonstrates how the RGB interference maxima are displayed in a top view. Optically they are linked by means of concurrence and provide the basis for overlapping receptive fields.

Fig. 15.17 illustrates how one ought to imagine the development of optical pillars by means of hierarchically ordered cell gratings with graduated grating constants. Such pillar-like information processing has been proven particularly in the visual cortical centers CGL and V1. The Fresnel diffraction images behind two hexagonal gratings with graduated grating constants demonstrate which field structures are to be expected in the Fresnel space.

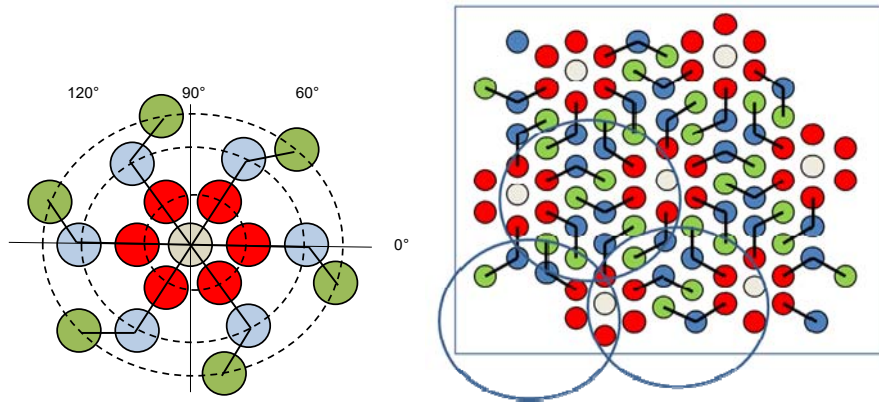


Fig. 15.16. (left) RGB-interferences in the top view; (right) distribution in the area. Optical basis for overlapping receptive fields.

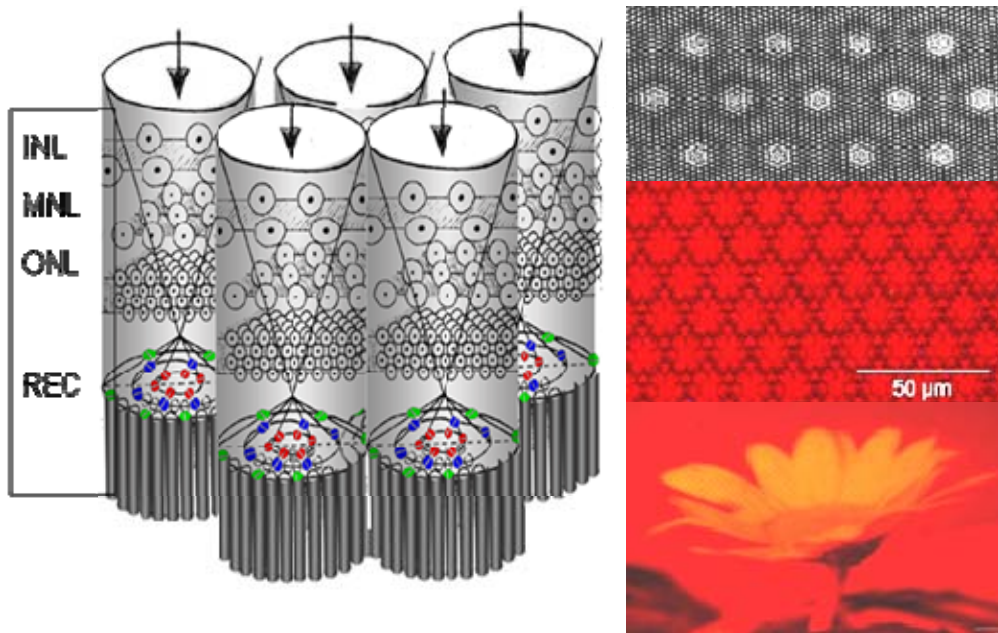


Fig. 15.17. (left) hierarchically ordered hexagonal gratings with graduated grating constants develop optical pillars. (right) Fresnel interferences behind two hexagonal phase gratings; (top) 100μm- with a 20μm grating, (center) 16μm- with a 8μm grating, (below) image of an object behind the 2nd grating.

The most important argument in favor of Fresnel nearfield interferences in the retina is based upon the fact that the local image information is retained while it is lost in the farfield and/or is present having been transformed there into Fourier-optical data constellations.

15.7. The Space Grating Optical Explanation of the Purkinje-Shift: RG(B) in Dimlight Vision

The Purkinje-Shift describes the transition from colored daylight to colorless dimlight vision in decreasing brightness. Whereas the photo pigment of the rod receptors reaches its peak in its brightness sensitivity curve at 512 nm and therefore in comparison to the RG-curves there is a transition to a shorter wavelength, a flower which is perceived as being blue during the day is then perceived to be brighter and a red rose to be darker. Using the data in Fig. 15.18, space grating optics explains the Purkinje-Shift as an adaptive shortening of the third grating constant g_z in the space grating. Equation (15.6) with $g_z = 0.913\mu\text{m}$ and $s = \sqrt{(2^2 + 0.913^2)} = \sqrt{4.834}$ leads to:

$$\lambda_{111}(R) = \frac{2 \times 1 \times 0.913 \times \sqrt{2^2 + 0.913^2}}{4 + 3 + 0.913^2} = 0.512 \mu\text{m}$$

$$\lambda_{123}(G) = \frac{2 \times 3 \times 0.913 \times \sqrt{2^2 + 0.913^2}}{4 + 12 + (3 \times 0.913)^2} = 0.512 \mu\text{m}$$

$$\lambda_{122}(B) = \frac{2 \times 2 \times 0.913 \times \sqrt{2^2 + 0.913^2}}{4 + 12 + (2 \times 0.913)^2} = 0.415 \mu\text{m}$$

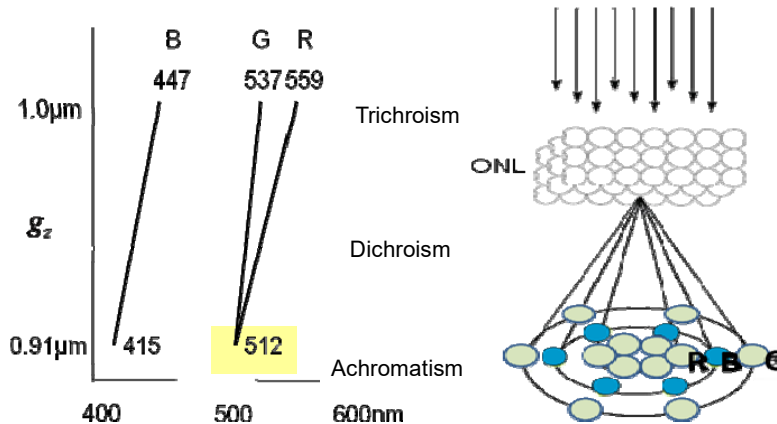


Fig. 15.18. Purkinje-Shift (left) with the merger of the G and R diffraction orders at 512 nm; (right) equal tuning of R and G.

This corresponds to a transformation of the round-shaped cells into a flattened rotational ellipsoid. In this way, the RG-diffraction orders fuse at 512 nm which is the peak of the spectral brightness sensitivity curve of the rods. At the same time, the contribution of B becomes nearly insignificant. The transition from Trichroism to Achromatism occurs via the instable blue-yellow Dichroism in both directions.

15.8. The Space Grating Chromatic Adaptivity in Daylight Vision

The adaptive tuning of the space grating not only occurs by means of a shortening of g_z in the direction of the shorter wavelengths as in the Purkinje-Shift, but rather by means of a prolongation of g_z - as proven in Fig. 15.19 - also in the direction of longer wavelengths with corresponding fundamental and harmonic wave RGB-triples in daylight vision.

This in turn corresponds to a transformation of the cells of the space grating into an elongated rotational ellipsoid. A calculation example for the RGB-triple at 612 nm (R), 560 nm (G), 482 nm (B) is shown in the figure on the right. In this way, the function of the third grating constant g_z has been completely explained. It plays a significant role in determining the RGB-chromatics from 512 nm over 559 nm and up to and beyond 612 nm.

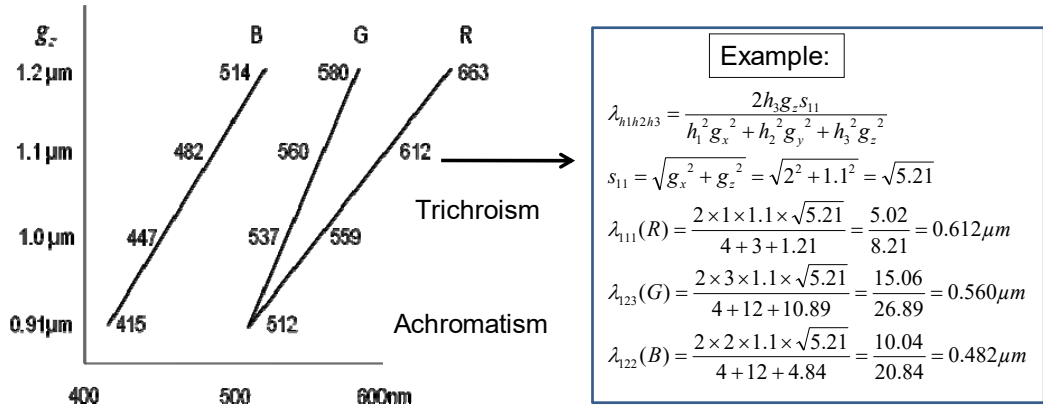


Fig. 15.19. (left) Space grating optical chromatic adaptation in daylight vision with g_z -prolongation to long-wave spectral ranges; (right) calculation example for a long-wave RGB-triple.

15.9. The Aperture Angle of the Light Cones of the Diffracted Light

While the third grating constant g_z determines the RGB-chromatics significantly, the two grating constants g_x, g_y are of decisive importance for the aperture angle of the light cones. Here, the angle γ being the angle related to the z-axis is of the highest importance. The calculation of the α -, β - and γ angles is carried out via the three segments of equation (15.7) in which the value s_{h1h2h3} occurs in their denominator and has to be calculated via equation (15.4).

$$\cos\alpha = \sqrt{\left(\frac{h_1g_x}{s_{h1h2h3}}\right)^2}; \cos\beta = \sqrt{\left(\frac{h_2g_y}{s_{h1h2h3}}\right)^2}; \cos\gamma = \sqrt{\left(\frac{h_3g_z}{s_{h1h2h3}} - 1\right)^2}, \quad (15.7)$$

Fig. 15.20 demonstrates the full range of the space grating optical available variation of the γ angles and the RGB-chromatics. In it, the γ angles of the light cones are indicated for the three RGB-chromatic locations (R559nm; R512nm; R612nm) which were previously discussed. The left half of the illustration, the smallest space grating cell with $g_x: g_y: g_z = 2: \sqrt{3}: 1$ (μm) provides the γ angle in the middle line; in the right half, the γ angle of the cell enlarged with $V = 6x$ in g_x, g_z to $g_x: g_y: g_z = 12: 6\sqrt{3}: 1$. The RGB-chromatics remains the same on each side of the line, the γ angles become considerably smaller due to the enlargement of $V = 6x$ of the g_x, g_y - grating constants. The same holds true for the top and bottom line which in each case result from a decrease of g_z to $g_z = 0.913\mu\text{m}$ in the Purkinje-Shift and/or an enlargement from g_z to $g_z = 1.1\mu\text{m}$ in the case of an adaptation to the longer R612nm wave.

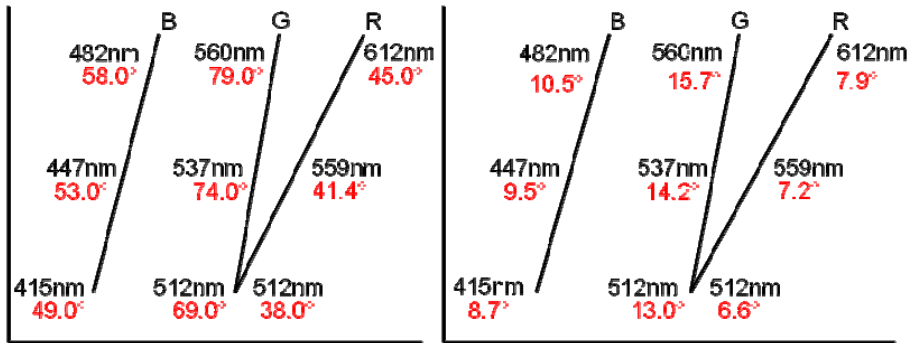


Fig. 15.20. γ angle of the diffraction order light cones in relation to the z-axis and RGB-chromatics for the three selected RGB-triples with R = 512, 559 and 612nm based on a change in size of the 3rd grating constant g_z (left) for the smallest hexagonal space grating cell in each case, (right) for the by $V = 6x$ in g_x, g_y enlarged space grating cell.

Amongst other things, the γ angles prove the RBG-sequence of light cones shown in Fig. 15.13 and the stronger widening of the γ angles as the g_z - grating constants increase. In the case of a percentage-wise equal enlargement or diminution of the three grating constants, i.e. in the case of a preservation of the linking of the three grating constants, the corresponding γ angles and RGB-chromatics of each line remain constant; the space grating then acts like an inflatable ball. This changes when the 3rd grating constant is released (shift from the middle to the top or bottom line) and g_x, g_y is kept constant. The space grating cell at $g_z < 1$ acts like a flattened and at $g_z > 1$ like an elongated rotational ellipsoid. The adaptive deformations of the cell geometry have a direct functional significance so that particular importance is placed on the decoupling of g_z and g_x, g_y .

15.10. The Retinal Space Grating Becomes a ‘Living Crystal’

By means of the deformation into a flattened or elongated rotational ellipsoid in association with the decoupling of the grating constants, the hexagonal space grating cell of the cortical retina becomes a ‘living crystal’ [Geguzin, 13] which has the ability to achieve adaptive accomplishments in this way. Similar adaptive cell geometry

deformations have been found in acoustics research in the human inner ear. The interpretation of the retina as a crystal-like brain layer is close to the statement made by the brain researcher Creutzfeld [14. p.118] He emphasized that the uniformity of the construction plan of the brain is not a case of construction elements or ‘chips’ separated from each other, but a continually repeating structure of interwoven single elements, “similar to a crystal grating in complex molecule systems”. The cell body layers have so far been completely neglected in brain research in comparison with the neuronal nets. This is all the more tragic because the cell bodies are the real producers of the action potentials upon which the neuronal nets rely.

15.11. The Individual and the Whole: The Relativization of Color of the Local on to the Color of the Global

E. Land [5] and J. J. McCann [15] agreed due to the retinex experiments that “Human color vision is a spatial calculation involving the whole image”. The necessary calculation work was attributed to the cortex due to a lack of alternatives. Now, however, it seems to become possible to attribute this relativization of the local to the global in RGBs to the optics in the eye. In the pupil - the aperture space of imaging optics - the eye possesses the information about the sum of the radiation of all self-radiators and reflectors which make their way into the eye from the object space. In 1948, E. Zernike [16] described it as follows: „The image that will be formed in a photographic camera is present in an invisible, mysterious way in the aperture of the lens, where the intensity is equal at all points”. There is one possibility how this Information which has never been measured in a targeted manner is used: that it is scattered by means of polygonal gratings in the cornea into the image space and is transformed there together with the local information of the object image into RGB so that the individual is related to the whole as a ‘coherent background’ in RGB. This is supported by: “The human eye exhibits considerable scattering“ [J. J. McCann,15]. Fig. 15.21 gives an impression of the complete grating optical correlator of the eye where irregular polygonal cell gratings in the aperture space undertake the scattering of the light. In this way, the diffraction of the light in the space grating of the cortical retina is supplemented by means of refractive optics.

15.12. Summary and Outlook

The results presented here prove that with Fresnel nearfield space grating optics a new interpretation of human vision in three relevant aspects succeeds; namely in the transformation of the visible spectrum in daylight vision into RGB-space, in the Purkinje-Shift to the RG(B)-dimlight vision and in the optical relativization of the colors from the local to the global. Therefore it can be claimed to have been explained that it is not the photo pigments of the receptors, but rather the diffractive Fresnel nearfield space grating optics which transforms the visible spectrum into RGB-color space. In this, the retinal hexagonal space grating resonator represents a transmission and receiver system with an RGB-sender, RGB space waves and RGB receiver antennae. The sender being the resonator produces interference optically standing RGB-space waves in the Fresnel nearfield, the nano antennas of the outer segments of the receptors represent the receivers.

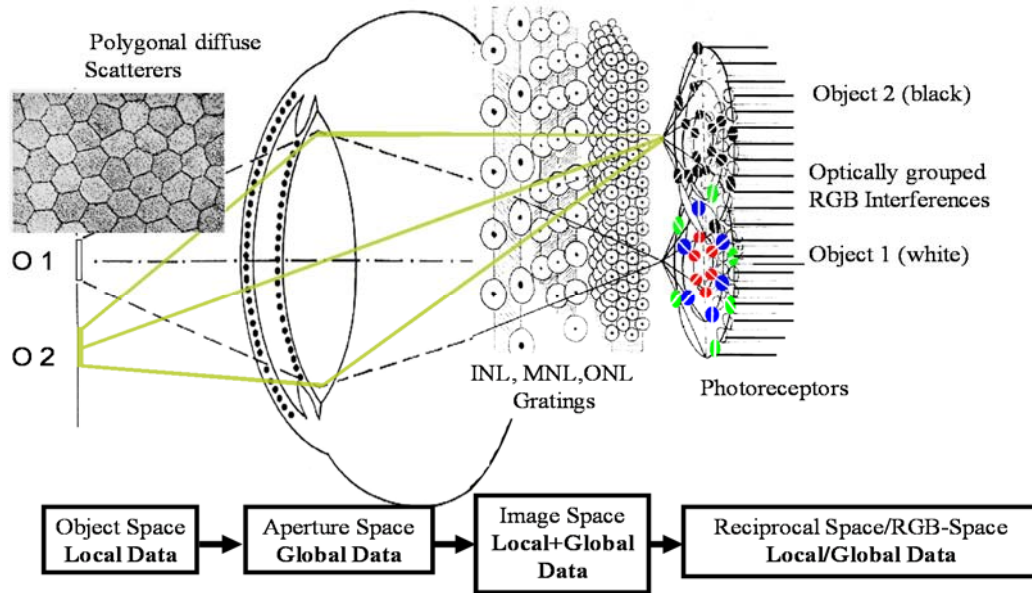


Fig. 15.21. The grating-optical correlator: By means of the scattering of the light at the gratings in the aperture space, local information could be relativized to global information in RGB.

Since the ‘inverted’ retina represents a brain layer, with its three granular layers - analogue to CGL and to V1 - as an optical multilayer image processor, it may be understood to be a building block of optical nearfield computing. By means of the diffraction of the visible into the RGB and thus into the ‘reciprocal grating space’ of the ‘living crystal’, the ‘inversion’ of the retina has decisive importance in evolution as a ‘jump function’ because it places visual experience in the eye on the (adaptive) RGB-basis of crystal optics. The sentence by Bergmann/Schäfer [2] is still valid for the practical implementation of the results: “Diffraction at a space grating is really of no importance in actual optics since there have been hardly any successes in creating a space grating of sufficient preciseness for the visible light”. The mathematical calculation merely shows a way for the technical reconstruction of the retinal space grating structures. In the NAMIROS-BMBF-Project [17] in which the author was involved as the initiator and leader from 2006-13 with competent partners (Fraunhofer-IAP/Golm, mrt Berlin, TU Optisches Institut Berlin, Physik at the University of Siegen, Vitronic/Wiesbaden), progress in the development of polymers was successfully made which made an optical inscription of phase gratings possible. Unfortunately, the optical inscription procedures were only able to produce high-performance space gratings for farfield interference optics, but not for Fresnel nearfield interferences which rely on short lengths of the 3rd grating constants. Accordingly, space gratings which can be used experimentally remain to be an open target which may be achievable with 3D print and Nano imprint procedures in the near future. Space grating optics will continue to be practically relevant especially for retina implants. The experimental reappraisal of the diffractive Fresnel nearfield space grating optics is a necessity and a chance for the further development of optics. Due to the shift of the visual

information processing into the ‘reciprocal grating space’ of crystal physics, it has a key function for a better understanding of the cortical layers. Only due to the fact that in the image space of the eye, the optical RGB-transformation of the visible occurs in the ONL space grating which is located anterior to the receptors, can the colors of things become perceivable on the objects. In this, the local RGB-shares are at all times simultaneously available and therefore do not require a neuronal linking.

References

- [1]. P. P. Ewald, Crystal optics for visible light and X-rays, *Reviews of Modern Physics*, Vol.37, Issue 1, 1965, pp. 46-56.
- [2]. L. Bergmann, C. Schaefer, Lehrbuch der Experimental-Physik, Bd.III,1. Wellenoptik, *Walter de Gruyter*, 3/1962.
- [3]. W.A.H. Rushton, Visual pigments and color blindness, *Scientific American*, Vol. 232, Issue 3, 1975.
- [4]. A. Schopenhauer, Ueber das Sehn und die Farben: Eine Abhandlung, *Brockhaus*, Leipzig, 1870.
- [5]. E. Land, The retinex theory of color vision, *Scientific American*, Vol. 237, Issue 6, 1977, pp. 108-128.
- [6]. N. Lauinger, The Human Eye: An Intelligent Optical Sensor (The inverted human retina: a diffractive-optical correlator), *IFSA Publishing*, 2014.
- [7]. I. Mann, The Development of the Human Eye, *British Medical Assoc.*, London, 1949.
- [8]. W. Bargmann, Histologie und mikroskopische Anatomie des Menschen, *Thieme-Verlag*, Stuttgart, 1967.
- [9]. E. Blechschmidt, Die Entwicklungsbewegungen der menschlichen Retina zur Zeit der Irisentstehung, *Ophthalmologica*, Vol. 154, 1967, pp.531-550.
- [10]. S. Ramon y Cajal, Cajal on the Cerebral Cortex, *Oxford University Press*, 1988.
- [11]. R. W. Rodieck, The First Steps in Seeing, *Sinauer Assoc.*, 1998.
- [12]. M. Tomandl, Realisierung von optischen Talbot- und Talbot-Lau-Teppichen, *Diploma Univ. Wien/Physik*, 2010.
- [13]. Ja. E. Geguzin, Lebender Kristall, *Verlag Harri Deutsch*, Thun/Frankfurt, 1985.
- [14]. O. Creutzfeld, Gehirn und Geist, *Verlag Göttinger Tageblatt, Göttingen*, 1986.
- [15]. J. J. McCann, Color imaging systems and color theory: past, present and future, *Proceedings of SPIE*, Vol. 3299, 1998, pp.38-46.
- [16]. F. Zernike, Diffraction and optical image formation, *Proceedings of the Phys. Soc.*, Vol.61, 1948, pp.158-164.
- [17]. NAMIROS-BMBF-Projekt 13N9040, *Nano- und Mikro-Raumgitter für die Optische Sensorik*, pp. 2006-2012.

Chapter 16

Characteristics of Unbalanced Mach-Zehnder Interferometers in Metal/Insulator/Metal Plasmonic Waveguides

Shun Kamada, Toshihiro Okamoto and Masanobu Haraguchi

16.1. Introduction

The integration of nanoscale optoelectronic devices is rapidly approaching the limitations below the diffraction limit of light. The diffraction limit of light does not allow the confinement of light in nanoscale regions much smaller than the wavelength of light. Plasmonic waveguides (PWGs) are able to concentrate and manipulate light below the diffraction limit. Plasmonic devices have attracted a great deal of attention for developing next-generation integrated optical systems. Plasmonic devices provide the facility for ultra-compact optical integrated circuits with the bandwidth of photonic networks [1-3].

Optical modulators are one of the most effective and important parts in optical signal processing of optical integrated circuits. Many plasmonic modulators based on PWGs have been reported [4-9]. In particular, Mach-Zehnder interferometers (MZIs) are typical structures used in electro-optic modulators. MZIs based on gap PWGs [10-13], ridge PWGs [14], and V-groove PWGs [15] have been investigated. Gap PWGs allow the plasmonic mode to be confined and propagates in the middle of the gap. In recently reported MZIs based on gap PWGs, the surface plasmon polariton (SPP) propagates not in the center but on the edges of the gap [10]. In such cases, the SPP will be subjected to some scattering loss due to the waveguide curvature (i.e., bending losses) or some defects and surface roughness of the fabricated structure [16]. In multilayer metal/insulator/metal (MIM) PWGs, there are no such the problems [1, 17, 18]. MIM PWGs for propagating SPP have recently been proposed [17-23]. In this study, we propose MIM PWGs based on the deposition of a metal thin film and dielectric layer. The proposed structure possesses the strong stability to confine light [24] and has low loss in bending waveguides

[25]. Our proposed structure combines a simple design with a rapid fabrication process. A number of researchers have numerically investigated the optical properties of MZI using MIM PWGs [26-28]. To the best of our knowledge, no experimental investigations on MZIs have been reported using multilayer MIM PWGs.

Furthermore, the propagation length of the plasmon mode in MIM PWGs, defined as the distance for the SPP intensity to decay by a factor $1/e$, is short ($\sim 10 \mu\text{m}$ at near-infrared region) [1] compared with those of guided modes in the dielectric waveguide. In order to realize a large change in light intensity using an MZI modulator, it is necessary to have interference in the optical path length that can apply a π phase shift. However, the mode length must be short to be compatible with the short propagation length of the SPP. Hence, we cannot obtain enough change in light intensity for a typical MZI structure if the electro-optic material has low electro-optic coefficients $r_{33} \sim 32 \text{ pm/V}$ [29]. Therefore, in our study we focus on a normally OFF-type unbalanced MZI, which provides a theoretically high ON/OFF ratio. We experimentally confirmed the interference of SPPs propagating in each propagation path of the proposed structure. The proposed structure can be used in a modulator by replacing the electro-optic materials in the insulator. In this chapter, we report a passive unbalanced MZI as the first step in the development of a compact electro-optic modulator.

16.2. Design and Analysis of the Unbalanced MZI

First, the configuration and performance of the MZI structure were analyzed by numerical simulations. To evaluate the optical properties of the unbalanced MZI, a 2D finite difference time domain (FDTD) method was employed to stimulate the transmission characteristics [30].

When the length in depth of MIM PWGs is a finite ($\sim \mu\text{m}$), the SPP mode propagating in PWGs will become multimodes [24]. It is difficult and complex to analyze the propagating multimode SPP. This multimode also leads to a biased electric field distribution in PWGs [31]. For this reason, propagated light easily scatters unless we carefully design the distance from the input port to the branch and bend of the MZI. Furthermore, because the biased electric field distribution depends on incident wavelength, this MZI has a wavelength dispersion property. We consider slub-type PWGs to clarify the fundamental properties of the MZI based on an MIM PWG. The mode of SPPs propagating in slub-type PWGs become single mode as shown by the following dispersion relation [1]:

$$\tanh\left(\frac{t\sqrt{k_{SPP}^2 - \varepsilon_i k_0^2}}{2}\right) = -\frac{\varepsilon_i \sqrt{k_{SPP}^2 - \varepsilon_m k_0^2}}{\varepsilon_m \sqrt{k_{SPP}^2 - \varepsilon_i k_0^2}}, \quad (16.1)$$

where k_0 and t are the vacuum wavenumber of the incident light and insulator thickness, respectively; and ε_m and ε_i are the dielectric constants of the metal and insulator, respectively.

The spatial cell size in the 2D FDTD method has been set to be 2 nm. Fig. 16.1 shows a typical schematic diagram of the MZI based on MIM PWGs used in FDTD simulations. Silver was used as the upper and lower metal layers of the MIM structure, where the dielectric constant of silver is employed from previous work [32]. The dielectric layer of the MIM structure is composed of a sodium p-styrenesulfonate homo-polymer (PSSNa) with refractive index of $n = 1.395$. Because this refractive index is similar to the refractive index of other polymers, the simulation does not lose generality. The PSSNa thickness are $d_0 = d_1 = d_2 = 100$ nm. The input port was illuminated by light with E_z component to excite the plasmon. The input light is confined in the dielectric layer and will propagate along the waveguides. SPP waves will be obtained and guided along the branch waveguide, and then divided into paths P_1 and P_2 . Transmittances were calculated from the ratio $(E_{\text{Out}}/E_{\text{In}})^2$ of the electric field intensity in the input port E_{In} and output port E_{Out} . The difference in path length ΔL between P_1 and P_2 has three different Ag thickness d_3 values of 400 nm, 600 nm, and 800 nm, as shown in Fig. 16.1. To confirm constructive and destructive interference at wavelengths 1100–1600 nm, different path lengths ΔL were chosen at 400 nm, 600 nm, and 800 nm.

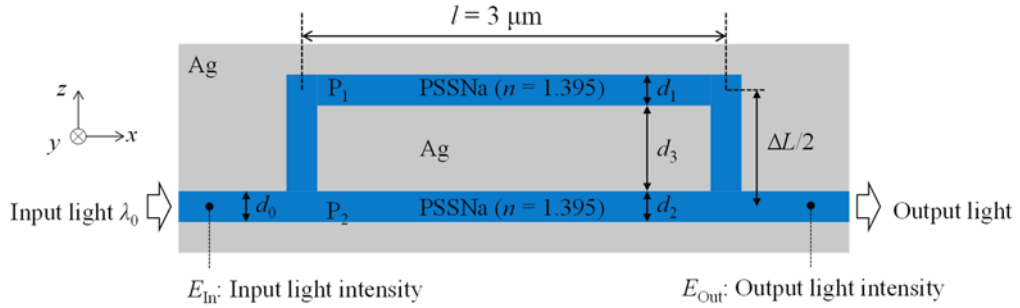


Fig. 16.1. A typical schematic diagram of the MZI based on MIM PWGs used in the 2D FDTD simulations. ($d_0 = d_1 = d_2 = 100$ nm, $d_3 = 100$ nm, 200 nm, 300 nm).

The results of the FDTD simulations were used to evaluate propagation loss in bends and splitters of PWGs. The transmittance in 90° bends of PWGs was 80 % compared with straight PWGs. The transmittance ratio between straight and 90° bends and splitters of PWGs were 55 % and 45 %, respectively. Fig. 16.2(a) shows the results of our calculations of the transmittance spectra. Transmittances varied with the wavelength of input light λ_0 . The transmittance at $\lambda_0 = 1420$ nm was found to be a maximum for $\Delta L = 800$ nm. The maximum transmittance for $\Delta L = 800$ nm, as shown Fig. 16.2, is 79 %. This transmittance is similar to the theoretical transmittance in straight PWGs with the same path length. The transmittance at $\lambda_0 = 1300$ nm was found to be a minimum for $\Delta L = 400$ nm. Fig. 16.2 (b) shows the transmission spectra of path length differences of 400 nm, 600 nm, and 800 nm deduced by the analytical solution. The spectral shapes determined by the numerical solution were quite similar to those determined by the analytical solution. These indicate that SPPs propagating in each path of MZI at $\lambda_0 = 1300$ nm for $\Delta L = 400$ nm undergo destructive interference. For SPPs with wavelength λ_{SPP} propagating in MIM PWGs, the condition of destructive interference can be expressed as the following equation:

$$\Delta L \approx N\lambda_{\text{SPP}} + \lambda_{\text{SPP}}/2, N = 1, 2, 3, \dots \quad (16.2)$$

The wavelength λ_{SPP} of SPPs propagating in the waveguide, as shown Fig. 16.1, is 750 nm at $\lambda_0 = 1300$ nm, which was calculated from the MIM SPPs dispersion relation expressed as Eq. 1. Therefore, according to Eq. 2, we may conclude that for destructive interference, Eq. 2 approximately tends to ($\lambda_{\text{SPP}}/2 = 375$ nm). Thus, by increasing the path length difference ΔL , the wavelength of input light λ_0 for maximum (minimum) transmittance will shift to longer wavelengths. Therefore, the phase difference between P_1 and P_2 leads to a change in transmittance, which is clearly shown in Fig. 16.2.

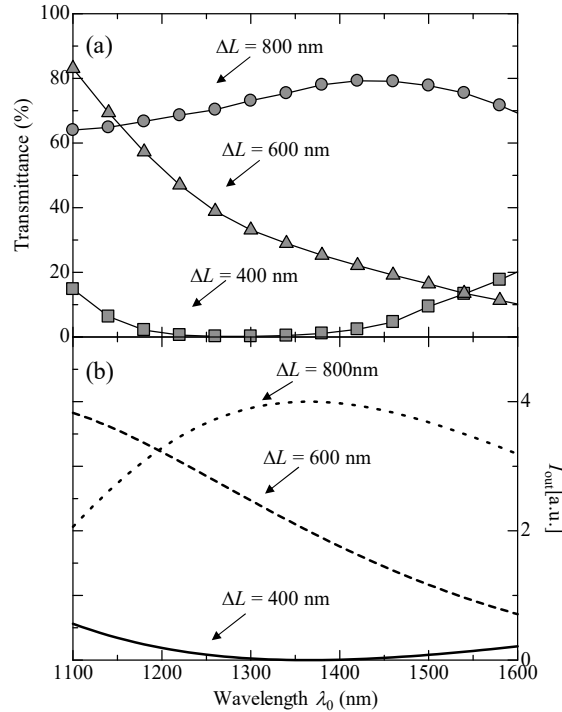


Fig. 16.2. (a) Transmittance spectra of proposed unbalanced MZI by numerical analysis. (b) Transmittance spectra of path length difference of 400 nm, 600 nm, and 800 nm deduced by the analytical solution.

16.3. Fabrication and Experimental Results in Straight MIM PWGs

In this section, we evaluate the propagation properties of MIM PWGs. The fabrication process is shown in Fig. 16.3, where a PSSNa diluted by pure water is adapted for an insulator material. The thickness of PSSNa film on a substrate was controlled by the rotational speed of the spin coating method. At first, we determined the optimum rotational speed for spin coating of the PSSNa film. The PSSNa concentrations were 100 wt %, 50 wt %, and 25 wt %. The PSSNa was dropped onto the Si substrate and then rotated at high speed in order to spread the fluid. It was annealed at 120 °C to evaporate

the solvent. Then, PSSNa thickness was measured by scanning probe microscopy. The relationship between rotational speed and PSSNa thin film thickness is shown in Fig. 16.4. The thickness values were obtained from the average of three thickness values measured on the same substrate. The error bars shown in Fig. 16.4 express standard deviations. It is found that PSSNa film thickness decreases with increasing rotational speed and decreasing concentration. The PSSNa concentration of 50 wt % and rotational speed of 6000 rpm are the optimum values for setting film thickness to 100 nm.

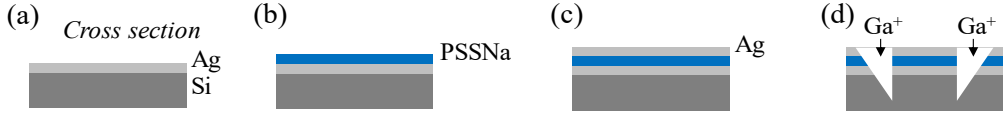


Fig. 16.3. Fabrication process of single-layer PWG. (a) Ag film (100 nm) of the bottom layer was deposited on a Si substrate by thermal evaporation; (b) PSSNa films were used to coat the Ag film by the spin coating method; (c) Ag film (100 nm) of the upper layer was deposited; (d) Input/output ports were fabricated using focus ion beam milling.

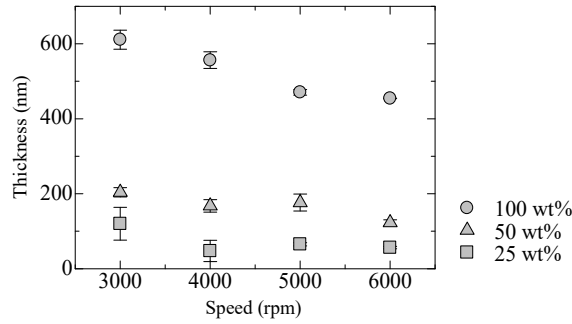


Fig. 16.4. Relationship between spin speed and PSSNa thin film thickness. PSSNa was coated on Si wafers using different speeds and baked at 120 °C for 60 s. Film thickness was measured by scanning probe microscopy.

V-like channels were fabricated by a focus ion beam (FIB) to serve as the input and output ports. If the ion beam is scanned in one direction, the sputtered atoms were deposited on the wall and became V-like channels. The V-like channels were designed to have an angle of 45°.

Fig. 16.5 shows a schematic diagram of the experimental setup in order to evaluate input/output light properties of the fabricated PWG. The tunable laser source (wavelength λ_0) is an optical parametric oscillator (OPO) that pumped a Ti: Sapphire laser (wavelength = 835 nm), which was utilized as the incident light source. A $\lambda/4$ plate was used to turn plane-polarized light into circularly-polarized light. The laser polarization was controlled by a polarizer. A laser light was illuminated at the input port of the unbalanced MZI through an objective lens. The intensities of the scattered light at the output port of the unbalanced MZI were recorded by an infrared charge-coupled device (IR-CCD) camera through an objective lens and a half mirror.

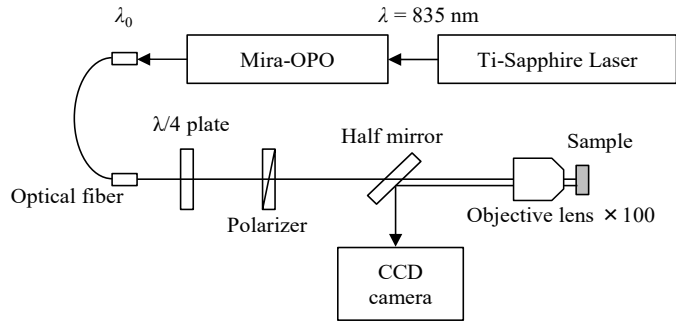


Fig. 16.5. A schematic diagram of the experimental setup for input/output light observation.

We confirmed the propagation of SPPs in the fabricated MIM PWG. A CCD camera image of an MIM PWG, with a distance $25\ \mu\text{m}$ between input/output ports, is shown in Fig. 16.6 (a). The incident light at wavelength $1.3\ \mu\text{m}$ was irradiated toward the input port. The incident light propagating in the z -direction refracts at the input port and propagated in the y -direction. The yz plane was assumed as the incident plane in this experiment. Transverse magnetic (TM) polarized light is characterized by its magnetic field, which was perpendicular to the incident plane. For TM light, the electric field was parallel to the incident plane. For transverse electric (TE) light, the electric field was perpendicular to the incident plane. CCD camera images of incident TM and TE polarized light are shown in Figs. 16.6 (b) and (c), respectively. Scattered lights are observed at the output port (Fig. 16.6 (b)), whereas there were no scattered lights at the output port (Fig. 16.6 (c)). Incident lights that illuminate the input port were reflected to an edge of the MIM PWG. In this case, when the incident light has TM or TE polarization, the incident light that illuminates the edge of the MIM PWG has a polarization direction that is vertical or parallel to the direction of a substrate, respectively. In general, the polarization direction of SPPs that propagate in layered MIM PWGs will be in the vertical direction of the metal-insulator interface [1]. When the incident light in Fig. 16.6(b) leads to an observed scattered light at the output port, with a polarization direction in the vertical direction of the MIM PWGs, the propagation condition of SPPs is satisfied. Therefore, the propagation of SPPs in the fabricated straight PWG has been verified.

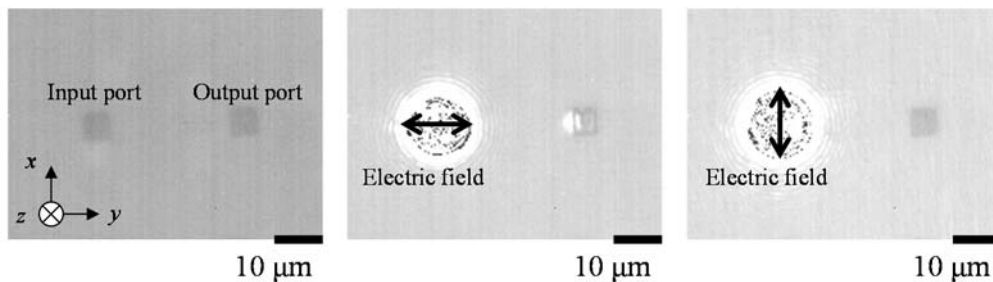


Fig. 16.6. CCD images of the MIM PWG. (a) The irradiation of halogen lamp; (b) Illumination of TM polarized light at the input port; (c) Illumination of TE polarized light at the input port.

Next, the propagation length of the fabricated MIM PWG was estimated. Light scattering intensities at the output port were evaluated by changing the distance between the input and output ports from 10 μm to 22 μm . Light scattering intensities at the output port were measured according to the brightness of images taken by the CCD camera. Light scattering intensities at the output port with changing distance between the input/output ports are shown in Fig. 16.7. The wavelength of incident light was tuned at a wavelength of 1.5 μm . As shown in Fig. 16.7, output light intensity weakens with increasing distance between input/output ports. We assumed that output light intensity decreased exponentially. The output light intensity curve was fitted by the least squares method. The results show an SPP propagation length of 13.5 μm . The theoretical value of the propagation length estimated by dispersion relation in Eq. 1 is 16.0 μm . It is presumed that the experimental propagation distance is less than theoretical value because of the roughness of the Ag surface. It is inferred that the propagation distance became shorter because the propagating SPP was scattered by the roughness of the Ag surface.

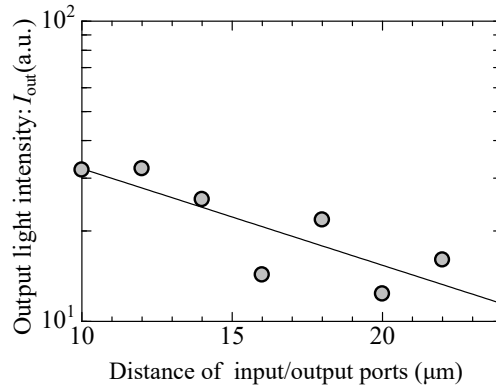


Fig. 16.7. Relationship between distance of input/output ports and light scattering intensity at output port. The incident wavelength is 1.5 μm tuned by the OPO.

The coupling efficiency of the SPP and the space propagation of light at the input/output ports were estimated. Light scattering intensity at the output port I_s expressed as

$$I_s = \eta^2 I_0 e^{-y/L} \quad (16.3)$$

where η is the coupling efficiency at the input/output ports, I_0 and L are the incident light intensity and propagation length, respectively, and y is the distance in the PWGs. I_s and I_0 were measured by CCD images. The coupling efficiency η was estimated using Eq. (16.3). As a result, the coupling efficiency η is approximately 0.118. Hence, approximately 11.8 % of incident light is coupled to the PWG at the fabricated input port.

16.4. Fabrication and Experimental Results in MZI Based on MIM PWGs

Fig. 16.8 shows a schematic diagram of the whole fabrication process of the MZI structure. The process begins with the thin film deposition of Ni (approximately 10 nm) on a Si

substrate using a sputtering method to improve adhesion, as shown Fig. 16.8 (a). Then, Ag thin films (100 nm) were deposited by a thermal evaporation method. The PSSNa film of the first layer was used as a dielectric layer, which was deposited using a spin coating method. The Ag thin film of the middle layer (300 nm) was deposited by a thermal evaporation method, as shown in Fig. 16.8 (b). The PSSNa film of the second layer was deposited as shown in Fig. 16.8 (c). Finally, a second Ag thin film (approximately 72 nm) was also deposited by thermal evaporation method at an incident angle of 45° and -45° , as shown in Fig. 16.8 (d). V-like channels were fabricated by FIB milling to serve as input and output ports, as shown in Fig. 16.8 (e). To fabricate a modulator in this structure, the PSSNa will be replaced to electro-optic polymer materials. The structure using electro-optic materials can be fabricated using the same fabrication process shown in Fig. 16.8.

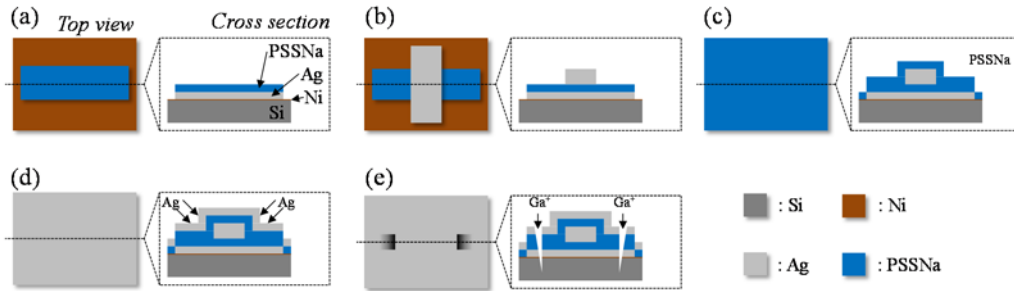


Fig. 16.8. Fabrication process of the unbalanced MZI based on MIM PWGs. (a) The patterns of the Ag film (100 nm) and PSSNa film were formed by electron beam (EB) lithography; (b) The pattern of the Ag film (300 nm) was formed by overlay exposure of EB lithography; (c) The PSSNa film was coated by a spin coating method; (d) The Ag film (72 nm) was deposited by the thermal evaporation of silver at incident angles 45° and -45° , respectively; (e) Input/output ports were fabricated using focus ion beam milling.

Fig. 16.9 shows scanning electron microscope (SEM) images of the fabricated unbalanced MZI structure based on MIM PWGs. Fig. 16.9 (a) shows the top-view of the fabricated structure. Fig. 16.9 (b) shows the cross-section image of the fabricated structure along the region from A to B in Fig. 16.9 (a). Fig. 16.9 (b) shows that the thickness of the PSSNa film before a branch (d_0) is approximately 350 nm. The thickness of PSSNa films after a branch (d_1 , d_2) are the same at approximately 130 nm. The SPPs can be confined and propagate within the thicknesses PSSNa with Ag cladding. The propagation length of the fabricated MIM PWGs constructed by Ag/PSSNa/Ag films was evaluated. The propagation length at $\lambda_0 = 1500$ nm was $13.5 \mu\text{m}$.

Ag-1 and Ag-2 films shown in Fig. 16.9 (b) are in a conductive state. If we apply an electric field between Ag-1 (Ag-2) and Ag-3, the structure can be used as a modulator owing to the phase shift.

The experimental setup shown in Fig. 16.5 was employed to measure input/output light properties in the fabricated MZI. The wavelength of the incident light was varied from 1150 nm to 1550 nm (50-nm step) by the OPO. We confirmed that only the propagating

TM polarized light is in the structure. Because SPPs were only excited by TM polarized light, we believe that SPPs propagate in our structure.

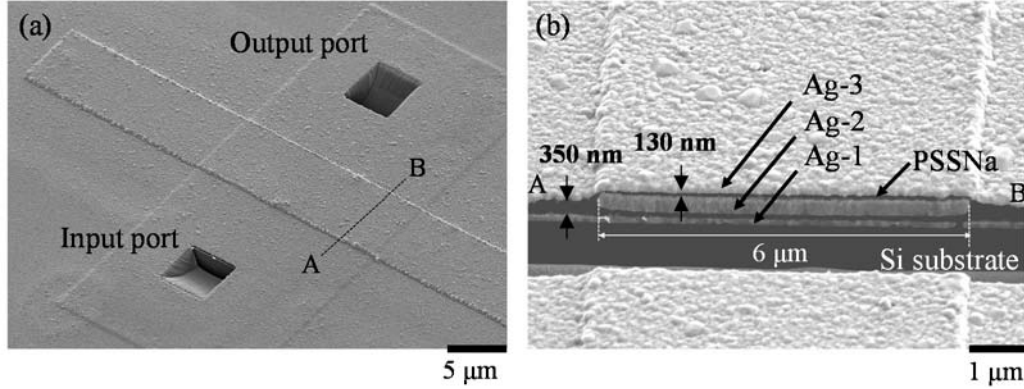


Fig. 16.9. SEM images of the fabricated unbalanced MZI. (a) SEM bird's-eye view image; (b) A cross section of the fabricated structure produced through focus ion beam milling from A to B.

Fig. 16.10 shows the experimental spectrum of light intensities scattered at the output port as a function of the wavelength of incident light λ_0 . The experimental results were obtained from the average values of 15 data points from three samples measured five times at the same substrate and fabrication conditions. The error bars in Fig. 16.10 express standard deviations. Light intensities at the output port were normalized by the intensities of incident light. Normalized light intensity I_0 was found to change as a result of the changing wavelength of incident light, as shown in Fig. 16.10. I_0 is maximum at $\lambda_0 = 1250$ nm and gradually decreased at longer wavelengths, reaching its minimum at $\lambda_0 = 1500$ nm. In the fabricated structure, it was assumed that constructive and destructive interference were achieved at $\lambda_0 = 1250$ nm and 1500 nm, respectively.

As previously shown in Fig. 16.2, the experimental results of the spectrum of light intensities at the output port were not in agreement with numerical simulation results. This may be related to the film thickness of PSSNa for a branch of incoming and outgoing waveguides. The film thickness variation resulted when the PSSNa was spin coated twice. To investigate the effect of thickness d_0 , the transmittance spectra of the unbalanced MZI was calculated using 2D FDTD simulation by changing the parameter from an SEM image, as shown in Fig. 16.9. The PSSNa thickness before a branch d_0 was found to have different thickness values of 130 nm, 240 nm, and 350 nm. Other sizes of the unbalanced MZIs are $d_1 = d_2 = 130$ nm, $l = 6$ μ m, and $d_3 = 300$ nm based on the SEM image shown in Fig. 16.9 (b).

Transmittance spectra for different thicknesses are shown in Fig. 16.11. Transmittance spectra varied with changing thickness d_0 . The transmittance at a wavelength of input light $\lambda_0 = 1350$ nm was found to be a maximum for thickness $d_0 = 130$ nm. By increasing the thickness d_0 , the transmittance spectra shifted to the shorter wavelength.

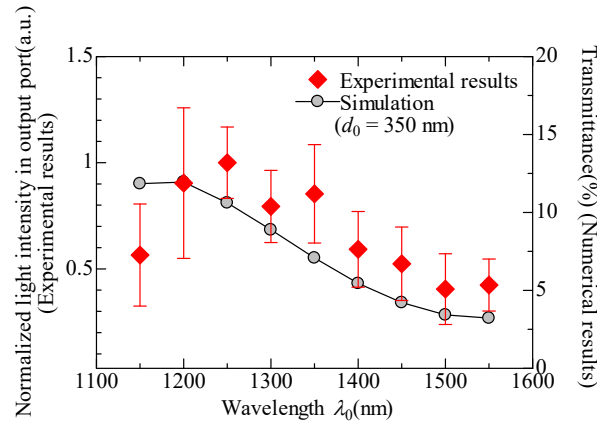


Fig. 16.10. Closed diamonds show the experimental results of normalized light intensity in output port of the unbalanced MZI. Open circles show the numerical simulation results of transmittance in the MZI.

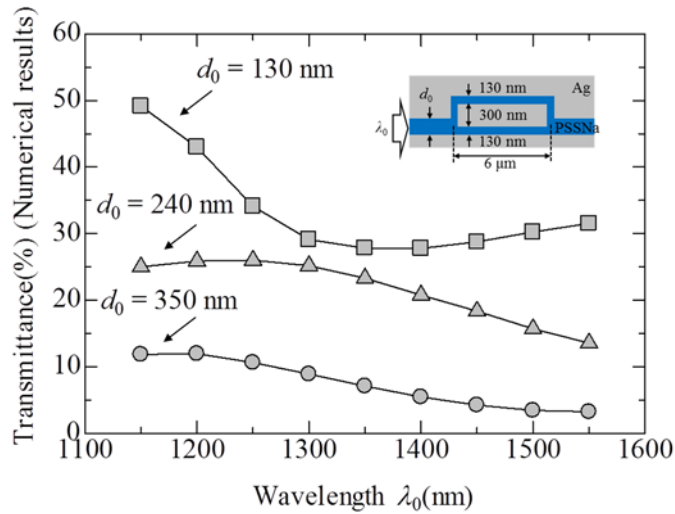


Fig. 16.11. Numerical analysis of the transmittance spectra as a function of the thickness d_0 of PSSNa.

If the thickness of the PSSNa layer d_0 increases, the SPPs were quicker to reach the upper branch waveguide. Therefore, transmittance spectra will be subject to shifts toward shorter wavelengths owing to the change in interference condition.

The thickness of the PSSNa layer before a branch in the fabricated structure was close to $d_0 = 350$ nm. The transmittance spectrum from 2D FDTD simulation at such thickness is shown in Fig. 16.10. The experimental and numerical simulation results were found to be in good agreement and the fabricated structure behaves as an MZI.

16.5. Conclusion

In this study, we proposed and designed unbalanced MZIs based on MIM PWGs. The characteristics of the proposed structure were analyzed and discussed. We fabricated the proposed structure by electron beam lithography techniques. The optical measurement results show that the normalized light intensity at the output port varied with incident light wavelength, and was a maximum and minimum at $\lambda_0 = 1200$ nm and 1500 nm, respectively. The experimental and numerical spectra were in good agreement qualitatively. The proposed structure is expected to realize an ultra-compact modulator by replacing part of the insulator with an electric-optic polymer.

References

- [1]. S. I. Bozhevolnyi, Plasmonic Nanoguides and Circuits, *Pan Stanford Publishing Pte. Ltd.*, 2009.
- [2]. M. L. Brongersma, V. M. Shalaev, The case for plasmonics, *Science*, Vol. 328, 2010, pp. 440-441.
- [3]. N. Kinsey, M. Ferrera, V. M. Shalaev, A. Boltasseva, Examining nanophotonics for integrated hybrid systems: a review of plasmonic interconnects and modulators using traditional and alternative materials [Invited], *J. Opt. Soc. Am. B*, Vol. 32, 2015, pp. 121-142.
- [4]. J. A. Dionne, K. Diest, L. A. Sweatlock, H. A. Atwater, PlasMOSor: A Metal-Oxide-Si field effect plasmonic modulator, *Nano Lett.*, Vol. 9, 2009, pp. 897-902.
- [5]. R. P. Dwivedi, H.-S. Lee, J.-H. Song, S. An, E.-H. Lee, Plasmonic modulator utilizing three parallel metal-dielectric-metal waveguide directional coupler and elasto-optic effects, *Opt. Commun.*, Vol. 284, 2011, pp. 1418-1423.
- [6]. V. E. Babicheva, A. V. Lavrinenko, Plasmonic modulator optimized by patterning of active layer and tuning permittivity, *Opt. Commun.*, Vol. 285, 2012, pp. 5500-5507.
- [7]. S. Zhu, G. Q. Lo, D. L. Kwong, Phase modulation in horizontal metal-insulator-silicon-insulator-metal plasmonic waveguides, *Opt. Express*, Vol. 21, 2013, pp. 8320-8330.
- [8]. V. E. Babicheva, S. V. Zhukovsky, A. V. Lavrinenko, Bismuth ferrite as low-loss switchable material for plasmonic waveguide modulator, *Opt. Express*, Vol. 22, 2014, pp. 28890-28897.
- [9]. A. Melikyan, L. Alloatti, A. Muslija, D. Hillerkuss, P. C. Schindler, J. Li, R. Palmer, D. Korn, S. Muehlbrandt, D. Van Thourhout, B. Chen, R. Dinu, M. Sommer, C. Koos, M. Kohl, W. Freude, J. Leuthold, High-speed plasmonic phase modulators, *Nature Photonics*, Vol. 8, 2014, pp. 229-233.
- [10]. A. Melikyan, K. Koehnle, M. Lauermann, R. Palmer, S. Koeber, S. Muehlbrandt, P. C. Schindler, D. L. Elder, S. Wolf, W. Heni, C. Haffner, Y. Fedoryshyn, D. Hillerkuss, M. Sommer, L. R. Dalton, D. Van Thourhout, W. Freude, M. Kohl, J. Leuthold, C. Koos, Plasmonic-organic hybrid (POH) modulators for OOK and BPSK signaling at 40 Gbit/s, *Opt. Express*, Vol. 23, 2015, pp. 9938-9946.
- [11]. C. Haffner, W. Heni, Y. Fedoryshyn, J. Niegemann, A. Melikyan, D. L. Elder, B. Baeuerle, Y. Salamin, A. Josten, U. Koch, C. Hoessbacher, F. Ducry, L. Juchli, A. Emboras, D. Hillerkuss, M. Kohl, L. R. Dalton, C. Hafner, J. Leuthold, All-plasmonic Mach-Zehnder modulator enabling optical high-speed communication at the microscale, *Nature Photonics*, Vol. 9, 2015, pp. 525-528.
- [12]. C. Haffner, W. Heni, Y. Fedoryshyn, A. Josten, B. Baeuerle, C. Hoessbacher, Y. Salamin, U. Koch, N. Dordevic, P. Mousel, R. Bonjour, A. Emboras, D. Hillerkuss, P. Leuchtmann, D. L. Elder, L. R. Dalton, C. Hafner, J. Leuthold, Plasmonic organic hybrid modulators –

- scaling highest speed photonics to the microscale, *Proceedings of the IEEE*, Vol. 104, 2016, pp. 2362-2379.
- [13]. W. Heni, C. Haffner, B. Baeuerle, Y. Fedoryshyn, A. Josten, D. Hillerkuss, J. Niegemann, A. Melikyan, M. Kohl, D. L. Elder, L. R. Dalton, C. Hafner, J. Leuthold, 108 Gbit/s plasmonic Mach-Zehnder modulator with > 70 GHz electrical bandwidth, *Journal of Lightwave Technology*, Vol. 34, 2016, pp. 393-400.
 - [14]. J. Gosciniaik, S. I. Bozhevolnyi, Performance of thermo-optic components based on dielectric-loaded surface plasmon polariton waveguides, *Scientific Reports*, Vol. 3, 2013, 1803.
 - [15]. S. I. Bozhevolnyi, V. S. Volkov, E. Devaux, J. -Y. Lauluet, T. W. Ebbesen, Channel plasmon subwavelength waveguide components including interferometers and ring resonators, *Nature*, Vol. 440, 2006, pp. 508-511.
 - [16]. D. R. Mason, D. K. Gramotnev, K. S. Kim, Wavelength-dependent transmission through sharp 90 bends in subwavelength metallic slot waveguides, *Opt. Express*, Vol. 18, 2010, pp. 16139-16145.
 - [17]. G. Veronis, Z. Yu, S. E. Kocaba, D. A. B. Miller, M. L. Brongersma, S. Fan, Metal-dielectric-metal plasmonic waveguide devices for manipulating light at the nanoscale, *Chinese Optics Letters*, Vol. 7, 2009, pp. 302-308.
 - [18]. Y. Matsuzaki, T. Okamoto, M. Haraguchi, M. Fukui, M. Nakagaki, Characteristics of gap plasmon waveguide with stub structures, *Opt. Express*, Vol. 21, 2008, pp. 16314-16325.
 - [19]. C. Min, G. Veronis, Theoretical investigation of fabrication-related disorders on the properties of subwavelength metal-dielectric-metal plasmonic waveguides, *Opt. Express*, Vol. 18, 2010, pp. 20939-20948.
 - [20]. D. F. P. Pile, D. K. Gramotnev, Adiabatic and nonadiabatic nanofocusing of plasmons by tapered gap plasmon waveguides, *Appl. Phys. Lett.*, Vol. 89, 2006, 041111.
 - [21]. P. Bai, M.-X. Gu, X.-C. Wei, E.-P. Li, Electrical detection of plasmonic waves using an ultra-compact structure via a nanocavity, *Opt. Express*, Vol. 17, 2009, pp. 24349-2437.
 - [22]. G. Veronis, S. Fan, Bends and splitters in metal-dielectric-metal subwavelength plasmonic waveguides, *Appl. Phys. Lett.*, Vol. 87, Issue 13, 2005, 131102.
 - [23]. Z. Han, A. Y. Elezzabi, V. Van, Experimental realization of subwavelength plasmonic slot waveguides on a silicon platform, *Opt. Lett.*, Vol. 35, Issue 4, 2010, pp. 502-504.
 - [24]. D. F. P. Pile, T. Ogawa, D. K. Gramotnev, Y. Matsuzaki, K. C. Vernon, K. Yamaguchi, T. Okamoto, M. Haraguchi, M. Fukui, Two-dimensionally localized modes of a nanoscale gap plasmon waveguide, *Appl. Phys. Lett.*, Vol. 87, 2005, 26114.
 - [25]. L. Liua, Z. Hana, S. He, Novel surface plasmon waveguide for high integration, *Opt. Express*, Vol. 17, 2005, pp. 6645-6650.
 - [26]. Z. Han, L. Liu, E. Forsberg, Ultra-compact directional couplers and Mach-Zehnder interferometers employing surface plasmon polaritons, *Opt. Commun.*, Vol. 259, Issue 2, 2006, pp. 690-695.
 - [27]. M. Pu, N. Yao, C. Hu, X. Xin, Z. Zhao, C. Wang, X. Luo, Directional coupler and nonlinear mach-zehnder interferometer based on metal-insulator-metal plasmonic waveguide, *Opt. Express*, Vol. 18, 2010, pp. 21030-21037.
 - [28]. Y.-D. Wu, S.-R. Hong, T.-T. Shih, New all-optical switch based on the local nonlinear plasmonic mach-zehnder interferometer waveguides, in *Proceedings of the Progress in Electromagnetics Research Symposium (PIERS'13)*, Sweden, 12-15 Aug. 2013, pp. 452-454.
 - [29]. J. D. Zook, D. Chen, G. N. Otto, Temperature dependence and model of the electro-optic effect in LiNb₃, *Appl. Phys. Lett.*, Vol. 11, 1967, pp. 159-161.
 - [30]. K. S. Yee, Numerical solution of initial boundary value problems involving Maxwell's equations in isotropic media, *IEEE Antennas and Propagation*, Vol. 14, 1966, pp. 302-307.
 - [31]. Y.-J. Tsai, A. Degiron, N. M. Jokerst, D. R. Smith, Plasmonic multi-mode interference couplers, *Opt. Express*, Vol. 17, 2009, pp. 17471-17482.

- [32]. A. D. Rakic, A. B. Djuricic, J. M. Elazar, M. L. Majewski, Optical properties of metallic films for vertical-cavity optoelectronic devices, *App. Optics*, Vol. 37, 1998, pp. 5271-5283.

Chapter 17

Stable Solitons of Higher Order Cubic Quintic Nonlinear Schrödinger Equation with a \mathcal{PT} -Symmetric Potential

İlkay Bakırtaş and İzzet Göksel

17.1. Introduction

In the field of optics, a soliton denotes any optical field that does not change its shape during propagation because of the sensitive balance between linear and nonlinear effects in the medium [1]. Interest in optical solitons has grown steadily in recent years. The field has considerable potential for technological applications, and it presents many exciting research problems from both a fundamental and an applied point of view. Over the past thirty-five years, soliton research has been conducted in fields as diverse as particle physics, molecular biology, quantum mechanics, geology, meteorology, oceanography, astrophysics and cosmology [2]. Nonlinear Schrödinger (NLS) equation is usually defined by the nonlinear dynamics of pulses on a picosecond time scale. A considerable amount of research work has been devoted to the study of nonlinear Schrödinger equation with a variety of nonlinearities. Several methods, numerical and analytical, have been effectively used to handle these problems [3]. The inverse scattering method, Lax pair, Backlund transformation are some of these methods. These works and results have important scientific values and application prospects such as transmitting digital signals over long distances. Mathematical and numerical analysis of NLS equation with further application areas can be found in the reference [2].

Nonlinear Schrödinger equation is usually defined as the propagation of an optical pulse in optical materials and is given as

$$iu_z + u_{xx} + \alpha |u|^2 u = 0. \quad (17.1)$$

In optics, u corresponds to the differentiable complex-valued, slowly varying amplitude of the electric field, u_{xx} corresponds to diffraction, z is a scaled propagation distance and the coefficient α represents the cubic nonlinearity of the medium.

On the other hand, the dynamics of pulses with widths smaller than 1 picosecond cannot be governed by the cubic NLS equation. For example, in a solid state laser, pulses are generated with generated time durations shorter than 10 femtoseconds and the approximation of the standard NLS equation breaks down. In order to describe the dynamics in such systems, higher order dispersion terms are necessary. One needs to consider the third order dispersion (3OD) for performance enhancement along trans-oceanic and trans-continental distances. Also, for short pulse widths where the group velocity dispersion changes, within the spectral bandwidth of the signal, this necessity can no longer be neglected; one needs to take into account the presence of fourth order dispersion (4OD) [4]. From now on, fourth order dispersion term will be denoted as 4OD. Evolution of the ultrashort optical pulses for cubic NLS equation with the presence of the 4OD term was investigated by Karlsson and Höök [5]. Also, in [6] and [7], the dynamics and interactions of bright solitons in an optical fiber with 4OD are investigated and for particular values of the propagation constant (eigenvalue), the existence of the solitons with oscillating tails is reported. Soliton type solutions to cubic NLS equation with higher order corrections are derived in [8, 9]. In [10], 4OD cubic-quintic nonlinear Schrödinger equation without a potential is solved through the extended elliptic sub-equation method. Consequently, many types of exact traveling wave solutions are obtained including bell and kink profile solitary wave solutions, triangular periodic wave solutions and singular solutions. Band-gaps and stability properties of solitons for the higher order NLS equation with a periodic lattice are studied in [11]. Recently, soliton dynamics in the higher order multidimensional NLS equation is deeply analyzed in [12].

The Hamiltonians that are defined in quantum mechanics are required to be Hermitian to ensure the spectra to be real. In recent years, Bender et al. have demonstrated that non-Hermitian Hamiltonians can exhibit entirely real spectra provided that they have the so-called parity-time (\mathcal{PT}) symmetry property [13, 14]. The real part of a \mathcal{PT} symmetric complex potential must be an even function and the imaginary part must be odd. Numerical existence of the optical solitons in \mathcal{PT} -symmetric periodic potentials is shown to be stable over a wide range of potential parameters in [15]. Optical solitons in various \mathcal{PT} -symmetric optical lattices in NLS type equations are investigated by many researchers [16-22].

In this chapter, we numerically investigate higher order (we will refer to 4OD), cubic-quintic nonlinear Schrödinger equation (CQNLS) with a \mathcal{PT} -symmetric optical potential given below:

$$iu_z + u_{xx} + \gamma u_{xxxx} + \alpha |u|^2 u + \beta |u|^4 u + V_{PT} u = 0, \quad (17.2)$$

where γ is the fourth order diffraction coupling constant, α and β represent the cubic and the quintic nonlinearities of the medium respectively and V_{PT} is the \mathcal{PT} -symmetric external potential (lattice). Depending on the signs of the nonlinearity coefficients, four different media can be presented, namely self-focusing cubic, self-defocusing quintic ($\alpha > 0$, $\beta < 0$), self-focusing cubic, self-focusing quintic ($\alpha > 0$, $\beta > 0$), self-defocusing cubic, self-defocusing quintic ($\alpha < 0$, $\beta < 0$), self-defocusing cubic, self-focusing quintic ($\alpha < 0$, $\beta > 0$).

17.2. \mathcal{PT} -Symmetric Optical Lattices

Any measurement of a physical observable in our universe obviously yields a real quantity. In quantum mechanics, observables correspond to eigenvalues of operators. Hence, the reality requires all the eigenvalues of operators to be real.

Consider the Hamiltonian operator \hat{H} which is the sum of the kinetic energy operator \hat{T} and potential energy operator \hat{V} :

$$\hat{H} = \hat{T} + \hat{V} = \frac{\hat{p}^2}{2m} + \hat{V}(\hat{x}), \quad (17.3)$$

where \hat{p} is the momentum operator, m is the mass and \hat{x} is the position operator.

Real eigenvalues of (17.3) correspond to a real energy spectrum. To guarantee a real spectrum, it was postulated that all observables corresponded to eigenvalues of Hermitian (i.e. self adjoint) operators by recalling the result from linear algebra that Hermitian matrices have real spectra. In fact, a Hermitian Hamiltonian ensures a real energy spectrum. However, Bender et al. investigated non-Hermitian Hamiltonians in 1998 and found out that many of them have entirely real spectra given that they have the so called parity-time (\mathcal{PT}) symmetry property [13, 14]. Furthermore, they showed in many cases a threshold value above which the spectrum becomes complex. This threshold is the boundary between the \mathcal{PT} -symmetric and broken symmetry phases and in literature, the transition is referred to as spontaneous \mathcal{PT} -symmetry breaking.

\mathcal{PT} -symmetry is defined by means of the parity operator \hat{P} and time (reversal) operator \hat{T} whose actions are given as follows:

$$\begin{aligned} \hat{P}: \hat{p} &\rightarrow -\hat{p}, \hat{x} \rightarrow -\hat{x} & (P(a\psi + b\phi))(x) &= a\psi(-x) + b\phi(-x), \\ \hat{T}: \hat{p} &\rightarrow -\hat{p}, \hat{x} \rightarrow \hat{x}, i \rightarrow -i & (T(a\psi + b\phi))(x) &= a^* \psi^*(x) + b^* \phi^*(x). \end{aligned} \quad (17.4)$$

A Hamiltonian is said to be \mathcal{PT} -symmetric if it has the same eigenfunctions as the $\hat{P}\hat{T}$ operator and satisfies the commutativity

$$\hat{P}\hat{T}\hat{H} = \hat{H}\hat{P}\hat{T}. \quad (17.5)$$

On one hand

$$\begin{aligned} (PTH)(f(x,t)) &= (PT)\left(\frac{p^2}{2m}f(x,t) + V(x)f(x,t)\right) \\ &= P\left(\frac{(-p)^2}{2m}f^*(x,t) + V^*(x)f^*(x,t)\right) \\ &= \frac{p^2}{2m}f^*(-x,t) + V^*(-x)f^*(-x,t), \end{aligned} \quad (17.6)$$

and on the other hand

$$\begin{aligned} (HPT)(f(x,t)) &= (HP)(f^*(x,t)) \\ &= H(f^*(-x,t)) \\ &= \frac{p^2}{2m}f^*(-x,t) + V(x)f^*(-x,t). \end{aligned} \quad (17.7)$$

One speaks of broken \mathcal{PT} -symmetry if Eq. (17.5) is satisfied but the same eigenfunctions are not shared.

\mathcal{PT} -symmetric structures have been realized in optical models governed by NLS type equations by which the propagation distance z replaces time in quantum mechanics [10]. The necessary (but not sufficient) condition in Eq. (17.5) implies

$$\left. \begin{aligned} \hat{H}\hat{P}\hat{T} &= \frac{\hat{p}^2}{2m} + V(x) \\ \hat{P}\hat{T}\hat{H} &= \frac{\hat{p}^2}{2m} + V^*(-x) \end{aligned} \right\} \Rightarrow V(x) = V^*(-x). \quad (17.8)$$

Consider the complex potential

$$V_{PT}(x) = V(x) + iW(x) \quad , \quad V, W \in \mathbb{R}. \quad (17.9)$$

As

$$V_{PT}^*(-x) = V^*(-x) - iW^*(-x) = V(-x) - iW(-x), \quad (17.10)$$

the real part of the potential, $V(x)$ must be an even function and the complex part of the potential, $W(x)$ must be an odd function so that Eq. (17.8) holds.

17.3. Numerical Solution of Higher Order (1+1)D CQNLS Equation

In this subsection, we explain the Spectral Renormalization (SR) method which is modified to find the numerical solution to the (1+1)D higher order cubic-quintic NLS equation with a \mathcal{PT} -symmetric potential.

It is known that various techniques have been used to compute localized solutions such as soliton solutions to nonlinear evolution equations. In this chapter, numerical solutions to Eq. (17.2) are investigated by using the Spectral Renormalization method. This method is essentially a Fourier iteration method that was proposed by Petviashvili [24].

Later, Ablowitz and Musslimani improved this method to a generalized numerical scheme for computing solitons in nonlinear wave guides [25]. Transforming the governing equation into Fourier space and finding a nonlinear, nonlocal integral equation coupled to an algebraic equation is the essence of the method. This coupling prevents the numerical scheme from diverging.

The optical mode is then obtained from an iteration scheme, which converges rapidly. This method can be applied to a large class of problems, which include higher order nonlinear terms with different homogeneity.

Using the ansatz $u(x, z) = f(x)e^{i\mu z}$ where $f(x)$ is a complex-valued function and μ is the propagation constant (or eigenvalue), we have following expressions:

$$\begin{aligned} u_t(x, z) &= i\mu f e^{i\mu z}, \\ u_{xx}(x, z) &= f_{xx} e^{i\mu z}, \\ u_{xxxx}(x, z) &= f_{xxxx} e^{i\mu z}, \\ |u(x, z)|^2 &= f e^{i\mu z} \cdot f e^{-i\mu z} = |f|^2, \\ |u(x, z)|^4 &= |f|^4. \end{aligned} \tag{17.11}$$

Substituting the set of the terms in Eq. (17.11) into Eq. (17.2), the following nonlinear equation for f is obtained

$$-\mu f e^{i\mu z} + f_{xx} e^{i\mu z} + \alpha |f|^2 f e^{i\mu z} + \gamma f_{xxxx} e^{i\mu z} + \beta |f|^4 f e^{i\mu z} + V_{PT} f e^{i\mu z} = 0. \tag{17.12}$$

Canceling the exponential terms yields

$$-\mu f + f_{xx} + \alpha |f|^2 f + \gamma f_{xxxx} + \beta |f|^4 f + V_{PT} f = 0. \tag{17.13}$$

By applying Fourier transformation to Eq. (17.13), we obtain

$$\mathcal{F}\{-\mu f\} + \mathcal{F}\{f_{xx}\} + \mathcal{F}\{\alpha |f|^2 f\} + \mathcal{F}\{\gamma f_{xxxx}\} + \mathcal{F}\{\beta |f|^4 f\} + \mathcal{F}\{V_{PT} f\} = \mathcal{F}\{0\}, \quad (17.14)$$

where \mathcal{F} denotes Fourier transform. Considering the properties of the Fourier transform, we obtain

$$-\mu \hat{f} + (-ik_x)^2 \hat{f} + \alpha \mathcal{F}\{|f|^2 f\} + \gamma (-ik_x)^4 \hat{f} + \beta \mathcal{F}\{|f|^4 f\} + \mathcal{F}\{(V + iW)f\} = 0, \quad (17.15)$$

here $\mathcal{F}(f) = \hat{f}$ and k_x is the Fourier variable. Solving Eq. (17.15) for \hat{f} yields

$$\hat{f} = \frac{\alpha \mathcal{F}\{|f|^2 f\} + \beta \mathcal{F}\{|f|^4 f\} + \mathcal{F}\{(V + iW)f\}}{\mu + k_x^2 - \gamma k_x^4}. \quad (17.16)$$

In order to find $f(x)$, this equation could be indexed and utilized, but the scheme does not converge. At this point, we define a new field variable $f(x) = \lambda w(x)$ where $\lambda \in R^+$ is a parameter to be determined. The equation with the new variable w can be written as

$$\lambda \hat{w} = \frac{\alpha \mathcal{F}\{|w|^2 | \lambda|^2 w \lambda + \beta \mathcal{F}\{|w|^4 | \lambda|^4 w \lambda + \mathcal{F}\{(V + iW)\lambda w\}}{\mu + k_x^2 - \gamma k_x^4}. \quad (17.17)$$

After simplifying this equation, we get

$$\hat{w} = \frac{\alpha \mathcal{F}\{|w|^2 | \lambda|^2 w\} + \beta \mathcal{F}\{|w|^4 | \lambda|^4 w\} + \mathcal{F}\{(V + iW)w\}}{\mu + k_x^2 - \gamma k_x^4}. \quad (17.18)$$

In order to solve for w , Eq. (17.18) can be utilized in an iterative method. For this purpose, we can calculate \hat{w} using the following iteration approach:

$$\hat{w}_{n+1} = \frac{\alpha |\lambda|^2 \mathcal{F}\{|w_n|^2 w_n\} + \beta |\lambda|^4 \mathcal{F}\{|w_n|^4 w_n\} + \mathcal{F}\{(V + iW)w_n\}}{\mu + k_x^2 - \gamma k_x^4}, \quad n \in N, \quad (17.19)$$

with the initial condition taken as a Gaussian type function: $w_0 = e^{-x^2}$, where the convergence criterion is $|w_{n+1} - w_n| < 10^{-12}$.

By multiplying both sides of Eq. (17.18) by $(\mu + k_x^2 - \gamma k_x^4)$, we obtain

$$(\mu + k_x^2 - \gamma k_x^4) \hat{w} - |\lambda|^2 \alpha \mathcal{F}\{|w|^2 w\} - |\lambda|^4 \beta \mathcal{F}\{|w|^4 w\} - \mathcal{F}\{(V + iW)w\} = 0. \quad (17.20)$$

Multiplying Eq. (17.20) by the conjugate of \hat{w} , i.e. by \hat{w}^* yields

$$\begin{aligned} & (\mu + k_x^2 - \gamma k_x^4) |w|^2 - |\lambda|^2 \alpha \mathcal{F} \{|w|^2 w\} \hat{w}^* \\ & - |\lambda|^4 \beta \mathcal{F} \{|w|^4 w\} \hat{w}^* - \mathcal{F} \{(V + iW)w\} \hat{w}^* = 0. \end{aligned} \quad (17.21)$$

Integrating Eq. (17.21) leads to

$$\begin{aligned} & - \int_{-\infty}^{\infty} \left[-\mathcal{F} \{(V + iW)w\} \hat{w}^* + (\mu + k_x^2 - \gamma k_x^4) |w|^2 \right] dk \\ & + |\lambda|^2 \int_{-\infty}^{\infty} \alpha \mathcal{F} \{|w|^2 w\} \hat{w}^* dk + |\lambda|^4 \int_{-\infty}^{\infty} \beta \mathcal{F} \{|w|^4 w\} \hat{w}^* dk = 0. \end{aligned} \quad (17.22)$$

Eq. (17.22) is nothing but a fourth order polynomial in terms of λ , i.e. $P(\lambda) = a\lambda^4 + b\lambda^2 + c$ and λ can be calculated as

$$\lambda_{1,2} = \pm \sqrt{\frac{-b \pm \sqrt{b^2 - 4ac}}{2a}}, \quad (17.23)$$

where

$$\begin{aligned} a &= \beta \int_{-\infty}^{\infty} \mathcal{F} \{|w|^4 w\} \hat{w}^* dk, \\ b &= \alpha \int_{-\infty}^{\infty} \mathcal{F} \{|w|^2 w\} \hat{w}^* dk, \\ c &= - \int_{-\infty}^{\infty} \left[-\mathcal{F} \{(V + iW)w\} \hat{w}^* + (\mu + k_x^2 - \gamma k_x^4) |w|^2 \right] dk. \end{aligned} \quad (17.24)$$

Then, the required soliton will be computed as $f(x) = \lambda(wx) = \lambda \mathcal{F}^{-1}(\hat{w})$.

17.4. Linear Stability Analysis via Linear Spectrum

Next, we turn our attention to the question of stability. In order to investigate the stability of the obtained solitons with SR method that is explained in Section 17.3, we analyze the linear spectrum. In this section, we derive the linear spectrum analysis for NLS equation with fourth order dispersion, a potential V and general type of nonlinearities.

Consider the following (1+1)D nonlinear Schrödinger equation:

$$iu_z(x, z) + u_{xx}(x, z) + \gamma u_{xxxx}(x, z) + F(|u(x, z)|^2)u(x, z) + V(x)u(x, z) = 0, \quad (17.25)$$

where $F(\cdot) \in \mathbb{R}$ and $F(0) = 0$. As seen before, (17.25) admits soliton solutions of the form

$$u(x, z) = f(x) \cdot e^{i\mu z}, \quad (17.26)$$

where $f(x)$ is a complex-valued function and $\lim_{(x,y) \rightarrow \pm\infty} f(x) = 0$.

Substituting

$$\begin{aligned} u_z &= f \cdot e^{i\mu z} \cdot i\mu, \\ u_{xx} &= f_{xx} \cdot e^{i\mu z}, \\ u_{xxx} &= f_{xxx} \cdot e^{i\mu z}, \\ |u|^2 &= u \cdot u^* = f \cdot e^{i\mu z} f^* \cdot e^{-i\mu z} = f \cdot f^* = |f|^2, \end{aligned} \quad (17.27)$$

into Eq. (17.25) and multiplying by $e^{-i\mu z}$ leads to

$$-\mu f + f_{xx} + \gamma f_{xxx} + F(|f|^2)f + V \cdot f = 0. \quad (17.28)$$

To analyze the linear stability, the soliton solution is perturbed as follows

$$u(x, z) = \left[f(x) + g(x)e^{\lambda z} + h^*(x)e^{\lambda^* z} \right] e^{i\mu z}, \quad (17.29)$$

where g and h are perturbation eigenfunctions and λ is the eigenvalue.

$$\begin{aligned} u_z &= \left(\lambda g \cdot e^{\lambda z} + \lambda^* h^* e^{\lambda^* z} + i\mu f + i\mu g \cdot e^{\lambda z} + i\mu h^* e^{\lambda^* z} \right) e^{i\mu z}, \\ u_{xx} &= \left(f_{xx} + g_{xx} e^{\lambda z} + h_{xx}^* e^{\lambda^* z} \right) e^{i\mu z}, \\ u_{xxx} &= \left(f_{xxx} + g_{xxx} e^{\lambda z} + h_{xxx}^* e^{\lambda^* z} \right) e^{i\mu z}, \end{aligned} \quad (17.30)$$

$$\begin{aligned} |u|^2 &= u \cdot u^* = \left(f + g \cdot e^{\lambda z} + h^* e^{\lambda^* z} \right) e^{i\mu z} \left(f^* + g^* e^{\lambda^* z} + h \cdot e^{\lambda z} \right) e^{-i\mu z} \\ &= f \cdot f^* + f \cdot g^* e^{\lambda^* z} + f \cdot h \cdot e^{\lambda z} + f^* g \cdot e^{\lambda z} + g \cdot g^* e^{(\lambda + \lambda^*)z} \\ &\quad + g \cdot h \cdot e^{2\lambda z} + f^* h^* e^{\lambda^* z} + g^* h^* e^{2\lambda^* z} + h \cdot h^* e^{(\lambda + \lambda^*)z} \\ &\simeq |f|^2 + \left(g^* e^{\lambda^* z} + h \cdot e^{\lambda z} \right) f + \left(g \cdot e^{\lambda z} + h^* e^{\lambda^* z} \right) f^*. \end{aligned} \quad (17.31)$$

Using linear Taylor expansion $F(x + h) = F(x) + h \cdot F'(x) + O(h^2)$,

$$\begin{aligned} F(|u|^2) &= F(|f|^2 + \left[\left(g^* e^{\lambda^* z} + h \cdot e^{\lambda z} \right) f + \left(g \cdot e^{\lambda z} + h^* e^{\lambda^* z} \right) f^* \right]) \\ &\simeq F(|f|^2) + \left[\left(g^* e^{\lambda^* z} + h \cdot e^{\lambda z} \right) f + \left(g \cdot e^{\lambda z} + h^* e^{\lambda^* z} \right) f^* \right] F'(|f|^2). \end{aligned} \quad (17.32)$$

Hence,

$$\begin{aligned}
 F(|u|^2)u &= \left\{ \begin{aligned} &F(|f|^2)f + \left[\left(g^* e^{\lambda^* z} + h \cdot e^{\lambda z} \right) f^2 + \left(g \cdot e^{\lambda z} + h^* e^{\lambda^* z} \right) |f|^2 \right] F'(|f|^2) \\ &+ F(|f|^2) g \cdot e^{\lambda z} \\ &+ \left[\left(g \cdot g^* e^{(\lambda+\lambda^*)z} + g \cdot h \cdot e^{2\lambda z} \right) f + \left(g^2 \cdot e^{2\lambda z} + g \cdot h^* e^{(\lambda+\lambda^*)z} \right) f^* \right] F'(|f|^2) \\ &+ F(|f|^2) h^* e^{\lambda^* z} \\ &+ \left[\left(g^* h^* e^{2\lambda^* z} + |h|^2 e^{(\lambda+\lambda^*)z} \right) f + \left(g \cdot h^* e^{(\lambda+\lambda^*)z} + (h^*)^2 e^{2\lambda^* z} \right) f^* \right] F'(|f|^2) \end{aligned} \right\} e^{i\mu z} \\
 &\simeq \left\{ F(|f|^2) \left[f + g \cdot e^{\lambda z} + h^* e^{\lambda^* z} \right] + F'(|f|^2) \left[\left(f^2 h + |f|^2 g \right) e^{\lambda z} + \left(f^2 g^* + |f|^2 h^* \right) e^{\lambda^* z} \right] \right\} e^{i\mu z}.
 \end{aligned} \tag{17.33}$$

By substituting Eq. (17.29), (17.30) and (17.33) into Eq. (17.5), we obtain

$$\begin{aligned}
 &i \left(\lambda g \cdot e^{\lambda z} + \lambda^* h^* e^{\lambda^* z} + i\mu f + i\mu g \cdot e^{\lambda z} + i\mu h^* e^{\lambda^* z} \right) e^{i\mu z} \\
 &+ \left(f_{xx} + g_{xx} e^{\lambda z} + h_{xx}^* e^{\lambda^* z} \right) e^{i\mu z} \\
 &+ \gamma \left(f_{xxxx} + g_{xxxx} e^{\lambda z} + h_{xxxx}^* e^{\lambda^* z} \right) e^{i\mu z} \\
 &+ \left\{ \begin{aligned} &F(|f|^2) \left[f + g \cdot e^{\lambda z} + h^* e^{\lambda^* z} \right] \\ &+ F'(|f|^2) \left[\left(f^2 h + |f|^2 g \right) e^{\lambda z} + \left(f^2 g^* + |f|^2 h^* \right) e^{\lambda^* z} \right] \end{aligned} \right\} e^{i\mu z} \\
 &+ V \cdot \left(f + g e^{\lambda z} + h^* e^{\lambda^* z} \right) e^{i\mu z} = 0.
 \end{aligned} \tag{17.40}$$

Grouping the terms yields

$$\begin{aligned}
 &\left[-\mu f + f_{xx} + \gamma f_{xxxx} + F(|f|^2)f + V \cdot f \right] \\
 &+ \left[i\lambda g - \mu g + g_{xx} + \gamma g_{xxxx} + F(|f|^2)g + \left(f^2 h + |f|^2 g \right) F'(|f|^2) + V \cdot g \right] e^{\lambda z} \\
 &+ \left[i\lambda^* h^* - \mu h^* + h_{xx}^* + \gamma h_{xxxx}^* + F(|f|^2)h^* + \left(f^2 g^* + |f|^2 h^* \right) F'(|f|^2) + V \cdot h^* \right] e^{\lambda^* z} = 0.
 \end{aligned} \tag{17.41}$$

Here, the first bracket is identically zero as f is a solution. For Eq. (17.41) to hold true, the factors of the exponentials must be equal to zero. Hence, one has

$$i\lambda g - \mu g + g_{xx} + \gamma g_{xxxx} + F(|f|^2)g + \left(f^2 h + |f|^2 g \right) F'(|f|^2) + V \cdot g = 0, \tag{17.42}$$

which can be rewritten as

$$g_{xx} + \gamma g_{xxxx} + \left[F(|f|^2) + F'(|f|^2)|f|^2 - \mu + V \right] g + F'(|f|^2) f^2 h = -i\lambda g, \quad (17.43)$$

and on the other hand

$$i\lambda^* h^* - \mu h^* + h_{xx}^* + \gamma h_{xxxx}^* + F(|f|^2) h^* + (f^2 g^* + |f|^2 h^*) F'(|f|^2) + V \cdot h^* = 0, \quad (17.44)$$

which can be rewritten as

$$h_{xx}^* + \gamma h_{xxxx}^* + \left[F(|f|^2) + F'(|f|^2)|f|^2 - \mu + V \right] h^* + F'(|f|^2) f^2 g^* = -i\lambda^* h^*. \quad (17.45)$$

Taking the conjugate of Eq. (17.45) gives

$$h_{xx} + \gamma h_{xxxx} + \left[F(|f|^2) + F'(|f|^2)|f|^2 - \mu + V^* \right] h + F'(|f|^2) (f^2)^* g = i\lambda h. \quad (17.46)$$

Multiplying Eq. (17.46) by -1 gives

$$-h_{xx} - \gamma h_{xxxx} - \left[F(|f|^2) + F'(|f|^2)|f|^2 - \mu + V^* \right] h - F'(|f|^2) (f^2)^* g = -i\lambda h. \quad (17.47)$$

Writing Eq. (17.43) and Eq. (17.47) in matrix form yields

$$\begin{bmatrix} \partial_{xx} + \gamma \partial_{xxxx} + F(|f|^2) + F'(|f|^2)|f|^2 - \mu + V & F'(|f|^2) f^2 \\ -F'(|f|^2) (f^2)^* & -\partial_{xx} - \gamma \partial_{xxxx} - F(|f|^2) - F'(|f|^2)|f|^2 + \mu - V^* \end{bmatrix} \begin{bmatrix} g \\ h \end{bmatrix} = -i\lambda \begin{bmatrix} g \\ h \end{bmatrix}, \quad (17.48)$$

or in a shorter form as

$$i \begin{bmatrix} L_1 & L_2 \\ -L_2^* & -L_1^* \end{bmatrix} \begin{bmatrix} g \\ h \end{bmatrix} = \lambda \begin{bmatrix} g \\ h \end{bmatrix}, \quad (17.49)$$

where

$$\begin{aligned} L_1 &= \partial_{xx} + \gamma \partial_{xxxx} + F(|f|^2) + F'(|f|^2)|f|^2 - \mu + V \\ L_2 &= F'(|f|^2) f^2. \end{aligned} \quad (17.50)$$

For the cubic-quintic nonlinearity,

$$\begin{aligned} F(x) &= \alpha x + \beta x^2, \\ F'(x) &= \alpha + 2\beta x. \end{aligned} \quad (17.51)$$

Using (17.51) in (17.48) yields

$$\begin{aligned} & \begin{bmatrix} \partial_{xx} + \gamma \partial_{xxxx} + 2\alpha |f|^2 + 3\beta |f|^4 - \mu + V & (\alpha + 2\beta |f|^2) f^2 \\ -(\alpha + 2\beta |f|^2)(f^2)^* & -\partial_{xx} - \gamma \partial_{xxxx} - 2\alpha |f|^2 - 3\beta |f|^4 + \mu - V^* \end{bmatrix} \begin{bmatrix} g \\ h \end{bmatrix} \\ &= -i\lambda \begin{bmatrix} g \\ h \end{bmatrix}, \end{aligned} \quad (17.52)$$

or in a shorter form as

$$i \begin{bmatrix} L_1 & L_2 \\ -L_2^* & -L_1^* \end{bmatrix} \begin{bmatrix} g \\ h \end{bmatrix} = \lambda \begin{bmatrix} g \\ h \end{bmatrix}, \quad (17.53)$$

where

$$\begin{aligned} L_1 &= \partial_{xx} + \gamma \partial_{xxxx} + 2\alpha |f|^2 + 3\beta |f|^4 - \mu + V, \\ L_2 &= \alpha f^2 + 2\beta |f|^2 f^2. \end{aligned} \quad (17.54)$$

For only the cubic nonlinearity, by taking $\beta = 0$, Eq. (17.54) reduces to

$$\begin{aligned} L_1 &= \partial_{xx} + \gamma \partial_{xxxx} + 2\alpha |f|^2 - \mu + V, \\ L_2 &= \alpha f^2. \end{aligned} \quad (17.55)$$

For only the quintic nonlinearity, by taking $\alpha = 0$, Eq. (17.54) reduces to

$$\begin{aligned} L_1 &= \partial_{xx} + \gamma \partial_{xxxx} + 3\beta |f|^4 - \mu + V, \\ L_2 &= 2\beta |f|^2 f^2. \end{aligned} \quad (17.56)$$

If the soliton and potential are real i.e. $f, V \in \mathbb{R}$, Eq. (17.52) becomes

$$\begin{aligned} & \begin{bmatrix} \partial_{xx} + \gamma \partial_{xxxx} + 2\alpha f^2 + 3\beta f^4 - \mu + V & \alpha f^2 + 2\beta f^4 \\ -\alpha f^2 - 2\beta f^4 & -\partial_{xx} - \gamma \partial_{xxxx} - 2\alpha f^2 - 3\beta f^4 + \mu - V \end{bmatrix} \begin{bmatrix} g \\ h \end{bmatrix} \\ &= -i\lambda \begin{bmatrix} g \\ h \end{bmatrix}, \end{aligned} \quad (17.57)$$

or in short

$$i \begin{bmatrix} L_1 & L_2 \\ -L_2 & -L_1 \end{bmatrix} \begin{bmatrix} g \\ h \end{bmatrix} = \lambda \begin{bmatrix} g \\ h \end{bmatrix}, \quad (17.58)$$

where

$$\begin{aligned} L_1 &= \partial_{xx} + \gamma \partial_{xxxx} + (2\alpha + 3\beta f^2) f^2 - \mu + V, \\ L_2 &= (\alpha + 2\beta f^2) f^2. \end{aligned} \quad (17.59)$$

In any case, the eigenvalues λ can be calculated numerically by some eigenvalue algorithm [21]. If any of the calculated eigenvalues in the spectrum has a positive real part, then the soliton will blow up as z grows due to the exponential term in the perturbation in Eq. (4.5); in other words, the soliton is linearly unstable. If however the spectrum only consists of pure imaginary eigenvalues, the perturbations will only cause oscillations and in this case, one can speak of linear stability [22].

17.5. Nonlinear Stability Analysis

If a soliton is considered as nonlinearly stable, then it should preserve its shape, location and maximum amplitude during direct simulations. To analyze the nonlinear stability of solitons, we directly compute Eq. (17.2) over a long distance ($z = 100$). In order to do this, split-step Fourier method is employed to advance in z [26].

The split-step Fourier method is a pseudo-spectral numerical method for solving nonlinear PDE's like the NLS equation due to its easy implementation and speed compared to other methods, notably finite difference methods [27]. It is actually a split-step method whose linear step is taken in the frequency domain while the nonlinear step is taken in the time domain. The name comes from the Fourier and inverse Fourier transforms which are necessary for going back and forth between these domains. Eq. (17.2) can be rewritten as

$$u_z = i(\partial_{xx} + \gamma \partial_{xxxx})u + i(\alpha |u|^2 + \beta |u|^4 + V_{PT})u, \quad (17.60)$$

and hence can be split with the linear operator $M = i(\partial_{xx} + \gamma \partial_{xxxx})$ and the nonlinear operator $N = i(\alpha |u|^2 + \beta |u|^4 + V_{PT})$. The linear step $u_z = Mu$ is solved by means of Fourier transform. Taking the Fourier transform of both sides of

$$u_z = i(\partial_{xx} + \gamma \partial_{xxxx})u, \quad (17.61)$$

gives

$$\hat{u}_z = i((-ik_x)^2 + \gamma(-ik_x)^4)\hat{u} = i(-k_x^2 + \gamma k_x^4)\hat{u}. \quad (17.62)$$

This is nothing but an ordinary differential equation (ODE) of \hat{u} and its exact solution is given by

$$\hat{u} = \hat{C}_1 \exp\left[i(-k_x^2 + \gamma k_x^4)z\right] \Rightarrow u = \mathcal{F}^{-1}\left(\hat{C}_1 \exp\left[i(-k_x^2 + \gamma k_x^4)z\right]\right). \quad (17.63)$$

The nonlinear step $u_z = Nu$, i.e.

$$u_z = i(\alpha|u|^2 + \beta|u|^4 + V_{PT})u, \quad (17.64)$$

has the exact solution

$$u = C_2 \exp\left[i(\alpha|u|^2 + \beta|u|^4 + V_{PT})z\right]. \quad (17.65)$$

Having found solutions to both parts, the split-step Fourier method can now be employed for the CQNLS with 4OD equation by using any splitting scheme.

17.6. Cubic Nonlinear Schrödinger Equation with Fourth Order Dispersion

In this subsection, we briefly explain the previous findings about cubic NLS equation with 4OD.

In [5], cubic NLS equation with negative fourth order dispersion term is analyzed for $\varepsilon > 0$:

$$iu_z + u_{xx} - \varepsilon u_{xxxx} + |u|^2 u = 0. \quad (17.66)$$

Eq. (17.66) has an exact, stationary solution:

$$u(x, z) = \sqrt{\frac{3}{10\varepsilon}} \operatorname{sech}^2\left(\frac{x}{\sqrt{20\varepsilon}}\right) \exp\left(i \frac{4}{25\varepsilon} z\right). \quad (17.67)$$

The effect of the fourth order dispersion on the shape and stability of the soliton solutions is investigated in [5] and [6, 7]. In [28, 29], the effect of the 4OD on the shape of the soliton is demonstrated and solitons with oscillating tails are reported. We should note that there are no localized solutions to Eq. (17.66) for $\varepsilon < 0$. In Fig. 17.1, mode profiles corresponding to various eigenvalues ($\mu = 0.16, 1, 5$) are depicted for fixed $\varepsilon = 1$. As it can be seen from this figure, as the propagation constant (eigenvalue) μ increases, maximum amplitude increases as well and the soliton becomes more localized; moreover, oscillatory behavior of the soliton tails becomes more pronounced.

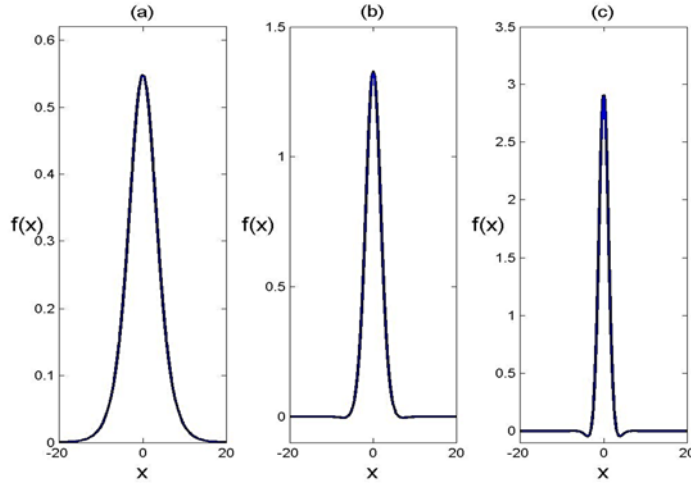


Fig. 17.1. Soliton profiles for NLS with 4OD, fixed $\varepsilon = 1$ and various eigenvalues:
(a) $\mu = 0.16$, (b) $\mu = 1$, (c) $\mu = 5$.

17.7. Cubic Quintic Nonlinear Schrödinger Equation with Fourth Order Dispersion

In this subsection, we investigate the effect of 4OD on the band-gap regions and on the stability properties of the numerically obtained solitons of the cubic-quintic NLS equation (CQNLS) given in Eq. (17.2). In this study, we only consider self-focusing cubic, self-focusing quintic ($\alpha = 1$, $\beta = 1$) medium and negative 4OD ($\gamma < 0$). Following that, the model equation in this study is given as

$$iu_z + u_{xx} + \gamma u_{xxxx} + |u|^2 u + |u|^4 u + V_{PT} u = 0. \quad (17.68)$$

In Fig. 17.2, we demonstrate the soliton solutions of CQNLS-4OD without a potential for fixed 4OD: $\gamma = -1$ and various eigenvalues: $\mu = 1, 4, 7$. As the propagation constant (eigenvalue) increases, the soliton amplitude increases and soliton becomes more localized. On the other hand, compared to NLS-4OD solitons, the amplitudes are smaller.

Recently in [30], one-dimensional cubic-quintic NLS (CQNLS) equation with a \mathcal{PT} -symmetric potential (given in Eq. (17.69)) is investigated both numerically and analytically and an exact soliton solution is obtained. The stability properties of the obtained solitons are also investigated and it is shown that, depending on the signs of the cubic and quintic terms (α and β respectively), there exist stable and unstable regions for various potential depths. In the aforementioned work, for self-focusing cubic, self-focusing quintic media ($\alpha = 1$, $\beta = 1$), majority of the solitons are found to be both linearly and nonlinearly unstable due to the presence of the quintic nonlinearity and non-periodic \mathcal{PT} -symmetric potential.

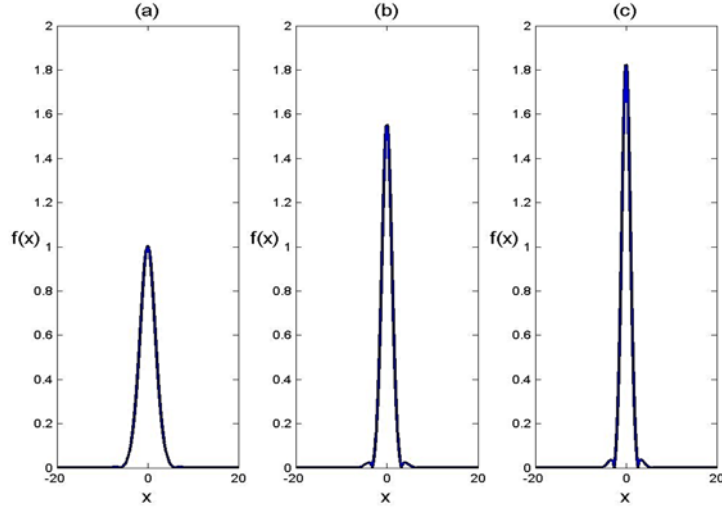


Fig. 17.2. Soliton profiles for CQNLS with 4OD, fixed $\gamma = -1$ and various eigenvalues: (a) $\mu = 1$, (b) $\mu = 4$, (c) $\mu = 7$.

In order to observe the effect of fourth order dispersion on the band-gap regions and on the stability properties of the numerically obtained solitons, we will compare our results with the results found in [30].

The \mathcal{PT} -symmetric potential considered in this work is as follows:

$$V_{\mathcal{PT}} = V(x) + iW(x) = [V_0 \operatorname{sech}^2(x) + V_1 \operatorname{sech}^4(x)] + [W_0 \operatorname{sech}(x) \tanh(x)], \quad (17.69)$$

where V_0 , V_1 and W_0 represent the depths of the real and imaginary parts of \mathcal{PT} -symmetric potential, respectively. This potential is essentially an extension of the so-called complexified Scarff II potential for a Kerr medium and it is of interest due to its physical significance [31]. For more detail and restrictions for the \mathcal{PT} -symmetric potential given in Eq. (17.69) cf. [30]. In Fig 17.3 the real part, imaginary part and cross sections of the \mathcal{PT} -symmetric potential are depicted.

In Fig. 17.4, we plotted the gap regions of CQNLS equation with \mathcal{PT} -symmetric potential for various potential depths and fourth order dispersion parameter γ . One can see from the figure that, as γ gets smaller (stronger fourth order dispersion), gap size increases drastically. In Fig. 17.5, the soliton profiles of CQNLS in a \mathcal{PT} -symmetric potential are plotted on top of the real and the imaginary parts of the potential for various 4OD coefficients. In all three figures, potential depths are taken as $V_0 = 2$, $W_0 = 1.9$, and the eigenvalue is taken as $\mu = 1$. This figure reveals that, as the coefficient of 4OD gets smaller (more negative), soliton becomes less localized.

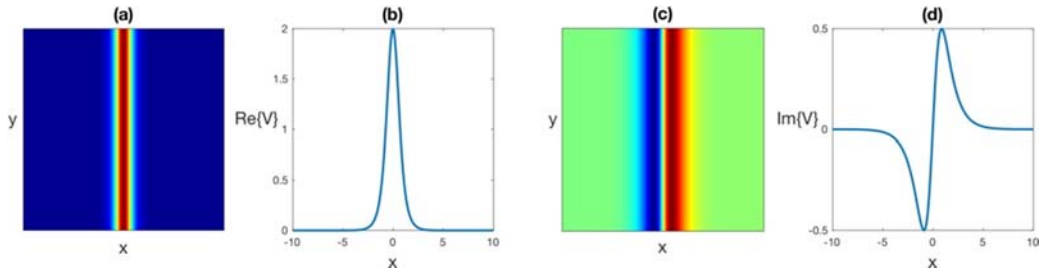


Fig. 17.3. (a) Real part; (b) Cross section of the real part; (c) Imaginary part; and (d) Cross section of the imaginary part of the \mathcal{PT} -symmetric potential with $V_0 = V_1 = W_0 = 1$.

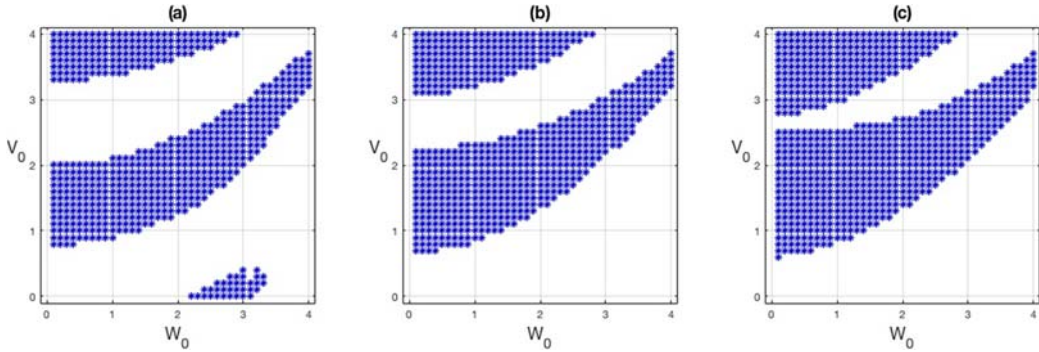


Fig. 17.4. Numerically obtained solitons of (a) CQNLS equation without 4OD; (b) CQNLS equation with 4OD for $\gamma = -0.5$; and (c) CQNLS equation with 4OD for $\gamma = -1$ for varying potential depths of the \mathcal{PT} -symmetric potential.

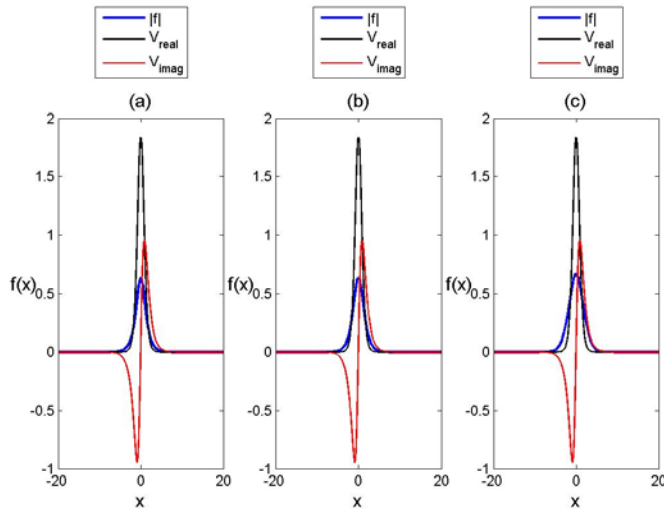


Fig. 17.5. Numerically obtained solitons on top of the \mathcal{PT} -symmetric potential with $V_0 = 2$ and $W_0 = 1.9$ for (a) CQNLS equation without 4OD; (b) CQNLS equation with 4OD for $\gamma = -0.1$, and (c) CQNLS equation with 4OD for $\gamma = -1$.

In order to demonstrate the effect of the increasing eigenvalue and 4OD on the soliton shape and amplitude, we plotted the soliton profiles corresponding to $\mu=1$ and $\mu=7$ in Fig. 17.6 (a) and (b), respectively. As it is seen from the figure, as the eigenvalue increases, soliton amplitude increases as well and as a result of this fact, soliton becomes more localized. This figure also reveals that, if 4OD effect is considered ($\gamma=-1$), soliton becomes less steep. On the other hand, oscillatory tails of the higher order dispersion soliton corresponding to $\gamma=-1$ become more pronounced as the eigenvalue increases.

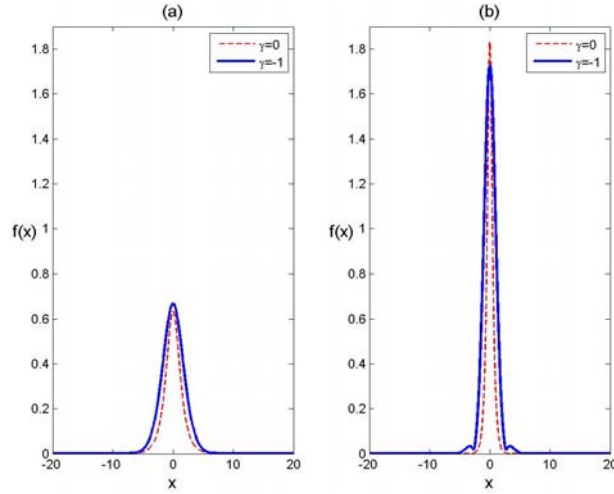


Fig. 17.6. Numerically obtained solitons of CQNLS equation with/without 4OD in \mathcal{PT} -symmetric potential with $V_0 = 2$ and $W_0 = 1.9$ for varying eigenvalues: (a) $\mu=1$; (b) $\mu=7$.

Now, the important question about the stability arises: What is the effect of the higher order negative dispersion on the stability of the solitons obtained in a \mathcal{PT} -symmetric potential? To investigate the effect of 4OD to nonlinear stability, obtained solitons are computed over a long distance ($z=100$). If a soliton preserves its shape, location and maximum amplitude during direct simulations, it is considered to be nonlinearly stable. The results of the nonlinear stability analyses are depicted in Fig. 17.7. This figure reveals that, as the effect of the 4OD gets stronger, more stable solitons can be obtained. So, one can conclude from this figure that, negative 4OD improves the nonlinear stability of the solitons and for $\gamma=-1$, majority of the solitons become nonlinearly stable in comparison to $\gamma=0$ (without 4OD).

To demonstrate this fact, we picked a nonlinearly unstable CQNLS (without 4OD) soliton shown in Fig. 17.7 (a) corresponding to $\mu=1$ and potential depths $V_0 = 2$ and $W_0 = 1.9$. In Fig. 17.8, we demonstrate the nonlinear evolution of this soliton and it can be seen from this figure that, as the propagation distance increases, the maximum amplitude of the

soliton becomes highly oscillatory and after $z = 50$, the shape of the soliton is deteriorated which leads to nonlinear instability for the chosen potential depths.

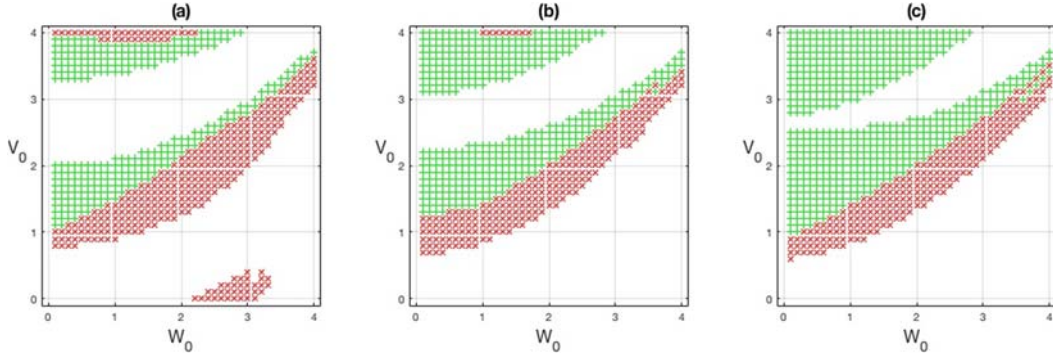


Fig. 17.7. Nonlinearly stable (marked as green) and nonlinearly unstable (marked as red) solitons of (a) CQNLS equation without 4OD; (b) CQNLS equation with 4OD for $\gamma = -0.5$; (c) CQNLS equation with 4OD for $\gamma = -1$, for varying potential depths of the \mathcal{PT} -symmetric potential.

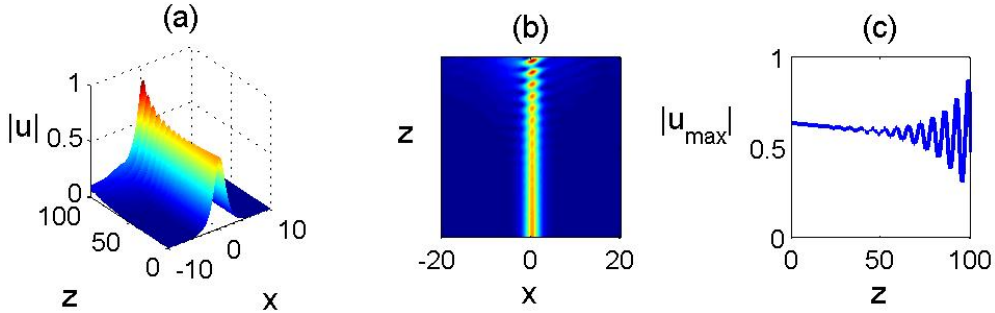


Fig. 17.8. (a) Nonlinear evolution of the CQNLS (without 4OD) soliton in a \mathcal{PT} -symmetric potential for $V_0 = 2$ and $W_0 = 1.9$; (b) View from the top; (c) Maximum amplitude as a function of the propagation distance z .

In order to see the positive effect of the 4OD to the nonlinear stability, now we set $\gamma = -0.1$ and keep all other parameters fixed. The evolution of this soliton is depicted in Fig. 17.9 and it is observed that the highly oscillatory behavior of the soliton amplitude is now suppressed for a long distance, but after $z \approx 50$, the maximum amplitude of the soliton starts to decrease and a small oscillation in the amplitude is observed. Although this soliton cannot be assumed as a stable soliton, one can see the dramatic change between the Fig. 17.8 and Fig. 17.9.

Now, the effect of the fourth order dispersion is strengthened by setting $\gamma = -1$. In Fig. 17.10, we show the nonlinear evolution of the corresponding soliton. This figure reveals that, for a stronger 4OD, the soliton preserves its shape, location and maximum

amplitude for a long distance. It should be noted that the behavior of the soliton during the evolution does not change drastically for a stronger 4OD, namely $\gamma = -2$ (see Fig. 17.11).

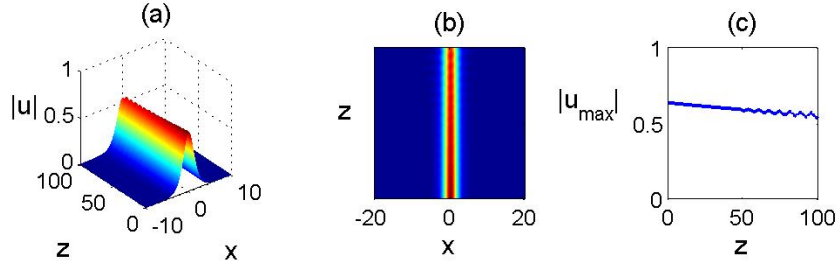


Fig. 17.9. (a) Nonlinear evolution of the CQNLS with 4OD ($\gamma = -0.1$) soliton in a \mathcal{PT} -symmetric potential for $V_0 = 2$ and $W_0 = 1.9$; (b) View from the top; (c) Maximum amplitude as a function of the propagation distance z .

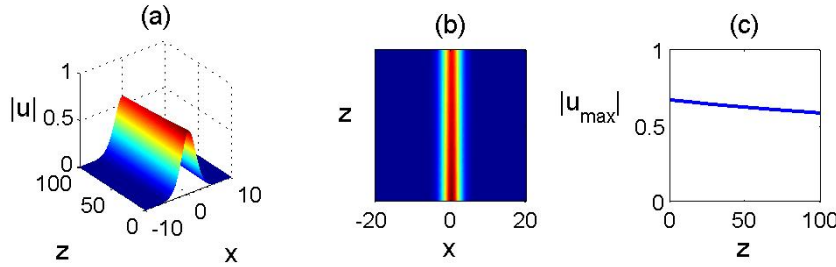


Fig. 17.10. (a) Nonlinear evolution of the CQNLS with 4OD ($\gamma = -1$) soliton in a \mathcal{PT} -symmetric potential for $V_0 = 2$ and $W_0 = 1.9$; (b) View from the top; (c) Maximum amplitude as a function of the propagation distance z .

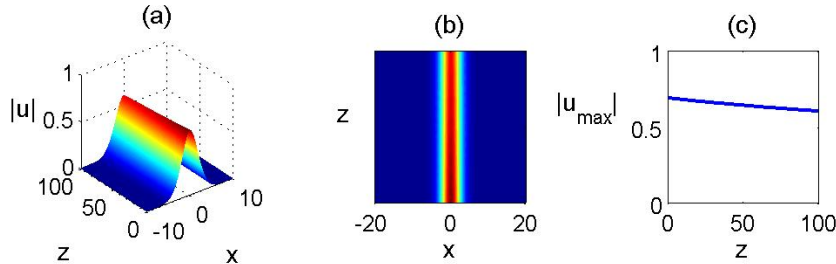


Fig. 17.11. (a) Nonlinear evolution of the CQNLS with 4OD ($\gamma = -2$) soliton in a \mathcal{PT} -symmetric potential for $V_0 = 2$ and $W_0 = 1.9$; (b) View from the top; (c) Maximum amplitude as a function of the propagation distance z .

To investigate the effect of the 4OD to the linear stability of CQNLS solitons, linear spectra of the solitons are investigated starting from solely cubic-quintic nonlinearity without 4OD and extending this to CQNLS with 4OD gradually by setting $\gamma=0$, $\gamma=-0.1$ and $\gamma=-1$. It can be concluded from Fig. 17.12 that, as the effect of the 4OD gets stronger, the eigenvalues with non-zero real parts in the spectra begin to disappear. As a result, all solitons computed for $\gamma=-1$ are found to be linearly stable and this result is in good agreement with the results that are obtained for nonlinear stability.

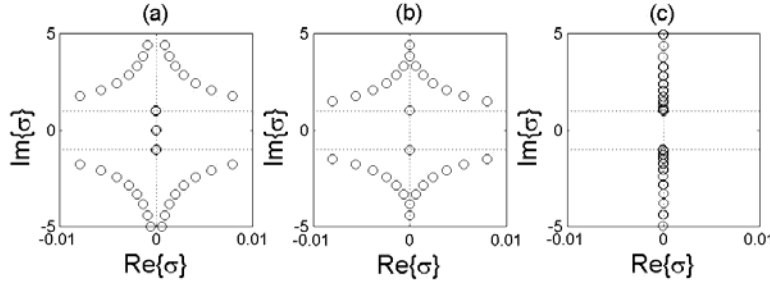


Fig. 17.12. Linear spectrum of (a) CQNLS equation without 4OD; (b) CQNLS equation with 4OD for $\gamma=-0.1$; (c) CQNLS equation with 4OD, for $\gamma=-1$ for $V_0 = 2$ and $W_0 = 1.9$.

17.8. Conclusion

In conclusion, in this chapter, the solitons of CQNLS equation in a \mathcal{PT} -symmetric potential with 4OD are computed numerically by means of the Spectral Renormalization method for various potential depths. These solitons tend to have larger amplitudes and as a result, they become more localized as the eigenvalue (propagation constant) increases. An interesting feature of these solitons is the existence of the oscillatory tails and these tails become more pronounced for larger eigenvalues. It is numerically shown that, adding a stronger 4OD to the model equation enlarges gap size; meaning that the number of the obtained solitons is increased. The linear and nonlinear stability properties of obtained solitons are investigated by means of linear spectrum analysis and by direct simulations. Nonlinear stability regions are depicted for various potential depths. It is observed that 4OD affects both nonlinear and linear stabilities in a positive way.

In this study, only solitons in self-focusing cubic, self-focusing quintic media ($\alpha=1$, $\beta=1$) are investigated. For future studies, solitons in different media and with different \mathcal{PT} -symmetric potentials can be analyzed. Moreover, the methods applied in this study can be extended to higher order models and different type of nonlinearities. Our results can be expanded by adding third order dispersion (3OD) to the model equation. Considering other complex potentials that may provide pulse shaping mechanisms, may also be very interesting and fruitful for future studies.

References

- [1]. Y. S. Kivshar, G. Agrawal, Optical Solitons: From Fibers to Photonic Crystals, *Academic Press*, USA, 2003.
- [2]. M. J. Ablowitz, Nonlinear Dispersive Waves Asymptotic Analysis and Solitons, *Cambridge University Press*, New York, 2011.
- [3]. A. B. Wazwaz, Exact solutions for the fourth order nonlinear Schrodinger equations with a cubic and a power law nonlinearities, *Math. Comp. Mod.*, Vol. 43, 2006, pp. 802-808.
- [4]. R. Kohl, A. Biswas, D. Milovic, E. Zerrad, Optical soliton perturbation in a non-Kerr law media, *Opt. Laser Tech.*, Vol. 40, 2008, pp. 647-662.
- [5]. M. Karlsson, A. Höök, Soliton-like pulses governed by fourth order dispersion in optical fibers, *Opt. Commun.*, Vol. 104, 1994, pp. 303-307.
- [6]. A. V. Buryak, N. N. Akhmediev, Stability criterion for stationary bound states of solitons with radiationless oscillating tails, *Phys. Rev. E*, Vol. 51, 1995, 3572.
- [7]. A. V. Buryak, N. N. Akhmediev, Interactions of solitons with oscillating tails, *Opt. Comm.*, Vol. 121, 1995, pp. 109-114.
- [8]. A. Biswas, D. Milovic, E. Zerrad and F. Majid, Optical solitons in a Kerr law media with fourth order dispersion, *Adv. Studies Theor. Phys.*, Vol. 2, 2008, pp. 1007-1012.
- [9]. H. Triki, T. R. Taha, Solitary wave solutions for a higher order nonlinear Schrödinger equation, *Math. Comp. Sim.*, Vol. 82, 2012, pp. 1333-1340.
- [10]. G. Q. Xu, New types of exact solutions for the fourth-order dispersive cubic-quintic nonlinear Schrödinger equation, *Appl. Math & Comp.*, Vol. 217, 2011, pp. 5967-5971.
- [11]. J. T. Cole, Z. H. Musslimani, Band gaps and lattice solitons for the higher-order nonlinear Schrödinger equation with a periodic potential, *Phys. Rev. A*, Vol. 90, 2014, 013815.
- [12]. J. T. Cole, Z. H. Musslimani, Spectral transverse instabilities and soliton dynamics in the higher-order multidimensional nonlinear Schrodinger equation, *Physica D*, Vol. 313, 2015, pp. 26-36.
- [13]. C. M. Bender, S. Boettcher, Real spectra in Non-Hermitian Hamiltonians having \mathcal{PT} symmetry, *Phys. Rev. Lett.*, Vol. 80, 1998, pp. 5243-5250.
- [14]. C. M. Bender, S. Boettcher, P. N. Meisinger, \mathcal{PT} -symmetric quantum mechanics, *J. Math. Phys.*, Vol. 40, 2008, pp. 2201-2229.
- [15]. Z. H. Musslimani, K. G. Makris, R. El Ganainy, D. N. Christodoulides, Optical solitons in \mathcal{PT} periodic potentials, *Phys. Rev. Lett.*, Vol. 100, 2008, 030402.
- [16]. Z. H. Musslimani, K. G. Makris, R. El Ganainy, D. N. Christodoulides, Analytical solutions to a class of nonlinear Schrödinger equations with \mathcal{PT} -like potentials, *J. Phys. A: Math. Theor.*, Vol. 41, 2008, 244019.
- [17]. K. G. Makris, R. El Ganainy, D. N. Christodoulides, Z. H. Musslimani, \mathcal{PT} -symmetric periodic optical potentials, *Int. J. Theor. Phys.*, Vol. 50, 2011, pp. 1019-1041.
- [18]. K. G. Makris, R. El Ganainy, D. N. Christodoulides, Z. H. Musslimani, \mathcal{PT} -symmetric optical lattices, *Phys. Rev. A*, Vol. 81, 2010, 063807.
- [19]. A. Khare, S. M. Al-Marzoug, H. Bahlouli, Solitons in \mathcal{PT} -symmetric potential with competing nonlinearity, *Phys. Lett. A*, Vol. 376, 2012, pp. 2880-2886.
- [20]. J. Yang, Detecting vortices in superconductors: Extracting one-dimensional topological singularities from a discretized complex scalar field, *Phys. Rev. E*, Vol. 91, 2015, 023201.
- [21]. J. Yang, Symmetry breaking of solitons in one-dimensional parity-time-symmetric optical potentials, *Opt. Lett.*, Vol. 39, 2014, pp. 5547-5550.

- [22]. S. Nixon, L. Ge, J. Yang, Stability analysis for solitons in \mathcal{PT} -symmetric optical lattices, *Phys. Rev. A*, Vol. 85, 2012, 023822.
- [23]. Y. Chen, Z. Yan, D. Mihalache, B. A. Malomed, Families of stable solitons and excitations in the \mathcal{PT} -symmetric nonlinear Schrödinger equations with position-dependent effective masses, *Scientific Reports*, Vol. 7, 2017, 1257.
- [24]. V. I. Petviashvili, Equation of an extraordinary soliton, *Sov. J. Plasma Phys.*, Vol. 2, 1976, pp. 257-258.
- [25]. M. J. Ablowitz, Z. H. Musslimani, Spectral renormalization method for computing self-localized solutions to nonlinear systems, *Opt. Lett.*, Vol. 30, 2005, pp. 2140-2142.
- [26]. J. Yang, Nonlinear Waves in Integrable and Nonintegrable Systems, *SIAM*, PA, USA, 2010.
- [27]. B. Washburn, Dispersion and nonlinearities associated with supercontinuum generation in fibers, (Doctoral dissertation), *Georgia Institute of Technology*, Atlanta, GA, USA, 2002.
- [28]. V. I. Karpman, Stabilization of soliton instabilities by higher-order dispersion: Fourth-order nonlinear Schrödinger-type equations, *Phys Rev. E*, Vol. 53, 1996, pp. R1336-R1339.
- [29]. V. I. Karpman, A. G. Shagalov, Stability of solitons described by nonlinear Schrödinger-type equations with higher-order dispersion, *Physica D*, Vol. 144, 2000, pp. 194-210.
- [30]. İ. Göksel, N. Antar, İ. Bakırtaş, Solitons of (1+1)D cubic-quintic nonlinear Schrödinger equation with \mathcal{PT} -symmetric potentials, *Opt. Comm.*, Vol. 354, 2015, pp. 277-285.
- [31]. B. Bagchi, C. Quesne, An update on the \mathcal{PT} -symmetric complexified Scarf II potential, spectral singularities and some remarks on the rationally extended supersymmetric partners, *J. Phys. A: Math. and Theor.*, Vol. 43, 2010, 305301.

Index

2N-Component Bright Solitons, 183
2N-Component Dark Solitons, 182

A

Aberration, filament, 302
a-dependent evolution, 347, 350, 353, 360, 362, 370, 375, 380
Airy theory revisited, 224
Amplitude
 of rays. *see* Ray, amplitude of
 vectors, 99
 zero, 345, 350, 351, 366,
 see also OV core
Anderson localization, 269-271, 282, 298
Arising, filaments, **310**
asymptotic analytical model, 347, 358, 360, 361, 363, 370, 371, 381, 385-387
attosecond streaking, 42, 43, 45, 51
average power flow, 108

B

Band-gap regions, 452, 464, 465
Bragg, 30, 41, 45, 48, 53, 253-255, 257-259
Branch points, 392, 407
Bright Soliton Pair, 175, 186
Bright-Dark Soliton Pair, 193

C

central projection, 69
centrosymmetric photorefractive media,
 161-165, 167, 173, 197, 198
chirped mirror, 31, 47
chirped-pulse-amplified system
 CPA, 325, 338
circular
 OV beam, 345-347, 352, 356, 357, 380, 381
 symmetry, 346, 349, 363, 380
circularly polarized incident case, 111
Coherence, 260, 262, 265

 Longitudinal, 262
 Transverse, 260
Coherent lasing, 278
coherently coupled soliton, 173
colloids, 233-235, 238, 241-244, 247
conical mounting, 97
constructive interference, 439, 445
convergence, 255, 259, 456
Convolution, 393, 406
coupled-wave equations, 98
Coupling, 233
 efficiency, 443
critical
 angle, 114
 percolation value, 274, 282
cubic crystal, 234, 238, 241, 243, 246, 247
Curl, 392, 403-407
curvature, 255
Curvature matrices
 of dioptric surface, 215
 of wave front, 215
cut-off, 23, 39, 40-42, 45
CXDI, 251-255, 257-260, 262-265

D

Dark Soliton Pair, 171, 188
Deformable Mirror, 391
destructive interference, 439, 440, 445
dielectric
 constants, 235, 237, 239, 240
 ellipsoid, 238, 239, 243, 247
 grating, 97, 112, 114
 Properties, 233, 235
 tensor, 234-237, 241
difference-frequency generation
 DFG, 319
diffraction, 251, 253-255, 258, 260-263, 265, 452, 453
 amplitudes, 108
diffraction imaging, 251, 258, 262, 263, 265
Diffraction of white light on a line grating,
 415

colors of the rainbow in the Fraunhofer
 farfield, 415
 colors of the rainbow in the Fresnel
 nearfield, 415
 Diffraction of white light on a space grating
 colors in the Fraunhofer farfield, 415
 gap of the physics of colors lies in the
 Fresnel nearfield, 415
 Dimlight vision and Purkinje-Shift
 fusion of R+G in RGB at 512 nm, the
 peak of the rod's spectral brightness
 sensitivity curve, 430
 space grating nearfield optics explains
 Purkinje-Shift to rod vision as
 adaptive shift (shortening of the
 grating constant g_z), 430
 Disordered optical materials, 269
 Dispersion, 452, 457, 463, 465, 467, 468
 relation, 438, 440, 443
 fourth order, 452, 463-470
 third order, 452, 470
 Divergence factor, 210, 211, 218
 in VCRM, 218
 Divergence, SC, 312
 Downsampling, 391, 393, 394, 401, 410

E

edge
 diffraction, 345, 356, 380, 382
 wave, 347, 357, 359, 360, 382
 effective refractive index, 320, 321
 Eigenfunction, 454, 458
 eigenvalue, 101, 453, 462-465, 467, 470
 matrix eigenvalue, 104
 degeneracy, 96
 eigenvalues degenerating, 115
 eigenvectors in extended eigenspace, 102,
 103
 matrix, 99, 101
 electric field, 22, 24-27, 42, 45, 50, 51
 Electric field, 452
 electromagnetic wave, 24, 26, 28
 electron
 beam evaporation, 21
 density, 310
 electro-optic
 coefficient, 438
 material, 438
 polymer, 444

energy
 circulation, 345, 348, 349, 351, 353, 354,
 371-373, 377, 378
 conservation, 111
 flux, 275
 kinetic, 453
 potential, 453
 error of approximation, 121
 evanescent incident wave, 110
 Evolution, nonlinear, 455, 467, 468
 existence curve, 173, 176, 180, 187, 190,
 192, 195
 extinction coefficient, 24, 25

F

FDTD technique, 273, 278, 282
 ferrofluids, 233, 234, 240, 244, 247
 Filament,
 conditions, 302
 supercontinuum, 301
 filtering, 21, 23, 24, 27, 37, 42, 45, 47
 finite difference time domain, FDTD, 438,
 439, 445, 446
 flux density, 116
 Focal lines, 211, 217
 phase of, 211, 219
 Fourier
 expansion, 98
 limit, 39, 40, 44, 45
 series, 98
 space, 455
 transform, 391, 392, 401, 456, 462
 Frequency spectrum, 276
 frequency-resolved optical gating
 FROG, 42
 Fresnel equations, 26, 34
 Fresnel formulas, 205, 214
 in vector form, 214
 Fresnel–Kirchhoff integral, 347, 350, 366,
 381
 Fried
 geometry, 397, 398, 401, 404

G

Gaussian
 beam, 327
 function, 152
 pulse, 30, 45

genetic algorithms, 123, 129
Geometrical optics, 204
goethite, 233
Gouy phase shift, 317
Green function, 144, 146
Grey Soliton Pair, 178, 191
group delay dispersion
 GDD, 31, 35, 41, 47

H

Hamiltonian, 452, 453
hard X-ray, 251, 263, 264
Helmholtz equation, 142, 146, 164, 166
Hermite polynomial, 142
Hermite–Gaussian beams, 142
Hermite–Laguerre–Gaussian (EHLG)
 waves, 144
Hermite–Laguerre–Gaussian beams, 137
Hermitian, 452, 453
High Harmonic Generation, 21, 22, 23
homogeneous
 coordinates, 69
 matrix, 69
hybrid genetic algorithm, 130

I

imaging, 251, 257, 260, 262-265
Incipient percolating cluster, 272
incoherently coupled
 multicomponent, 162, 163, 166, 182-184,
 197
 coupled spatial solitons, 162
index of refraction, 24-28
Initial condition, 456
interdiffusion, 32
interface roughness, 32-35, 53
Interference optics (RGB diffraction orders)
 in the Fresnel-nearfield of the ONL-
 space grating, 421
 crystal-optical resonance factor $v\lambda$, a
 condition for resonances in the space
 grating, 424
 double cone directional cosine, 422
 Fresnel-/Talbot interferences on a
 hexagonal space grating, 421
 Fresnel-/Talbot-interferences on a line
 grating, 421

Fresnel-cosine reciprocal to the von-
 Laue cosine, 423
hexagonal space grating cell for the
 fundamental wave $h_1h_2h_3 = 111$, 424
light beam perpendicularly incident into
 the space grating, 423
resonator wavelength of the 111-
 fundamental wave = 559 nm (R in
 RGB), 425
RGB interference maxima located in
 intersections of three pairs of
 hyperbolas, 427
RGB triple corresponds to the peak
 wavelengths of the cone's spectral
 sensitivity curves, 427
RGB-Fresnel nearfield transmission
 system with space grating resonator
 (sender), RGB space waves and RGB-
 receiver antennae, 426
RGB-triple with RGB-wavelength 559
 nm (R), 537 nm (G), 447 nm (B), 425
spatial Pythagoras, a condition for
 resonant wavelengths in the space
 grating, 423
two harmonic resonance wavelengths
 with $h_1h_2h_3 = 122$ and 123 (B+G in
 RGB), 425
intermixing, 32
Inverse participation ratio, 271, 284, 285,
 288
inverse rule, 98
Inverted retina, three nuclear layers
 lightwards before the photoreceptors,
 418
 layered information processing in
 cortical visual areas CGL/V1, 420
 outer nuclear layer ONL, a hexagonal
 cellular space grating, 418
 prenatal differentiation of the three
 nuclear layers INL, MNL, ONL, 418
Ioffe- Regel criterion, 281
ion beam sputtering, 21
irrational angle, 85
isolated attosecond pulses, 23, 37, 42, 46,
 52
iterative improvement, 129

J

jump criterion, 347, 373, 376, 380, 381

K

Kerr

- lens effect, 321, 336
- medium, 465

Kronecker's delta, 99

- Kummer beam, 346-348, 351, 353-357, 359-362, 370, 371, 374, 376, 377, 381, 382, 387

L

- Laguerre Gaussian Beams, 137, 142, 346-349, 352, 355-357, 359-370, 374-379, 381, 385

Lasing

- modes, 269, 276
- start time, 274, 276, 279

Lattice, optical, 452, 453

LED array, 122

left-circularly (LC) polarized wave, 111

Levels of emitters, 271

linearly polarized incident case, 111

local search algorithms, 123

Localization criterion, 283, 298

Localized field, 278, 281, 298

low grazing limit, 95, 109, 112

M

- Mach-Zehnder interferometer, MZI, 437-439, 441, 443-447

magnetic and dielectric properties, 234, 240, 244

magnetic field, 233, 234, 236-239, 242-247

magnetization, 235, 236, 241, 243, 245, 247

magneto-optical effects, 233-235, 238, 240, 243, 247

magnetron sputtering, 21, 36

matrix

- diagonalization, 96
- eigenvalue problem, 99
- eigenvalues method, 95, 97

MAX error, 121

Maxwell equations, 140, 271

Maxwell's equations in dimensionless form, 98

Medium, self-defocusing quintic, 453

cubic, 453, 464, 470

Medium, self-focusing quintic, 453, 464, 470

Meridional, sagittal, focus, 306

merit function, 35, 36, 49

metal/insulator/metal (MIM) PWG, 437-440, 442-444, 447

MI gain. *See* Modulation Instability

Mode, optical, 455, 463

modified diffracted

waves, 95

diffraction amplitudes, 107

diffraction efficiencies, 111

modulation instability, 163, 167-169, 186, 197

modulator, 437, 438, 444, 447

moire effect

lowest spatial frequencies, 74

Moiré Effect, 61

Moire Effect in

3D bodies, 68

3D vodies, 64

cylinder, 82

cylinder (horizontal component), 83

cylindr (vertical component), 84

parallel gratings, 78

parallelepiped, 68, 78

sphere, 68

wedge, 67

Moiré Effect in

3D bodies, 78

Moire Effect,

examples, 64

interesting issues, 85

logical halves of a 3D object, 62

lowest spatial frequwny, 73

planar case, 69

rays of divverent nature, 62

spatial case, 78

Moiré Effect,

around Us, 64

Moiré Lens, 87

Moire

magnification factor, 74

mirror, 82

orientation, 74

patterns, 61

period, 65, 73

phenomenon, 61

probability, 85

spectra,

overlapped, 76
 spectrum, 74
 moire spectrum,
 octagon, 88
 spectral peaks, 75
 Spectral Trajectory, 76
 square, 88
 moire statistics, 85
 moiré wavevector, 63, 72
 Momentum, 453
 multicharged OV, 346, 353, 354, 356, 380
 decomposition, 346, 353, 354, 368, 371
 Multi-dimensional, 393
 multilayer stack, 28, 29, 31, 37, 48

N

N Component Grey Solitons, 184
 nanoparticles, 233, 234, 239, 241, 243, 246, 247
 Nature, appearance, two, zones, 309
 needle optimization, 35, 36, 41, 47
 Névot-Croce, 33, 36, 48
 Newton's method, 123
 NLS equation, 451, 452, 454, 455, 457, 462-464
 NLS equation, cubic, 452, 463
 NLS equation, cubic-quintic, 452, 455, 461, 463-470
 Noble identities, 394, 410
 nonlinear
 microscopy, 317
 refractive index, 319, 321, 328, 330, 337, 339
 nonlinearity
 electrostriction, 321
 fast electronic, 317, 318, 323, 330, 334, 338, 339, 341
 nuclear, 318, 330
 orientational, 317, 330, 338, 339, 341
 thermal, 321
 Nonlinearity, cubic, 452, 461
 Nonlinearity, quintic, 461, 464
 Nonparaxial
 Beams, 138
 parabolic rotational coordinate beams, 148
 Parabolic Rotational Coordinate Beams, 147
 Pearcey Gaussian Beams, 146

Radially Polarized Beams, 138

O

observation plane, 347, 348, 350, 357, 379, 385
 octagon in moire spectrum, 88
 Optical beam propagation, 137
 optical glass
 B270, 330-333, 335, 337
 BK7, 325, 326, 338
 F2, 325, 326, 328
 silica, 325, 326, 328, 330-333, 335, 337-339
 SK11, 325, 326
 Optical
 localized modes, 278
 parametric amplifier
 OPA, 325, 331, 338
Optilayer, 35, 47, 48
 optimization algorithms, 123
 organic solvent
 acetone, 330, 332, 335, 339
 chloroform, 330, 332, 339
 DMSO, 330, 332, 333, 339
 OV (optical vortex), 345-382
 OV core, 345, 346, 349-351, 353-357, 359, 361-365, 369-373, 375, 376, 379, 382, 385
 OV dipole, 347, 372, 374, 378, 379
 annihilation, 346, 347, 353, 372, 375, 378, 379, 381
 birth (emergence), 346, 347, 372, 379, 381
 OV filament (OV line), 347, 354, 362, 378, 380
 OV migration, 345, 346, 353, 356, 361, 362, 372, 376, 378, 380, 382
 OV trajectory, 346, 347, 353-356, 358, 360, 362-381
 3D trajectory, 347, 362-364, 369, 376, 378, 379
 discontinuity, 347, 356, 360, 375, 381, 382
 loops, 356, 360, 376
 pulsations, 356, 360, 376, 381
 spiral, 346, 347, 353, 354, 356, 358, 360, 362-368, 372-377, 380, 381

P

paraxial
 approximation, 138
 beams, 148
 equation, 164, 166
 theory, 138
 partial coherence, 260, 262
 Pedestrian bridges
 parallel gratings, 66
 penetration depth, 26, 39
 Percolating threshold, 270, 278
 period
 number, 29, 30, 35
 thickness, 29, 30, 38, 53
 Perturbation, 458, 462
 Phase
 error, 393
 field, 392
 function, 392
 points, 397, 398, 401
 reconstruction, 391-393, 397, 398, 401
 signal processing, 394
 wrapped, 393, 405
 Phase
 of rays. see Ray, phase of
 wrapping, 403, 404, 406
 phase-matching integral, 323, 324, 331
 photometry, 122
 Photorefractive effect, 161
 photovoltaic solitons, 161, 162
 Plain Coplanar Case
 planar coplanar gratings, 69
 plasmonic waveguide, PWG, 437-444, 447
 Position, 453
 Potential,
 complexified Scarff II, 465
 external, 451-455, 457, 461, 464-470
 primary excitation, 95, 106
 oblique primary incidence, 96
 Prismoire effect in
 prism, 80
 probability of the moire effect, 85
 projected period, 70
 Projection
 matrices, 215
 of grating, 70
 transformation, 70
 propagating incident wave, 109

Propagation, 451, 452, 454, 455, 463, 464, 467-470
 Propagation
 constant, 452, 455, 463, 464, 470
 length, 438, 443, 444
 loss, 439
 matrix, 96, 100, 106
 improved propagation matrix, 104
 PT-symmetry, 452-455, 464-470
 Ptychography, 251, 257-259
 pulse
 train, 51
 duration, SC, 311
 optical, 451, 452, 470
 pulsed laser deposition, 21
 pyroelectric
 effect, 162, 163, 169, 170, 187-190, 192, 195-198
 Photorefractive Materials, 186
 spatial solitons, 162

Q

quadratic electro-optic coefficient, 164
 Quadrature Mirror Filter, 395, 396, 398, 401, 402

R

RABBITT, 23, 48, 51
 radial polarization, 141
 Radially polarized beams, 137
 Random lasing, 270
 rational angle, 85
 Ray, 204
 amplitude in VCRM, 218
 amplitude of, 204, 208, 217
 direction of, 204
 phase of, 205, 208, 211, 219
 polarization of, 204
 properties of, 204, 214
 Rayleigh
 parameter, 327
 range, 317, 323, 325, 327-329
 Rayleigh range, 347, 348, 352, 368, 377
 reconstruction, 254, 255, 258-260
 reflected diffraction efficiencies, 109
 refractive index, 320, 321, 323, 324, 326, 332

Relativization of the color of the local on to the color of the global light conditions in object space
at the aperture of the lens the global information is available, the intensity is equal on all points, 433
grating-optical correlator with global and local information processing, 434
polygonal gratings in the cornea exhibit considerable scattering, 433
scattered global information could be transformed into RGB space together with local image information, 433
Resolution, 262
right-circularly (RC) polarized wave, 111
rigorous coupled wave analysis method, 95
RMS error, 121
Rotational Fields, 391

S

Scalar potential, 393, 403
scattering, 21, 23, 25, 26, 34
scattering factor, 95, 107
screen edge, 346-349, 351-357, 360-362, 364, 366, 370, 371, 373, 375-381
screening solitons, 161, 162
screening-photovoltaic solitons, 162
secondary OV, 346, 356, 358, 360, 361, 367-374, 376-378, 382
second-harmonic generation
SHG, 319
self trapping, 161-163, 173, 176, 180, 187
self-defocusing, 321
self-focusing
SF, 317, 318, 321-323, 328-331, 333-339, 341
SF effect, 317-319, 321-323, 325, 328-330, 333-335, 338, 339, 341
SF model, 339
SF process, 322
Sellmeier, 324, 332
Shack-Hartmann, 397, 401
shadow theory, 95, 102
single interface reflectivity, 27, 29, 38, 53
single mode, 438
single-charged OV, 346, 353, 354, 356, 362, 368, 370, 380
singular skeleton, 346, 347, 350, 353, 357, 370, 374-376, 379, 380, 382

Snell law, 205, 214
in vector form, 214
sodium p-styrenesulfonate homo-polymer, PSSNa, 439, 440, 441, 444-446
Solution, numerical, 455
space charge field, 161, 162, 164, 165, 169, 170, 173
Space grating ONL - and not photochemistry - transforms the visible spectrum into RGB space in color vision, 428
Fresnel nearfield interferences retain object information in the objects images, 429
nearfield optics programs the photochemistry, 428
RGB interference maxima become the optical bases for overlapping Receptive Fields, 429
three nuclear layers with hierarchically graduated cell gratings develop optical pillars, 428
Space, soliton, 309
Space-grating optical chromatic adaptivity from achromatism to trichromatism
retinal space grating become's a 'living crystal', 432
shortening or prolongation of the grating constant g_z , 431
spacing function, 86
spatial
harmonics, 98
light intensity distribution, 121
spectral
phase, 22, 23, 35, 38-41, 43, 45, 47, 48, 52
renormalization method, 455, 470
trajectory in ire spectrum, 76
trajectory in moire spectrum, 76
Spectrum, 452, 453, 457, 462, 469, 470
Spectrum, change, 305
spherical particle, 235, 236, 239
spin coating, 440, 444
Split-step Fourier method, 462, 463
square in moire spectrum, 88
Stability,
linear, 457, 458, 462, 469
nonlinear, 462, 467, 468, 470
steepest descent, 128
Strehl ratio, 392
sum-frequency generation

SFG, 319
 Supercontinuum, 301
 symmetry perturbation, 345, 370, 382
 Symmetry, broken, 453, 454

T

Tail, oscillatory, 463, 467, 468, 470
 thin film, 259
 third harmonic
 TH, 317, 318, 324-328, 331, 335, 338, 339, 341
 TH at interface, 317
 TH curve, 327
 TH intensity, 335, 339
 TH light, 325
 TH peak, 330, 333, 334, 336
 TH profile, 333
 TH signal, 334, 339
 TH wavelength, 332
 third-harmonic generation
 THG, 317-319, 321-324, 327, 328, 330, 331, 336-339, 341
 THG asymmetry, 318
 THG at interface, 317, 325, 329, 339, 341
 THG curve, 326
 THG intensity, 318, 323, 325, 330, 337, 341
 THG law, 328
 THG peak, 318, 326, 328, 333
 THG process, 319, 322
 THG response, 330
 THG signal, 318, 322, 325, 330, 331, 335, 341
 THG slope, 328, 335
 third-order susceptibility, 317-319, 322, 323, 331
 Threshold, 453
 tightly focused, 318, 319, 322, 323, 330
 topological
 charge, 349, 352-355, 361, 371, 374-378
 reactions, 346, 347, 375, 378, 380-382
 transformation matrix, 96, 106
 new transformation matrix, 104
 Transformation of the visible light into
 RGB color space, 416
 by photochemistry of cone and rod photopigments, 416

by space grating optics in the Fresnel nearfield, 417
 spectral brightness sensitivity curves of the photo pigments of cones and rods, 417
 transmitted diffraction efficiencies, 109
 Tree structure, 395, 398
 Two-dimensional, 391, 393, 396, 403, 404
 two-photon absorption, 331

U

upsampling, 259, 394

V

Variable aperture angle of the light cones of the diffracted light
 enlargement of the grating constants g_x and g_y , 431
 VCRM, 204
 for a prolate spheroid, 229
 for an ellipsoid, 227
 for an elliptical cylinder, 225
 for an oblate spheroid, 227
 hyperbolic umbilic foci, 227
 VCRMell2D, software, 227
 Vector
 potential, 392, 403
 vector sum in spectrum, 75
 VG (vortex-generating element), 348, 349, 351, 352, 353, 354, 356, 389
 'virtual' OV, 373-375, 378, 379
 visibility circle, 75, 77

W

water window, 27, 43, 47-49, 52, 53
 Wave front equation, 215
 in a symmetric plane, 216
 of a circular cylinder, 223
 of a plane dioptr, 221
 of a spherical dioptr, 222
 Wavefront
 irrotational, 392, 393, 404, 405, 406
 phase, 391, 394, 402
 reconstruction, 391, 392, 396, 403, 404
 rotational, 392, 393, 403-405
 Wavelet, 392, 393, 396-398, 399, 401, 402

Haar, 392, 396-398, 410
transform, 392, 396, 399, 402
wavevector of grating, 63
wavevector of moire patterns, 63, 72
WDP (weak diffraction perturbation), 346,
356-358, 361-363, 367, 368, 370, 372,
380, 385

X

x-ray reflectometry, 35, 42, 45, 48

Z

z-dependent evolution, 347, 361, 363, 375,
376, 379, 380
Z-scan, 318, 323, 325, 331, 338
z-Transform, 393, 406

Advances in Optics: Reviews, Volume 1

Sergey Y. Yurish, Editor

'Advances in Optics: Reviews' Book Series is a comprehensive study of the field of optics, which provides readers with the most up-to-date coverage of optics, photonics and lasers with a good balance of practical and theoretical aspects. Directed towards both physicists and engineers this Book Series is also suitable for audiences focusing on applications of optics.

The Vol. 1 is devoted to various topics of optics and optic instrumentation, and contains 17 chapters written by 36 experts in the field from 15 countries: Brazil, China, Denmark, France, Germany, India, Japan, Mexico, Russia, Turkey, Slovenia, South Korea, UK, Ukraine and USA.

A clear comprehensive presentation makes these books work well as both a teaching resources and a reference books. The book is intended for researchers and scientists in physics and optics, in academia and industry, as well as postgraduate students.



ISBN 978-84-697-9436-4

